

Investigation of L-Amino Acid Hydroxylases



Tristan Jamie Smart

Christ Church

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2015

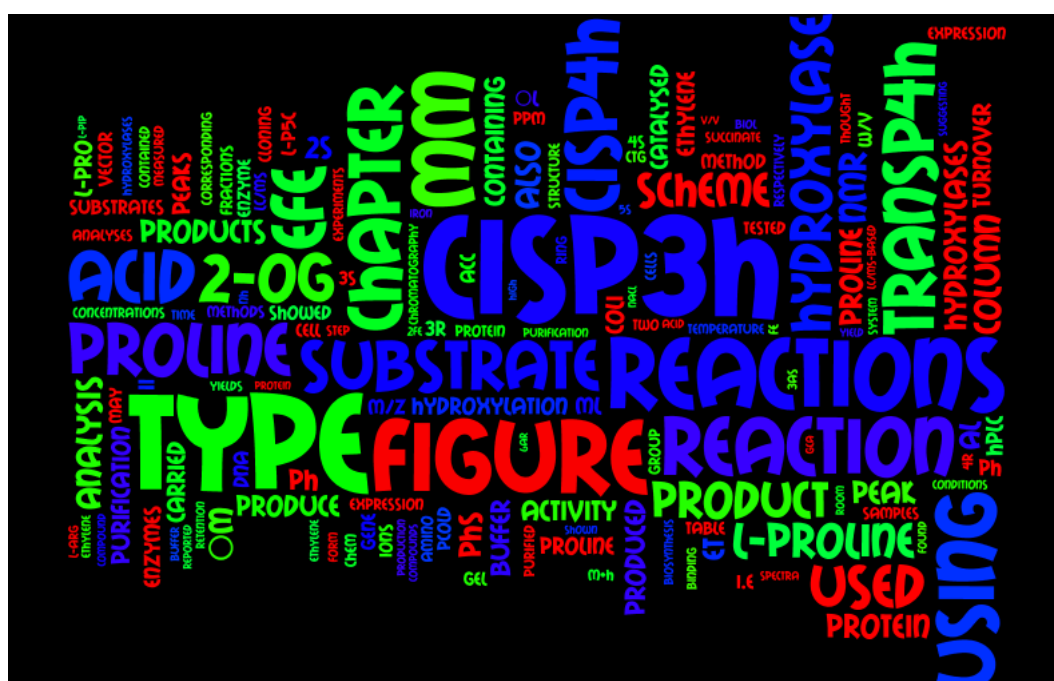
This thesis is
dedicated to my parents,
Linda and Nigel, for their
unwavering support and positive
influence. I am also grateful to
the rest of my family for
their love and support.

We keep moving forward, opening
new doors, and doing new things,
because we're curious, and curiosity
keeps leading us down new paths.

Walt Disney

Acknowledgements

Firstly, I'd like to thank Prof. Chris Schofield for the opportunity to work in a vibrant lab filled with many people; it is also a very 'free' lab, in which you can pursue many different projects and techniques, as well as work with a broad range of talented. Secondly, I'd like to thank the BBSRC for funding me and my project for 4 years (Doctoral Studentship in Molecular Biochemistry and Chemical Biology). Many notable members of the CJS lab (and related labs have worked with me, and I forever owe them. As a Part II student, during my undergraduate masters degree, I was trained by Akane Kawamura and Oliver King who were always helpful, without stifling my independence as a researcher. I also did interesting work with: Refaat Hamed (on the Proline Hydroxylases), Mike McDonough & Rashed Chowdury (Proline Hydroxylase crystallography), Ivanhoe Leung (NMR studies), Richard Hopkinson (Epigenetics and formaldehyde research), and Zhihong Zhang (on the Ethylene-Forming Enzyme), individuals I have learnt a tremendous amount from. Wendy Sobey and Dr Adam Hardy have also always been friendly and welcoming to me, and do a lot to maintain the day-to-day operations of the lab, from order lab consumables, fixing the network, to the general logistics of where people have to sit. I'd also like to thank the Mass Spectrometry department in Chemistry (in Oxford) (particularly James Wickens) and the NMR facilities managers (notably Prof Tim Claridge) for the continuous advice, and help when equipment is misbehaving. I'd like to thank the rest of the CJS lab for putting up with me, and allowing me carry out my research in an enjoyable and supportive atmosphere. Finally, I'd like to thank friends from outside of the lab particularly Dr Alex Drong and Dr Kate Roll, for being generally fun, and putting me up whilst I was (/am - at the time of writing this!) technically homeless (many thanks to: Paul Robertson (RIP), Meelis Lootus, Tristan Graye-Davies, Narin Hengrung, Chris Lander, Dr Jon Cleland, Andi Schmidt, Rich Davison, Weel-
iem Seah, Matthius Schmidt, James McCormick, ...).



Abstract

Ferrous-iron/2-oxoglutarate(2-OG)-dependent oxygenases have emerged as a widely distributed family of oxygenases. They have applications both in biocatalysis and as pharmaceutical/agrochemical targets. One of the first subfamilies of 2-OG oxygenases to be studied in detail were the prolyl hydroxylases. The work described in my thesis focused on substrate analogue studies on the proline hydroxylases and on a highly unusual type of 2-OG oxygenase that catalyses the production of ethylene from 2-OG.

The work involved the development of assays, in particular mass spectrometry and nuclear magnetic resonance spectroscopy, to screen a panel of substrate analogues. The results identified the need for an α -amino acid for productive catalysis, and were consistent with previous studies showing that the proline hydroxylases can accept different ring sizes, and can catalyse production of dihydroxylated products. New substrates for the proline hydroxylases including bicyclic compounds, which undergo a bifurcating pathway to give singly hydroxylated products, were also identified.

In a separate line of investigation, a novel ethylene forming enzyme was studied. Previous work had identified an arginine-oxidising 2-OG oxygenase from *Pseudomonas* spp that also produced ethylene. This observation is interesting given the key role of ethylene, as produced by aminocyclopropanecarboxylate (ACC) oxidase, in plant biology. ACC oxidase is structurally related to the 2-OG oxygenases but does not employ 2-OG during catalysis. Methods for the production of the recombinant ethylene-forming enzyme from *Pseudomonas* spp were developed as were assays for monitoring catalysis. The results reveal that the enzyme produces ethylene in an arginine-dependent manner and coproduces pyrroline-5-carboxylic acid (L-P5C) (the cyclic form of glutamate-5-semialdehyde). Crystal structures were obtained for the ethylene forming enzyme in the presence of a 2-OG analogue with and without arginine. The combined biochemical and structural studies imply that the enzyme has bifurcating pathway, with ethylene production arising from 2-OG and requiring arginine, but which is not oxidised; alternatively arginine can be oxidised to give Δ^1 -pyrroline-5-carboxylic acid (L-P5C) in a 'normal' 2-OG oxygenase type reaction with concomitant production of succinate. The results open up new possibilities for ethylene production in organisms ranging from bacteria to animals.

0.0 Abbreviations

Abbreviation	Meaning(s)
% (v/v)	Percentage Concentration by Volume (10% (v/v) $\equiv$ 10 ml in 100 ml (total) solvent)
% (w/v)	Percentage Concentration by 'Weight' (Mass) (10% (w/v) $\equiv$ 10 g in 100 ml (total) solvent)
1D	1-Dimensional
2D	2-Dimensional
2-OA	2-Oxoadipate
2-OG	2-Oxoglutarate/ α -Ketoglutarate
2,4-DNP	2,4-Dinitrophenylhydrazine
(3S)-OH-L-Arg	(3S)-Hydroxy-L-Arginine
4-HIL	(2S,3R,4S)-4-Hydroxy-L-Isoleucine
5-HT	5-Hydroxytryptamine
5-HTP	5-Hydroxy-L-Tryptophan
4Hyp	<i>trans</i> -4-hydroxy-L-proline
AAH(s)	L-Amino Acid Hydroxylase(s)
AAA(s)	Aromatic Amino Acid Hydroxylase(s)
ACC	1-Aminocyclopropane-1-Carboxylic Acid
ACCO	1-Aminocyclopropane-1-Carboxylic Acid Oxidase
ACE	Angiotensin-Converting Enzyme
ACN	Acetonitrile (CH ₃ CN)
AEC	Anion-Exchange Chromatography
AMP	Adenosine Monophosphate
ATP	Adenosine Triphosphate
L-Aze	L-Azetidine-2-Carboxylic Acid
L-Azp	L-Azepane-2-Carboxylic Acid
L-Azy	L-Azirine-2-Carboxylic Acid
BBO	Broadband Observe (in NMR Spectroscopy - Bruker probe type)
BBOX1	γ -Butyrobetaine Hydroxylase
Car	1-Carbapen-2-em-3-Carboxylic Acid
<i>car</i>	Carbapenem Gene Cluster
<i>carA</i>	Gene encoding the protein Carbapenam-3-Carboxylate Synthase (CarA)
<i>carB</i>	Gene encoding the protein Carboxymethylproline Synthase (CarB)
<i>carC</i>	Gene encoding the protein (5R)-Carbapenem-3-Carboxylate Synthase (CarC)
<i>cis</i> P3H Type I	L-Proline <i>cis</i> -3-Hydroxylase Type I
<i>cis</i> P4H Type I	L-Proline <i>cis</i> -3-Hydroxylase Type I
CMP	Carboxymethylproline (5-(Carboxymethyl)-L-Proline)
cCMP	<i>cis</i> -Carboxymethylproline (<i>cis</i> -5-(Carboxymethyl)-L-Proline)
tCMP	<i>trans</i> -Carboxymethylproline (<i>trans</i> -5-(Carboxymethyl)-L-Proline)
Me-tCMP	<i>trans</i> -Carboxymethylproline Methyl Ester (<i>trans</i> -5-(O-Methyl)-tCMP)
CNS	Central Nervous System
CODD	C-Terminal Oxygen-Dependent Degradation Domain
COSY	Correlation Spectroscopy
CS or CAS	Clavamate Synthase
DSS	4,4-Dimethyl-4-Silapentane-1-Sulfonic Acid (NMR Standard)
<i>efe</i>	The gene coding for EFE
EFE	Ethylene-Forming Enzyme
L-End	L-Enduracididine
ELSD	Evaporating Light Scattering Detector
ESI	Electrospray Ionisation

FA	Formic Acid
Fe(II)	Iron(II) ions with an oxidation state of +2 (ferrous iron)
Fe(III)	Iron(III) ions with an oxidation state of +3 (ferric iron)
Fe(VI)=O	Ferryl/Iron(IV)-oxo species (iron centre has +4 oxidation state)
FIH	Factor Inhibiting HIF
FMO(s)	Flavin-dependent Monooxygenase(s)
FPLC	Fast Protein Liquid Chromatography
L-G-5-SA	Glutamate-5-Semialdehyde
GloF	L-Proline <i>trans</i> -3/4-Hydroxylase
HIF	Hypoxia Inducible Factor
HIF- α	Hypoxia Inducible Factor α
HIF- β	Hypoxia Inducible Factor β
HPLC	High Performance Liquid Chromatography
IPTG	Isopropyl β -D-1-Thiogalactopyranoside
KMBA	2-Keto-4-Methylthiobutyric Acid
LC/MS	Liquid Chromatography/Mass Spectrometry
LdoA	Leucine 5-Hydroxylase (Npun F2185 from <i>Nostoc punctiforme</i> PCC 73102)
[M] ⁺	Molecular Ion (in mass spectrometry)
[M+H] ⁺	Protonated Molecular Ion (in mass spectra)
[M-H] ⁺	Deprotonated Molecular Ion (in mass spectra)
MALDI	Matrix-Assisted Laser Desorption/Ionisation
MALDI-TOF-MS	Matrix-Assisted Desorption/Ionisation Time-of-Flight Mass Spectrometry
MatB	Malonyl CoA Synthetase (Product of the <i>matB</i> Gene)
MPP(s)	Mannopeptimycin(s)
<i>mppO</i>	Gene encoding Enduracididine β -Hydroxylase
MppO	Enduracididine β -Hydroxylase
MS	Mass Spectrometry
NMR	Nuclear Magnetic Resonance
NODD	N-Terminal Oxygen-Dependent Degradation Domain
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
N-OG	N-Oxalylglycine
ODD(s)	2-Oxoglutarate/Iron(II)-Dependent Dioxygenase(s)
Ofd1	Prolyl 3,4-dihydroxylase Ofd1
OGFOD1	2-OG- and Fe(II)-Dependent Oxygenase Domain-Containing Protein 1
<i>orf5</i>	'Open reading frame 5' gene (<i>Nostoc sp</i> GSV224 nostopeptolides cluster)
ORF5	The product of <i>orf5</i>
L-P5C	L- Δ^1 -Pyrroline-5-Carboxylate
PAGE	Polyacrylamide Gel Electrophoresis
<i>pASA</i> / PAS	<i>para</i> -Aminosalicylic Acid
PCP	Peptidyl Carrier Protein
PDB	Protein Data Bank
PH(s)	Proline Hydroxylase(s)
L-PH(s)	L-Phenylalanine Hydroxylase(s)
PHD1	HIF-1 α Prolyl 4-Hydroxylase Domain 1
PHD2	HIF-1 α Prolyl 4-Hydroxylase Domain 2
PHD3	HIF-1 α Prolyl 4-Hydroxylase Domain 3
L-Phe	L-Phenylalanine
P _i	Inorganic Phosphate
PP _i	Inorganic Pyrophosphate
PLP	Pyridoxal Phosphate
PP3H	Procollagen Prolyl 3-Hydroxylase
PP4H	Procollagen Prolyl 4-Hydroxylase
L-Pro	L-Proline

L-Pip	L-Pipecolic Acid
RI	Refractive Index (Detection)
RNA	Ribonucleic Acid
RPS23	Small Ribosomal Subunit Protein
SadA	N-Substituted L-Amino Acid Hydroxylase A
SAM	S-Adenosyl-Methionine
SDS	Sodium Dodecyl Sulphate
Tbms	Tuberactinomycins
TCEP	<i>tris</i> (2-Carboxyethyl)Phosphine
TFA	Trifluoroacetic Acid
TFI	Transfer Function Inversion
<i>threo</i> -3-OH-L-Leu	<i>threo</i> -3-Hydroxy-L-Leucine
TOCSY	Total Correlation Spectroscopy
ge-1D TOCSY	(Selective) Gradient-Enhanced, 1D Total Correlation Spectroscopy
Tpa1p	Termination and Polyadenylation 1 Protein
TPH1	L-Tryptophan Hydroxylase 1
TPH2	L-Tryptophan Hydroxylase 2
<i>trans</i> P4H	L-Proline <i>trans</i> -4-Hydroxylase
L-Trp	L-Tryptophan
TSP- <i>d</i> ₄	3-(Trimethylsilyl)Propionic-2,2,3,3- <i>d</i> ₄ Acid, Sodium Salt (NMR Standard)
L-Tyr	L-Tyrosine
<i>vioC</i>	Gene encoding L-Arginine 3-Hydroxylase
pVHL	Von-Hippel-Lindau Tumour Suppressor Protein
VioC	L-Arginine 3-Hydroxylase
<i>vioD</i>	Gene encoding Capreomycin Synthase
VioD	Capreomycin Synthase

Collaborative Work

The crystallographic aspects of the work detailed in Chapter 3 were carried out in collaboration with Dr Rasheduzzaman Chowdhury, who aided crystal harvesting, data collection, and structure solution (by molecular replacement) for *Sinorhizobium meliloti* *cis*P4H.

Dr Refaat Hamed carried out some compound HPLC purifications and NMR studies detailed in Chapter 3.

Dr Zhihong Zhang carried out the crystallisation of EFE from *Pseudomonas syringae* pv. *phaseolicola* PK2; EFE crystal harvesting and crystal structures solutions were then carried out by Dr Michael McDonough. Some GC chromatograms were also measured by Dr Zhihong Zhang to detect gaseous product from EFE reactions.

Contents

1	Introduction	1
1.1	C—H Bond Activation	1
1.1.1	Cytochrome P450 Monooxygenases	2
1.1.2	Flavin-dependent Monooxygenases (FMOs)	3
1.1.3	2-Oxoglutarate/Iron(II)-Dependent Dioxygenases	4
1.2	L-Amino Acid Hydroxylases (AAHs)	8
1.2.1	Aromatic Amino Acid Hydroxylases (AAAHs)	9
1.2.1.1	L-Asparagine Hydroxylase (AsnO)	11
1.2.1.2	Aspartyl β -Hydroxylase (AspH)	12
1.2.1.3	Isoleucine Dioxygenase (IDO)	14
1.2.1.4	L-Ile Dihydroxylation Cascade (HilA and HilB)	15
1.2.1.5	Leucine 5-Hydroxylase (LdoA)	16
1.2.1.6	N-Substituted L-Amino Acid Hydroxylase A	18
1.2.1.7	L-Arginine 3-Hydroxylase (VioC)	18
1.2.1.8	Enduracididine β -Hydroxylase (MppO)	20
1.2.2	Proline Hydroxylases	22
1.2.2.1	Proline and its Derivatives	22
1.2.2.2	Hydroxyprolines in Nature	22
1.2.2.2.1	The Hypoxic Response	23
1.2.2.2.2	Collagen	24
1.2.2.2.3	Ribosome Hydroxylation	28
1.2.2.2.4	Cyclic Peptide Antibiotics	30
1.2.2.3	L-Proline <i>trans</i> -4-Hydroxylase (<i>trans</i> P4H)	32
1.2.2.4	L-Proline <i>cis</i> -3-Hydroxylase (<i>cis</i> P3H)	34
1.2.2.5	L-Proline <i>cis</i> -4-Hydroxylase (<i>cis</i> P4H)	39
1.2.2.6	L-Proline <i>trans</i> -3/4-Hydroxylase (GloF)	40
1.2.3	Ethylene Forming Enzymes	41
1.2.3.1	1-Aminocyclopropane-1-Carboxylate (ACC) Oxidase	42

CONTENTS

1.2.3.2	Ethylene-Forming Enzyme (EFE)	43
1.3	Project Aims	46
2	Cloning and Protein Expression	47
2.1	Proline Hydroxylase Production	47
2.1.1	<i>cis</i> P3H Type I Production	47
2.1.1.1	<i>cis</i> P3H Type I Expression Systems	48
2.1.1.2	The pET Expression System	49
2.1.2	<i>cis</i> P3H Type I Expression and Purification	52
2.1.2.1	Small-Scale <i>cis</i> P3H Type I Test Purification	54
2.1.2.2	Large-Scale <i>cis</i> P3H Type I Test Purification	57
2.1.2.3	<i>cis</i> P3H Type I Production by Auto-Induction	61
2.1.2.4	<i>cis</i> P3H Type I Purification Development	68
2.1.2.4.1	<i>cis</i> P3H Type I Affinity Purification	72
2.2	<i>cis</i> P4H and <i>trans</i> P4H Production	82
2.2.1	Takara pCOLD I Vector Cloning	83
2.2.2	<i>cis</i> P4H and <i>trans</i> P4H Nickel-Affinity Purification	88
2.2.2.1	<i>cis</i> P4H Purification Development	93
2.2.2.2	<i>trans</i> P4H Purification Development	99
2.2.3	Takara pCOLD I <i>cis</i> P3H Type I Expression	103
2.3	Ethylene-Forming Enzyme (EFE) Production	105
2.3.1	Subcloning the <i>efe</i> Gene into Takara pCOLD I	105
2.3.2	EFE Production	107
2.3.2.1	Recombinant EFE Purification	108
2.4	Summary	120
3	The L-Proline Hydroxylases (PHs)	121
3.1	PH Discoveries	121
3.2	Proline Hydroxylase Assays	124
3.2.1	[¹⁴ C]-Radioactivity Detection	124
3.2.2	Reverse-Phase HPLC Assays	127
3.3	PH Assay Development	130
3.3.1	Proline Hydroxylase NMR Assays	131
3.3.1.1	Proline Hydroxylase End-Point NMR Assays	132
3.3.1.2	PH NMR Timecourse Assay	133
3.3.2	LC/MS-Based PH Assays	144
3.4	PH Substrate Analogue Screening	153

3.4.1	Substrate Screening by [U - ^{14}C]-2-OG Turnover	153
3.4.2	Substrate Screening by HPLC	157
3.4.3	Robust Methods for L-Proline-Derivative Hydroxylation	159
3.4.4	Proline Analogue Substrate Screening	164
3.4.4.1	Effects of Amino Acid α -Carbon-Stereochemistry on PH Catalysis	165
3.4.4.1.1	Turnover Summary: D/L-Stereochemistry .	170
3.4.4.2	Effects of Amino- & Carboxylic-Functionalisation on PH Catalysis	171
3.4.4.2.1	Turnover Summary: Amino- & Carboxylic- Functionalisation	179
3.4.4.3	Effects of Carbon Backbone Substitutions	181
3.4.4.3.1	Alkylated and Isoprenylated Proline Ana- logues	183
3.4.4.3.2	Turnover Summary: Alkyl & Other Non- Polar Substituents	191
3.4.4.3.3	Halogenated Proline Analogues	192
3.4.4.3.4	Turnover Summary: Halogenated Proline Derivatives	210
3.4.4.3.5	Hydroxylated Proline Analogues	211
3.4.4.3.6	Turnover Summary: Hydroxylated Proline Derivatives	222
3.4.4.3.7	Carboxymethylproline (CMP) Derivatives .	223
3.4.4.3.8	Turnover Summary: Carboxymethylprolines (CMPs)	229
3.4.4.3.9	α,α' -Dicarboxy Proline Derivatives	230
3.4.4.4	Proline analogues of Varying Ring Size	230
3.4.4.4.1	Turnover Summary: Substrate Ring Size . .	241
3.4.4.5	Proline analogues with Fused/Bridged Ring Systems	242
3.4.4.5.1	Turnover Summary: Bicyclic Proline Ana- logues	257
3.5	Proline Hydroxylase (PH) Crystallography	259
3.5.1	PH Crystallisation Screening	259
3.5.1.1	<i>cis</i> P4H Crystallography	260
3.5.1.2	<i>cis</i> P4H D108 Variants	266
3.5.1.3	Summary of PH Crystallographic Work	268

3.6	Conclusions	269
4	Ethylene Biosynthesis	271
4.1	Industrial Ethylene Production	271
4.2	Pathways of Ethylene Biosynthesis	271
4.2.1	Plant Metabolism of L-Methionine to Ethylene	272
4.2.2	Microbial Ethylene Biosynthesis	277
4.2.2.1	Ethylene Biosynthesis from 2-OG	278
4.3	Studies on the Ethylene-Forming Enzyme (EFE) from <i>Pseudomonas syringae</i> pv. <i>phaseolicola</i> PK2	285
4.3.1	Solution-State NMR Detection Assays	285
4.3.2	Solution-State LC/MS-Detection Assays	293
4.3.2.1	L-Arg Substrate Analogue Studies	296
4.3.3	L-P5C Purification and Characterisation	304
4.3.4	Gas Chromatograph (GC) Assays	308
4.3.5	EFE Structural Studies	314
4.4	Conclusions	325
5	Materials and Methods	327
5.1	Materials	327
5.1.1	Cloning and Molecular Biology	327
5.1.1.1	DNA gel electrophoresis	328
5.1.2	Bacterial Growth Media	330
5.1.2.1	Bacterial Growth Media	331
5.1.3	Protein Purification	335
5.1.3.1	Buffers: Stock solutions	335
5.1.3.2	SDS-PAGE Gel Formulations	335
5.1.3.2.1	Staining and Destaining	337
5.1.3.3	FPLC: Columns and Equipment	338
5.1.3.3.1	Protein Liquid Chromatography (FPLC)	338
5.1.3.3.2	Columns	339
5.1.3.3.3	Purification Buffer Formulations	342
5.1.4	Crystallography	344
5.1.5	Mass Spectrometry	344
5.1.5.1	LC/MS Equipment	344
5.1.5.2	GC Detection Equipment	345
5.1.6	NMR Spectroscopy	346

5.1.6.1	Real-Time Detection	346
5.2	Methods	347
5.2.1	Microbiology	347
5.2.1.1	<i>E. coli</i> Competent Cell Preparation	347
5.2.1.2	Heat Shock DNA Transformation Protocol	348
5.2.1.3	Recombinant <i>E. coli</i> Growth for Enzyme Purification	350
5.2.1.3.1	Cell Growth for Large-Scale Production of <i>cisP3H</i> Type I	351
5.2.1.3.2	Large-Scale Production of <i>cisP4H</i> , <i>transP4H</i> , and EFE	351
5.2.1.4	Purification Methods	352
5.2.1.4.1	Ni-Affinity Chromatography	352
5.2.1.4.2	AEC: DEAE Sepharose FF	352
5.2.1.4.3	AEC: MonoQ	353
5.2.1.4.4	Size-Exclusion Chromatography	353
5.2.2	Enzymology: Detection Assays	354
5.2.2.1	LC/MS-Detection Assays	354
5.2.2.2	GC-Detection Assay	356
5.2.2.3	Preparative Scale Reactions: Product Purification	356
5.2.2.4	NMR-Detection Assays	359
A	Cloning, Sequences, and Purification	361
A.1	Purification Methods - IMPACT™-CN System	361
A.2	Proline Hydroxylase Sequences	362
A.2.1	<i>cisP3H</i> Type I Sequences	362
A.2.2	<i>cisP4H</i> Sequences	363
A.2.3	<i>transP4H</i> Sequences	364
A.2.3.1	Takara pCOLD I <i>N</i> -His ₆ -Tag Sequence (5'- <i>NdeI</i> and 3'- <i>BamHI</i> Cloning Sites)	366
A.2.3.1.1	<i>N</i> -His ₆ - <i>cisP4H</i> Theoretical Properties	366
A.2.3.1.2	<i>N</i> -His ₆ - <i>transP4H</i> Theoretical Properties	367
A.2.4	EFE Sequences	367
A.2.4.1	Takara pCOLD I <i>N</i> -His ₆ -Tag Sequence (5'- <i>SacI</i> and 3'- <i>PstI</i> Cloning Sites)	369
A.2.4.1.1	<i>N</i> -His ₆ -EFE Theoretical Properties	369
	References	370

CONTENTS

List of Figures

1.1	CYP450 enzymes have a wide variety of substrates with different structures; some of these products are displayed above, with the respective substrate names in blue.	2
1.2	The cytochrome P450 (CYP450) iron(III) binding centre (haem site): it contains a protophyrin-IX prosethetic group with a proximal cysteine ligand.	2
1.3	Examples of different reaction types catalysed by ODDs.	5
1.4	(a) Structure of 2-oxoglutarate and (b) a view of the ferrous iron binding site in ODDs.	6
1.5	The structure of CDA4a, a calcium-dependent antibiotic.	11
1.6	The structure of syringomycin E.	12
1.7	The structure of the <i>N</i> -butyryl nonapeptide mimic of a syringomycin lipononapeptide; it is a substrate of AspH.	13
1.8	The structure of nostopeptolides A1, A2, and A3.	17
1.9	The structure of Viomycin B, a non-ribosomal peptide antibiotic. . . .	19
1.10	The structure of mannopeptimycin β (MPP β).	21
1.11	(a) Substitutions on the L-proline ring induce relatively fixed stereochemistry owing to the relatively rigid nature of the ring; (b) L-proline adopts 2 major, envelope ring conformations (C4- <i>endo</i> and C4- <i>exo</i>) which rapidly interconvert.	22
1.12	The structures of (a) echinocandin B and (b) actinomycin I.	23
1.13	The chronic hypoxic response mechanism in animals: the HIF mechanism.	25
1.14	Molecular and structural aspects of collagen.	27
1.15	The structure of Etamycin.	30
1.16	The structure of Telomycin.	32

LIST OF FIGURES

1.17 A comparison between the deduced amino acid sequences of <i>cis</i> P3H Type I and Type II. The sequence alignment and conservation scoring was performed by using PRALINE. ^{202,203}	36
1.18 Views from X-ray crystal structures of <i>cis</i> P3H Type II (both are heterodimers): (a) the apo structure (PDB 1E5R) and (b) the iron(II)-bound form (PDB 1ESR), showing the active site.	38
1.19 The structure of ethylene (ethene).	41
2.1 The structures of (a) allolactose and (b) isopropyl β -D-1-thiogalactopyranoside (IPTG).	50
2.2 The Novagen pET-24-a(+) expression vector map.	51
2.3 The structures of polyethylenimine (PEI) isoforms; the simplest form of PEI is a linear chain, but PEI can also be a branched-chain polymer.	55
2.4 Small-scale, anion-exchange-chromatography (AEC) purification of <i>cis</i> P3H Type I.	56
2.5 Large-scale, anion-exchange-chromatography (AEC) purification of <i>cis</i> P3H Type I.	58
2.6 Preparative-scale size-exclusion chromatography of <i>cis</i> P3H Type I fractions.	60
2.7 Anion-exchange-chromatography (AEC) purification of <i>cis</i> P3H Type I produced by <i>E. coli</i> cells grown on ZYM-5052 auto-induction media.	64
2.8 Size-exclusion chromatography of <i>cis</i> P3H Type I produced by <i>E. coli</i> cells grown on ZYM-5052 auto-induction media.	65
2.9 Mono Q anion-exchange chromatography of <i>cis</i> P3H Type I produced by <i>E. coli</i> cells grown on ZYM-5052 auto-induction media.	67
2.10 Anion-exchange-chromatography (AEC) purification of <i>cis</i> P3H Type I produced by <i>E. coli</i> cells grown on ZYM-5052 auto-induction media.	69
2.11 The UV _{280nm} -absorbance FPLC chromatograms for the purification of <i>cis</i> P3H Type I by (a) hydrophobic-interaction-chromatography (HIC) and (b) size-exclusion chromatography (SEC).	71
2.12 The binding of a N-terminally hexahistidine-tagged protein to NTA-resin-immobilised Ni ²⁺ ions of an IMAC column.	72
2.13 The pNIC-28-Bsa4 <i>E. coli</i> Expression Vector. ²⁹⁶	74
2.14 Polymerase chain reaction (PCR) cycle for gene amplification. ³⁰⁵⁻³⁰⁷	77
2.15 The structures of betaine and DMSO.	78
2.16 Visualisation of the <i>cis</i> P3H Type I PCR product DNA.	79

2.17 Depiction of the agarose-gel-electrophoresis analysis of the pNIC-28-Bsa4 BsaI-digestion products and the gel extraction of the bands corresponding to linear, digested pNIC-28-Bsa4 (- <i>sacB</i>).	80
2.18 Agarose-gel-electrophoresis visualisation of products (<i>cis</i> P3H Type I inserts) from PCRs using colony plasmids as templates.	81
2.19 SDS-PAGE analysis of the fractions from the nickel-affinity purification of <i>N</i> -His ₆ - <i>cis</i> P3H Type I produced by <i>E. coli</i> cells grown on ZYM-5052 auto-induction medium.	82
2.20 The Takara pCOLD I <i>E. coli</i> cold shock expression vector. ³¹¹	84
2.21 Agarose-gel-electrophoresis visualisation of the pCOLD I NdeI/BamHI double restriction digestion products.	86
2.22 The gel purification of the linear pCOLD I vector and the <i>cis</i> P4H and <i>trans</i> P4H gene fragments.	87
2.23 Agarose-gel-electrophoresis analysis of NdeI/HindIII-digested colony plasmids.	88
2.24 Immobilised-nickel-affinity chromatography of <i>cis</i> P4H.	91
2.25 Immobilised-nickel-affinity chromatography of <i>trans</i> P4H.	92
2.26 Size-exclusion purification of <i>cis</i> P4H as a second column-step after IMAC purification.	94
2.27 Immobilised-nickel-affinity chromatography of <i>cis</i> P4H.	96
2.28 Mono Q anion-exchange chromatography of <i>cis</i> P4H.	97
2.29 Size-exclusion purification of <i>cis</i> P4H as a second column-step after IMAC purification.	98
2.30 DEAE Sepharose FF anion-exchange chromatography of <i>trans</i> P4H.	100
2.31 Size-exclusion purification of <i>trans</i> P4H as a second column-step after IMAC purification.	102
2.32 Agarose-gel-electrophoresis analysis of SacI/PstI-digested colony plasmids.	107
2.33 SDS-PAGE analysis of cell lysate from EFE expression trials.	108
2.34 Immobilised-nickel-affinity chromatography of EFE.	111
2.35 Size-exclusion purification of EFE as a second column-step after IMAC purification.	113
2.36 Immobilised-nickel-affinity chromatography of EFE.	115
2.37 DEAE Sepharose FF anion-exchange chromatography of EFE.	117
2.38 Size-exclusion purification of EFE as a second column-step after IMAC purification	119

LIST OF FIGURES

3.1	<i>trans</i> P4H inhibitors: aromatic 2-OG analogues including pyridine dicarboxylic acids (PDCAs) and 3,4-dihydroxybenzoic acid (3,4-DHBA). ¹⁹³	127
3.2	Proline hydroxylase Trial ¹ H-NMR Assay: a <i>cis</i> P3H Type I batch was tested for activity.	134
3.3	Example <i>cis</i> P3H Type I ¹ H-NMR timecourse spectrum.	136
3.4	<i>cis</i> P3H Type I NMR Timecourse experiment using a high concentration (15 μM) of <i>cis</i> P3H Type I and moderate substrate concentrations (1.0 mM L-Pro and 1.0 mM succinate).	138
3.5	<i>cis</i> P3H Type I NMR Timecourse experiment using a high concentration (10 μM) of <i>cis</i> P3H Type I and moderate substrate concentrations (1.0 mM L-Pro and 1.0 mM succinate).	139
3.6	<i>cis</i> P3H Type I NMR Timecourse experiment using a relatively high concentration (5.0 μM) of <i>cis</i> P3H Type I and high substrate concentrations (5.0 mM L-Pro and 5.0 mM succinate).	140
3.7	<i>cis</i> P3H Type I NMR Timecourse experiment using a relatively low concentration (1.0 μM) of <i>cis</i> P3H Type I and low substrate concentrations (250 μM L-Pro and 250 μM succinate).	142
3.8	<i>cis</i> P3H Type I NMR Timecourse experiment using a relatively low concentration (0.5 μM) of <i>cis</i> P3H Type I and low substrate concentrations (250 μM L-Pro and 250 μM succinate).	143
3.9	The architecture and properties of SiELC Technologies Primesep™ HPLC columns.	146
3.10	SiELC Primesep™ 100 column applications.	148
3.11	LC/MS-based analysis of proline hydroxylase reactions with L-Pro [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	151
3.12	LC/MS-based analysis of proline hydroxylase reactions with L-Pro [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	152
3.13	L-Proline analogues tested for stimulation of <i>S. griseoviridis</i> P8648 <i>trans</i> P4H 2-OG turnover. ^{191,193,256}	154
3.14	L- and D-stereochemistry nomenclature of biological molecules . . .	166
3.15	Proline analogues with D-stereochemistry tested for proline hydroxylase substrate activity.	168
3.16	LC/MS-based analysis of proline hydroxylase reactions with D-Pro [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	168

3.17 Amino- and carboxylic-acid-functionalised L-proline test substrate: the substrates tested for substrate activity by LC/MS and/or NMR assay.	171
3.18 LC/MS-based analysis of proline hydroxylase reactions with <i>N</i> -methyl- L-proline (<i>N</i> -Me-L-Pro) [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	174
3.19 LC/MS-based analysis of proline hydroxylase reactions with <i>N</i> -methyl- L-proline (<i>N</i> -Me-L-Pro) [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	175
3.20 COSY-NMR analysis of the product from the <i>cis</i> P3H Type I reaction with <i>N</i> -methyl-L-proline (<i>N</i> -Me-L-Pro).	176
3.21 NOESY-NMR analyses of the product from the <i>cis</i> P3H Type I reaction with <i>N</i> -methyl-L-proline (<i>N</i> -Me-L-Pro).	177
3.22 LC/MS-based analysis of proline hydroxylase reactions with L-proline methyl ester (L-Pro-OMe) [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	178
3.23 Proline analogues with a carbon-backbone functionality.	182
3.24 2-Methyl-L-proline (2-Me-L-Pro) and (<i>R</i>)-2-(4-methylbenzyl)-L-proline [(<i>R</i>)-2-(4-Me-Bn)-L-Pro] were tested for proline hydroxylase substrate activity.	183
3.25 LC/MS-based analysis of proline hydroxylase reactions with 2-methyl- L-proline (2-Me-L-Pro) [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	185
3.26 COSY-NMR analysis of the product from the <i>cis</i> P4H reaction with <i>N</i> - methyl-L-proline (2-Me-L-Pro).	186
3.27 A <i>trans</i> -4-(2-propynyl)-L-proline [<i>trans</i> -4-(HC≡C-CH ₂)-L-Pro] substrate with tested for proline hydroxylase turnover	187
3.28 LC/MS-based analysis of proline hydroxylase reactions with <i>trans</i> -4- (2-propynyl)-L-proline [<i>trans</i> -4-(HC≡C-CH ₂)-L-Pro] [(a) <i>cis</i> P3H Type I and (b) <i>cis</i> P4H].	188
3.29 COSY-NMR analysis of the product from the <i>cis</i> P3H Type I/ <i>cis</i> P4H reactions with <i>trans</i> -4-(2-propynyl)-L-proline [<i>trans</i> -4-(HC≡C-CH ₂)- L-Pro].	189
3.30 ¹ H-NMR analyses of proline hydroxylase reactions using a <i>cis</i> -4-fluoro- L-proline (<i>cis</i> -4-F-L-Pro) substrate [(a) <i>cis</i> P3H Type I].	194
3.31 ¹ H-NMR analyses of proline hydroxylase reactions using a <i>cis</i> -4-fluoro- L-proline (<i>cis</i> -4-F-L-Pro) substrate [(a) <i>trans</i> P4H].	195

LIST OF FIGURES

3.32 Proton-decoupled ^{19}F -NMR analyses of proline hydroxylase reactions using a <i>cis</i> -4-fluoro-L-proline (<i>cis</i> -4-F-L-Pro) substrate [(a) <i>cis</i> P3H Type I and (b) <i>trans</i> P4H].	197
3.33 LC/MS-based analysis of proline hydroxylase reactions with <i>cis</i> -4-fluoro-L-proline (<i>cis</i> -4-F-L-Pro) [(a) <i>cis</i> P3H Type I and (b) <i>trans</i> P4H].	198
3.34 ^1H - and TOCSY-NMR analyses of <i>cis</i> -4-F-L-Pro <i>cis</i> P3H Type I reaction products in both (a) and (b) complete reaction mixtures and in (c) and (d) LC/MS-purified forms.	200
3.35 COSY-NMR analysis of a product from the <i>cis</i> P3H Type I/ <i>cis</i> P4H reactions with <i>cis</i> -4-fluoro-L-proline (<i>cis</i> -4-F-L-Pro).	201
3.36 ^1H -NMR analyses of proline hydroxylase reactions using a <i>trans</i> -4-fluoro-L-proline (<i>trans</i> -4-F-L-Pro) substrate [(a) <i>cis</i> P3H Type I and (b) <i>cis</i> P4H].	204
3.37 Proton-decoupled ^{19}F -NMR analyses of proline hydroxylase reactions using a <i>trans</i> -4-fluoro-L-proline (<i>trans</i> -4-F-L-Pro) substrate [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) 'no enzyme' negative control].	206
3.38 LC/MS-based analysis of proline hydroxylase reactions with <i>trans</i> -4-fluoro-L-proline (<i>trans</i> -4-F-L-Pro) [(a) <i>cis</i> P3H Type I, (b) and (c) <i>cis</i> P4H].	208
3.39 COSY-NMR analysis of a product from the <i>cis</i> P3H Type I/ <i>cis</i> P4H reactions with <i>trans</i> -4-fluoro-L-proline (<i>trans</i> -4-F-L-Pro).	209
3.40 Hydroxylated proline derivatives tested for proline hydroxylase substrate activity	211
3.41 LC/MS-based analysis of proline hydroxylase (<i>trans</i> P4H) reactions with <i>trans</i> -3-hydroxy-L-proline (<i>trans</i> -3-OH-L-Pro).	212
3.42 COSY-NMR analysis of a product from the <i>trans</i> P4H reactions with <i>trans</i> -3-hydroxy-L-proline (<i>trans</i> -3-OH-L-Pro).	213
3.43 LC/MS-based analysis of proline hydroxylase (<i>cis</i> P3H Type I) reactions with <i>cis</i> -4-hydroxy-L-proline (<i>cis</i> -4-OH-L-Pro).	214
3.44 COSY-NMR analysis of a product from <i>cis</i> P3H Type I reactions with <i>cis</i> -4-hydroxy-L-proline (<i>cis</i> -4-OH-L-Pro).	215
3.45 LC/MS-based analysis of proline hydroxylase (<i>cis</i> P3H Type I) reactions with <i>cis</i> -4-hydroxy- <i>N</i> -methyl-L-proline (<i>cis</i> -4-OH- <i>N</i> -Me-L-Pro). .	217
3.46 LC/MS-based analysis of proline hydroxylase reactions with <i>trans</i> -4-hydroxy-L-proline (<i>trans</i> -4-OH-L-Pro) [(a) <i>cis</i> P3H Type I and (b) <i>cis</i> P4H].	219

3.47 Carboxymethylproline (CMPs) derivatives tested as proline hydroxylase substrates.	223
3.48 LC/MS-based analysis of proline hydroxylase (<i>cis</i> P3H Type I) reactions with <i>trans</i> -5-(2-methoxy-2-oxoethyl)-L-proline ("Me- <i>t</i> CMP"). . .	224
3.49 COSY and NOESY spectra of the HPLC-purified product of <i>cis</i> P3H-catalysed hydroxylation of <i>trans</i> -5-(2-methoxy-2-oxoethyl)-L-proline ("Me- <i>t</i> CMP") (as well as its ¹ H-, ge-1D-selective TOCSY-spectra). . . .	225
3.50 (2 <i>S</i> ,5 <i>S</i>)-5-[(<i>S</i>)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline ["(6 <i>S</i>)-6-(isopenten-2-yl)-Me- <i>t</i> CMP"] was tested as a proline hydroxylase substrate.	226
3.51 LC/MS-based analysis of proline hydroxylase (<i>cis</i> P3H Type I) reactions with (5 <i>S</i>)-5-[(<i>S</i>)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline ["(6 <i>S</i>)-6-(isopenten-2-yl)-Me- <i>t</i> CMP"].	228
3.52 α,α' -Dicarboxy proline derivatives were tested as proline hydroxylases; compounds included (5 <i>S</i>)-5-carboxy-L-proline [(2 <i>R</i> ,5 <i>S</i>)-5-(COOH)-L-Pro], (5 <i>S</i>)-5-carboxy- <i>N</i> -methyl-L-proline [(2 <i>R</i> ,5 <i>S</i>)-5-(COOH)- <i>N</i> -Me-L-Pro], and (6 <i>S</i>)-6-carboxy-L-pipecolic acid [(2 <i>R</i> ,6 <i>S</i>)-6-(COOH)-L-Pip].	230
3.53 Proline analogues with different ring sizes were tested as proline hydroxylase (<i>cis</i> P3H Type I, <i>cis</i> P4H, and <i>trans</i> P4H) substrates.	231
3.54 LC/MS-based analysis of proline hydroxylase [(a) <i>cis</i> P3H Type I and (b) <i>cis</i> P4H] reactions with L-azetidine-2-carboxylic acid (L-Aze). . . .	232
3.55 LC/MS-based analysis of proline hydroxylase reactions with L-pipecolic acid (L-Pip) [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	234
3.56 COSY-NMR analysis of a product from the <i>cis</i> P4H reactions with L-pipecolic acid (L-Pip).	235
3.57 LC/MS-based analysis of proline hydroxylase reactions with <i>N</i> -methyl-L-pipecolic acid (L-Pip) [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	237
3.58 LC/MS-based analysis of proline hydroxylase reactions with L-azepane-2-carboxylic acid (L-Azp) [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	239
3.59 COSY-NMR analysis of a product from the <i>cis</i> P3H Type I reaction with L-azepane-2-carboxylic acid (L-Azp).	240
3.60 Proline analogues with fused and bridged ring systems were tested as proline hydroxylase (<i>cis</i> P3H Type I, <i>cis</i> P4H, <i>trans</i> P4H)	242

LIST OF FIGURES

3.61 LC/MS-based analysis of proline hydroxylase reactions with (1 <i>S</i> ,2 <i>S</i> ,5 <i>R</i>)-3-azabicyclo[3.1.0]hexane-2-carboxylic acid [(1 <i>S</i> ,2 <i>S</i> ,5 <i>R</i>)-3,4-CH ₂ -L-Pro] [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	244
3.62 Ring conformations of proline analogues with a C4,C5-fused 5-membered ring	245
3.63 LC/MS-based analysis of proline hydroxylase reactions with (2 <i>S</i> ,3 <i>aS</i> ,6 <i>aS</i>)-octahydrocyclopenta[<i>b</i>]pyrrole-2-carboxylic acid [(2 <i>S</i> ,3 <i>aS</i> ,6 <i>aS</i>)-H ₈ -cPP2CA] [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	247
3.64 Proton-, TOCSY-, COSY-NMR analysis of the ‘major’ product from <i>cis</i> P3H Type I and <i>cis</i> P4H reactions using a (2 <i>S</i> ,3 <i>aS</i> ,6 <i>aS</i>)-H ₈ -cPP2CA substrate [(2 <i>S</i> ,3 <i>aS</i> ,4 <i>R</i> ,6 <i>aR</i>)-4-OH-H ₈ -cPP2CA].	249
3.65 LC/MS-based analysis of proline hydroxylase reactions with (2 <i>S</i> ,3 <i>aR</i> ,6 <i>aR</i>)-octahydrocyclopenta[<i>b</i>]pyrrole-2-carboxylic acid [(2 <i>S</i> ,3 <i>aR</i> ,6 <i>aR</i>)-H ₈ -cPP2CA] [(a) <i>cis</i> P3H Type I, and (b) <i>trans</i> P4H].	250
3.66 COSY-NMR analysis of the ‘major’ product from the <i>cis</i> P3H Type I reaction with (2 <i>S</i> ,3 <i>aR</i> ,6 <i>aR</i>)-octahydrocyclopenta[<i>b</i>]pyrrole-2-carboxylic acid [(2 <i>S</i> ,3 <i>aR</i> ,6 <i>aR</i>)-H ₈ -cPP2CA].	252
3.67 NOESY-NMR analyses of the ‘major’ product from the <i>cis</i> P3H Type I reaction with (2 <i>S</i> ,3 <i>aR</i> ,6 <i>aR</i>)-octahydrocyclopenta[<i>b</i>]pyrrole-2-carboxylic acid [(2 <i>S</i> ,3 <i>aR</i> ,6 <i>aR</i>)-H ₈ -cPP2CA].	253
3.68 LC/MS-based analysis of proline hydroxylase (<i>cis</i> P3H Type I) reactions with (2 <i>S</i> ,3 <i>aS</i> ,7 <i>aS</i>)-octahydro-1H-indole-2-carboxylic acid (H ₈ -L-I2CA).	254
3.69 COSY-NMR analysis of the ‘major’ product from the <i>cis</i> P3H Type I reaction with (2 <i>S</i> ,3 <i>aS</i> ,7 <i>aS</i>)-octahydro-1H-indole-2-carboxylic acid (H ₈ -L-I2CA).	255
3.70 NOESY-NMR analyses of the ‘major’ product from the <i>cis</i> P3H Type I reaction with (2 <i>textitS</i> ,3 <i>aS</i> ,7 <i>aS</i>)-octahydro-1H-indole-2-carboxylic acid (H ₈ -L-I2CA).	256
3.71 LC/MS-based analysis of proline hydroxylase [(a) <i>cis</i> P3H Type I and (b) <i>trans</i> P4H] reactions with decanhydroisoquinolone-3-carboxylic acid (H ₁₀ -IQ3CA).	257
3.72 3-Drop, 96-well crystallisation plate.	261
3.73 <i>cis</i> P4H is tetramer in the unit cell of <i>P</i> 2 ₁ 2 ₁ 2 ₁ protein crystals.	262
3.74 ‘Cartoon’ depiction of the secondary and tertiary structure of apo- <i>cis</i> P4H monomer units.	263

3.75	Molecular alignment of <i>cis</i> P3H Type II (Fe(II)-bound form - PDB accession number 1E5S) and <i>cis</i> P4H (apo) crystal structures.	264
3.76	Active site residues of the Apo- <i>cis</i> P4H crystal structure.	265
3.77	SDS-PAGE analysis of nickel-affinity purifications of D108 <i>cis</i> P4H variants.	266
3.78	LC/MS-based analysis of <i>cis</i> P4H variant reactions [(a) D108E (b) D108P].	267
3.79	<i>cis</i> P4H variant timecourse reactions [(a) D108E (b) D108P].	268
4.1	NMR analyses of EFE reactions.	288
4.2	Further NMR analyses of EFE reactions.	290
4.3	¹ H Spectrum of DL-P5C.	291
4.4	¹ H-Spectrum of substrate-uncoupled turnover of 2-OG by EFE: a comparison with spectra measured for a substrate-coupled EFE reaction (with L-Arg) and a ‘no-enzyme’ control sample.	292
4.5	LC/MS-based analysis of EFE reactions.	295
4.6	L-Arginine substrate analogues were tested for turnover by EFE.	296
4.7	D- and L-Arginine EFE Reactions: the LC/MS analysis of the reaction products.	297
4.8	EFE reactions with a L-canavanine substrate: the LC/MS analysis of the reaction products.	299
4.9	EFE reactions with a <i>N</i> ^G -Me-L-Arg substrate: the LC/MS analysis of the reaction products.	301
4.10	EFE Reactions with the <i>N</i> ^G , <i>N</i> ^G -DiMe-L-Arg substrate: the LC/MS analysis of the reaction products.	302
4.11	EFE Reactions with the <i>N</i> ^G -Nitro-L-Arg substrate: the LC/MS analysis of the reaction products.	303
4.12	LC/MS Purification of L-P5C from EFE reactions: the TIC (total-ion count) chromatograms of each purification run.	305
4.13	LC/MS Purification of L-P5C from EFE reactions: (a) the XIC (extracted-ion count) chromatograms of each purification run and (b) the mass spectra for the L-P5C peaks (for which the major peak <i>m/z</i> = ~114).	305
4.14	¹ H-NMR spectrum of LC/MS-purified L-P5C (compared EFE reaction spectra).	307
4.15	GC analysis of EFE reaction headspaces: signals were detected using an FID.	309

LIST OF FIGURES

4.16 GC analysis of EFE reaction headspaces; signals were detected using an FID.	311
4.17 Proposed fragmentation mechanism of EFE-catalysed fragmentation of (a) 2-oxoglutarate and (b) 2-oxoadipate (2-OA) (a compound analogous to 2-OG): stereoelectronic considerations.	312
4.18 Proposed potential EFE substrates: different 2-OG analogues could be transformed into specific E/Z stereoisomers.	313
4.19 View from a crystal structure of EFE showing BisTris bound to active-site Fe ^{II}	315
4.20 View of the active site of the EFE-Fe ^{II} -(2-OG) complex crystal structure; key residues are shown.	316
4.21 A comparison of active site residues of the EFE structures with 2-OG bound (green) and without 2-OG bound (blue): both structures were overlaid and analysed using PyMOL modelling software.	317
4.22 The active site residues of BBOX1 as seen in a crystal structure (PDB ID: 3MS5) containing complexed N-OG (N-oxalylglycine, a 2-OG mimic), Ni ²⁺ (in place of Fe ²⁺), and 2-(2-carboxyethyl)-1,1,1-trimethyl diazanium (2-(2-(COOH)Et)-1,1,1-TriMe-Diazanium, a γ -butyrobetaine substrate analogue). ⁴¹²	318
4.23 The deviation of EFE-bound 2-OG from a coplanar conformation. . .	319
4.24 Stereoelectronic considerations of EFE-mediated 2-OG decarboxylation.	319
4.25 The active site residues of a <i>cis</i> P4H crystal structure (PDB ID: 4P7W) containing complexed 2-OG, Co ²⁺ (in place of Fe ²⁺), and L-proline. [?]	320
4.26 View from a structure of EFE in complex with Fe(II), N-OG, and L-Arg.	322
5.1 GC Method for Ethylene Detection: outline of 11 min GC profile. . .	346
5.2 SiELC Primesep 100 HPLC Method for LC/MS analysis of assay samples.	355
5.3 SiELC Primesep 100 HPLC (30 min) method for LC/MS purification of ‘polar’ proline analogues.	357
5.4 SiELC Primesep 100 HPLC (30 min) method for LC/MS purification of ‘polar’ proline analogues.	358
A.1 Schematic illustration of the IMPACT TM -CN System. ²⁶⁶	361

List of Schemes

1.1	The catalytic cycle of cytochrome P450 enzymes.	3
1.2	Outline catalytic cycle of flavin-dependent monooxygenases (FMOs).	4
1.3	Outline of the catalytic cycle of the 2-oxoglutarate/iron(II)-dependent dioxygenases (ODDs).	7
1.4	The reaction rates of ODDs can be improved, in some cases, by including: (a) the enzyme catalase, which removes oxidative hydrogen peroxide that may damage ODDs, and (b) ascorbic acid ('Vitamin C'), a known reducing reagent thought help maintain ferrous iron(II) levels.	8
1.5	The general reaction catalysed by aromatic amino acid hydroxylases.	10
1.6	The catalytic cycle of the aromatic amino acid hydroxylases (AAAHs).	10
1.7	The reaction catalysed by AsnO.	12
1.8	The reactions catalysed by AspH.	13
1.9	The reaction catalysed by IDO: IDO hydroxylates L-Ile to produce (2 <i>S</i> ,3 <i>R</i> ,4 <i>S</i>)-4-HIL.	15
1.10	The reactions catalysed by HilA and HilB; (a) HilA transforms L-Ile into 4'-hydroxy-L-isoleucine and (b) HilB transforms the product of this reaction into 4,4'-dihydroxy-L-isoleucine.	16
1.11	The reaction catalysed by LdoA: L-leucine is hydroxylated to produce (2 <i>S</i> ,4 <i>S</i>)-5-hydroxyl-L-leucine.	17
1.12	The reaction catalysed by SadA: <i>N</i> -succinyl L-leucine is hydroxylated to produce <i>N</i> -succinyl- <i>threo</i> -3-hydroxy-L-leucine.	18
1.13	The reactions catalysed by VioC and VioD: (a) VioC converts L-Arg to (3 <i>S</i>)-hydroxy-L-arginine and (b) VioD transforms (3 <i>S</i>)-OH-L-Arg into capreomycinidine using PLP.	20
1.14	The reaction catalysed by MppO: L-enduracididine (L-End) is hydroxylated to produce (3 <i>S</i>)-hydroxy-L-End.	20

LIST OF SCHEMES

1.15 The reaction catalysed by PP4H: L-proline in Gly-X-Pro triplets (X is other L-amino acids other L-glycine) is hydroxylated to produce 4Hyp residues (giving Gly-X-(4Hyp) units).	26
1.16 The reaction catalysed by PP3H: L-proline in Gly-Pro-4Hyp-Gly sequences is hydroxylated to produce a <i>trans</i> -3-hydroxy-L-prolyl residue. 28	
1.17 The reaction catalysed by OGFOD1: the <i>trans</i> -3-hydroxylation of Pro ⁶² of RPS23.	29
1.18 The reaction catalysed by <i>trans</i> P4H: free L-proline is hydroxylated to produce <i>trans</i> -4-hydroxy-L-proline with a retention of stereochemistry. 33	
1.19 The reaction catalysed by <i>trans</i> P4H proceeds with a retention of stereochemistry.	33
1.20 The reaction catalysed by L-proline <i>cis</i> -3-hydroxylase Type I (<i>cis</i> P3H Type I): <i>cis</i> P3H Type I converts L-Pro to <i>cis</i> -3-hydroxy-L-proline. . . .	34
1.21 The reaction catalysed by L-proline <i>cis</i> -4-hydroxylase (<i>cis</i> P4H): <i>cis</i> P4H converts L-proline to <i>cis</i> -4-hydroxy-L-proline.	39
1.22 Turnover of L-pipecolic acid by WT <i>cis</i> P4H and the regioselective production of <i>cis</i> -5-Hypip by a V97F/V95W/E114G <i>cis</i> P4H triple-variant. 40	
1.23 The reactions catalysed by GloF, a L-proline <i>trans</i> -3/4-hydroxylase: GloF also converts <i>trans</i> -4-methyl-L-proline into <i>trans</i> -3-hydroxy- <i>trans</i> -4-methyl-L-proline.	41
1.24 The reaction catalysed by ACC Synthase: SAM is converted to ACC and S-thiomethylmethionine.	43
1.25 The reaction catalysed by ACC Oxidase involves the oxidative, step-wise cleavage of the ACC ring ^{231,232} to form ethylene and cyanide; ferrous iron is a cofactor and ascorbate is cosubstrate.	43
1.26 The transformation of 2-keto-4-methylthiobutyric acid (KMBA) to ethylene by <i>Escherichia coli</i> , <i>Cryptococcus albidus</i> etc.; KMBA is produced from L-methionine in NADH:Fe(III)(EDTA) oxidoreductase-mediated, transamination reaction.	44
1.27 The overall reaction catalysed by EFE: O ₂ , 2-OG, and L-arginine are consumed to yield L- Δ^1 -pyrroline-5-carboxylate (L-P5C), guanidine, succinate, and CO ₂	44

1.28	The two pathways catalysed by EFE: in reaction 1 (major), two reaction cycles produce 6CO ₂ , 2H ₂ O, and 2 ethylenes from two 2-OG molecules; in reaction 2 (minor), one cycle consumes both 2-OG and L-arginine molecules to generate L-P5C, guanidine, succinate, water, and CO ₂	45
2.1	The conversion of lactose to allolactose in a transglycosylation reaction catalysed by β -galactosidase (coded by <i>lacZ</i> gene in <i>lac</i> operon).	62
2.2	The levansucrase-catalysed polymerisation of D-sucrose molecules to form levans.	74
2.3	An overview of the steps involved in subcloning the full-length <i>cisP3H</i> Type I construct into a pNIC-28-Bsa4 expression vector.	75
2.4	The strategy used for subcloning <i>cisP4H</i> and <i>transP4H</i> gene a Takara pCOLD I vector.	85
2.5	The strategy for subcloning the <i>cisP3H</i> Type I gene into a Takara pCOLD I vector.	104
2.6	The strategy for subcloning the <i>efe</i> gene into a Takara pCOLD I vector.	106
3.1	The reaction catalysed by <i>transP4H</i> : free L-proline is hydroxylated to produce <i>trans</i> -4-hydroxy-L-proline with a retention of stereochemistry.	121
3.2	The reaction catalysed by L-proline <i>cis</i> -3-hydroxylase Type I (<i>cisP3H</i> Type I): <i>cisP3H</i> Type I converts L-Pro to <i>cis</i> -3-hydroxy-L-proline.	122
3.3	The reaction catalysed L-proline <i>cis</i> -4-hydroxylase (<i>cisP4H</i>): <i>cisP4H</i> converts L-proline to <i>cis</i> -4-hydroxy-L-proline.	122
3.4	The reactions catalysed by GloF, a L-proline <i>trans</i> -3/4-hydroxylase: GloF also converts <i>trans</i> -4-methyl-L-proline into <i>trans</i> -3-hydroxy- <i>trans</i> -4-methyl-L-proline.	123
3.5	Radioactive assays using [¹⁴ C]-labelled L-Pro.	125
3.6	Radioactive assay using [¹⁴ C]-labelled 2-OG.	126
3.7	Amino acid derivatisation methods for HPLC detection and quantification.	128
3.8	Proline derivatisation methods for HPLC detection and quantification.	131
3.9	Proline Hydroxylase (<i>transP4H</i> , <i>cisP3H</i> Type I and II) proline-analogue substrate activity assays.	158
3.10	Hydroxylation of <i>trans</i> -3-methyl-L-proline (<i>trans</i> -3-Me-L-Pro) by: (a) <i>cisP3H</i> Type II and (b) <i>cisP4H</i> . ²⁹⁵	164

3.11 Proposed hydroxylation reactions catalysed by the proline hydroxylases using an <i>N</i> -methyl-L-proline (<i>N</i> -Me-L-Pro) substrate [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	173
3.12 Proposed hydroxylation reactions catalysed by the proline hydroxylases using an 2-methyl-L-proline (2-Me-L-Pro) substrate [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	184
3.13 Proline hydroxylase reactions using a <i>trans</i> -4-(2-propynyl)-L-proline [<i>trans</i> -4-(HC≡C-CH ₂)-L-Pro] substrate [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	190
3.14 Turnover of <i>cis</i> -4-fluoro-L-proline (<i>cis</i> -4-F-L-Pro) by <i>trans</i> P4H.	197
3.15 Proposed reactions catalysed by the proline hydroxylases using a <i>cis</i> -4-fluoro-L-proline (<i>cis</i> -4-F-L-Pro) substrate [(a) <i>cis</i> P3H Type I and (b) <i>trans</i> P4H].	202
3.16 <i>trans</i> -4-fluoro-L-proline (<i>trans</i> -4-F-L-Pro) is not a substrate of <i>trans</i> P4H.	205
3.17 Turnover of <i>trans</i> -4-fluoro-L-proline (<i>trans</i> -4-F-L-Pro) by <i>cis</i> P4H.	207
3.18 Proposed reactions catalysed by the proline hydroxylases using a <i>trans</i> -4-fluoro-L-proline (<i>trans</i> -4-F-L-Pro) substrate [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	210
3.19 <i>trans</i> P4H catalyses the <i>trans</i> -4-hydroxylation of <i>trans</i> -3-hydroxy-L-proline (<i>trans</i> -3-OH-L-Pro) to produce (3 <i>R</i> ,4 <i>S</i>)-3,4-dihydroxy-L-proline [(3 <i>R</i> ,4 <i>S</i>)-3,4-DiOH-L-Pro].	212
3.20 <i>cis</i> P3H Type I catalyses the <i>cis</i> -4-hydroxylation of <i>cis</i> -4-hydroxy-L-proline (<i>cis</i> -4-OH-L-Pro) to produce (3 <i>S</i> ,4 <i>R</i>)-3,4-dihydroxy-L-proline [(3 <i>S</i> ,4 <i>R</i>)-3,4-DiOH-L-Pro].	214
3.21 <i>cis</i> P3H Type I catalyses the <i>cis</i> -4-hydroxylation of <i>cis</i> -4-hydroxy- <i>N</i> -methyl-L-proline (<i>cis</i> -4-OH- <i>N</i> -Me-L-Pro) to produce (3 <i>S</i> ,4 <i>R</i>)-3,4-dihydroxy- <i>N</i> -methyl-L-proline [(3 <i>S</i> ,4 <i>R</i>)-3,4-DiOH- <i>N</i> -Me-L-Pro].	216
3.22 Proposed reactions catalysed by the proline hydroxylases using a <i>trans</i> -4-hydroxy-L-proline (<i>trans</i> -4-OH-L-Pro) substrate [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	218
3.23 Potential product rearrangement pathways when <i>cis</i> P3H Type I and <i>cis</i> P4H use a <i>trans</i> -4-hydroxy-L-proline (<i>trans</i> -4-OH-L-Pro).	220
3.24 Potential product formation pathways when <i>cis</i> P3H Type I and <i>cis</i> P4H use a <i>trans</i> -4-hydroxy-L-proline (<i>trans</i> -4-OH-L-Pro).	221
3.25 Carboxymethylprolines (CMPs) are involved in carbapenem antibiotic biosynthesis.	223

3.26	<i>trans</i> -5-(2-methoxy-2-oxoethyl)-L-proline ("Me- <i>t</i> CMP") is a substrate of <i>cis</i> P3H Type I, and is <i>cis</i> -3-hydroxylated to produce (3 <i>R</i> ,5 <i>S</i>)-3-hydroxy-5-(2-methoxy-2-oxoethyl)-L-proline ["(3 <i>R</i>)-3-OH-Me- <i>t</i> CMP"].	224
3.27	The <i>in vitro</i> biosynthesis of (5 <i>S</i>)-5-[(<i>S</i>)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline ["(6 <i>S</i>)-6-(isopenten-2-yl)-Me- <i>t</i> CMP"].	227
3.28	(5 <i>S</i>)-5-[(<i>S</i>)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline ["(6 <i>S</i>)-6-(isopenten-2-yl)-Me- <i>t</i> CMP"] is a substrate of <i>cis</i> P3H Type I, and is <i>cis</i> -3-hydroxylated to produce (3 <i>R</i> ,5 <i>S</i>)-3-hydroxy-5-[(<i>S</i>)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline ["(3 <i>R</i> ,6 <i>S</i>)-3-OH-6-(isopenten-2-yl)-Me- <i>t</i> CMP"].	228
3.29	Proposed reaction of (a) <i>cis</i> P3H Type I and (b) <i>cis</i> P4H using a L-azetidine-2-carboxylic acid (L-Aze).	232
3.30	Proposed reactions catalysed by the proline hydroxylases using an L-pipecolic acid (L-Pip) substrate [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	233
3.31	Proposed reactions catalysed by the proline hydroxylases using an <i>N</i> -methyl-L-pipecolic acid (<i>N</i> -Me-L-Pip) substrate [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	236
3.32	Proposed reactions catalysed by the proline hydroxylases using an L-azepane-2-carboxylic acid (L-Apn) substrate [(a) <i>cis</i> P3H Type I, (b) <i>cis</i> P4H, and (c) <i>trans</i> P4H].	238
4.1	Apple tissue metabolises L-methionine to ethylene: L-methionine-[¹⁴ C] was shown to be converted to ¹⁴ C-labelled ethylene. ²²⁰	272
4.2	Reactions used to determine the identity of compound 'X', an intermediate produced from L-methionine in the plant biosynthesis of ethylene.	274
4.3	Use of isotopic-labelling experiments with radioactive L-methionine to follow the pathway of ethylene biosynthesis in apple tissue; [¹⁴ C] and [³⁵ S] labels were used. ²²⁰	275
4.4	The overall reaction catalysed by ACC oxidase and the outline mechanism of ACC turnover by ACC oxidase.	276
4.5	ACC oxidase preferentially converts mixtures of (1 <i>R</i> ,2 <i>S</i>)- and (1 <i>S</i> ,2 <i>R</i>)-AEC to but-1-ene over converting mixtures (1 <i>R</i> ,2 <i>R</i>)- and (1 <i>S</i> ,2 <i>S</i>)-AEC to but-1-ene. ²²⁷⁻²³⁰	277

LIST OF SCHEMES

4.6	The transformation of 2-keto-4-methylthiobutyric acid (KMBA) to ethylene by <i>Escherichia coli</i> , <i>Cryptococcus albidus</i> , and <i>Chromobacterium violaceum</i> etc.	278
4.7	Chemical identification of ethylene produced by <i>P. digitatum</i>	280
4.8	The reactions catalysed by EFE.	281
4.9	Proposed reactions catalysed by EFE. ²⁵⁵	283
4.10	Proposed reaction pathways catalysed by EFE. ²⁵⁵	284
4.11	Preparation of L-P5C from a <i>t</i> -Bu- <i>N</i> -Boc-5-OH-L-prolinate precursor. .	294
4.12	Proposed EFE reaction pathways	323
4.13	Proposed binding order of substrates and cofactors in EFE bifurcating reactions	324

List of Tables

3.1	<i>S. griseoviridis</i> P8648 <i>trans</i> P4H-induced carboxylative turnover of 2-OG in the presence of L-proline. ^{191,256}	155
3.2	Proline hydroxylase activity data for proline analogues: HPLC assays were used to measure substrate turnover for compounds such as L-azetidine-2-carboxylic acid (L-Aze), L-proline (L-Pro), 3,4-dehydro-L-proline (3,4-Dehydro-L-Pro), and L-pipecolic acid (L-Pip).	157
3.3	<i>In vivo</i> hydroxylation of L-proline (L-Pro) in <i>E. coli</i> cultures expressing L-proline <i>cis</i> -3-, <i>cis</i> -4-, and <i>trans</i> -4-hydroxylases (<i>cis</i> P3H, <i>cis</i> P4H, and <i>trans</i> P4H). ¹⁹⁸	161
3.4	<i>In vitro</i> hydroxylation of L-pipecolic acid (L-Pip) by FPLC-purified L-proline <i>cis</i> -3-, <i>cis</i> -4-, and <i>trans</i> -4-hydroxylases (<i>cis</i> P3H, <i>cis</i> P4H, and <i>trans</i> P4H). ¹⁹⁸	162
3.5	<i>In vivo</i> hydroxylation of L-proline (L-Pro) in <i>E. coli</i> cultures expressing L-proline <i>trans</i> -4-hydroxylase (<i>trans</i> P4H). ¹⁹⁸	163
3.6	Proline hydroxylase substrate analogue screening: the effects of substrate D/L-stereochemistry on proline hydroxylase activity.	170
3.7	Proline hydroxylase substrate analogue screening: the effects of substrate amino- and carboxylic functionalisation on proline hydroxylase activity.	180
3.8	Proline hydroxylase substrate analogue screening: the effects of substrate hydroxyl functionalities on proline hydroxylase activity.	191
3.9	Proline hydroxylase substrate analogue screening: the effects of substrate halogenation on proline hydroxylase activity.	211
3.10	Proline hydroxylase substrate analogue screening: the effects of substrate hydroxyl functionality on proline hydroxylase activity.	222
3.11	Proline hydroxylase substrate analogue screening: proline hydroxylase activity with carboxymethylproline (CMP) derivatives.	229

LIST OF TABLES

3.12 Proline hydroxylase substrate analogue screening: the effects of substrate ring size on proline hydroxylase activity.	241
3.13 Proline hydroxylase turnover of bicyclic proline derivatives.	258
5.1 1 L 50×TAE Tank Running Buffer Formulation	329
5.2 1 L TAE Running Buffer Preparation.	329
5.3 100 mL Sample Loading Dye Formulation.	329
5.4 Agarose DNA Gel Formulation.	330
5.5 1 L LB Formulation.	331
5.6 1 L 2-YT Broth Formulation.	331
5.7 1 L SOB Formulation.	332
5.8 1 L TB Formulation.	332
5.9 1 L ZYM-5052 Auto-induction Medium Formulation.	333
5.10 1 L ZF Formulation.	333
5.11 1 L 50×M Solution Formulation.	334
5.12 1 L 50×5052 Formulation.	334
5.13 1 L MgSO ₄ Formulation.	334
5.14 Stock solutions for buffer preparation.	335
5.15 Tris-Glycine SDS-PAGE Separating Gel Formulations.	336
5.16 Tris-Glycine SDS-PAGE Stacking Gel Formulation.	337
5.17 1 L Coomassie Blue Stain Formulation.	337
5.18 1 L Destain Solution Formulation: 30%(v/v) MeOH, 10%(v/v) AcOH in dH ₂ O.	338
5.19 GE Healthcare HisTrap FF (Ni-Sepharose) Maintenance Buffers. . . .	340
5.20 DEAE Sepharose FF Anion-Exchange Chromatography Buffer Components: <i>cis</i> P3H Type I Purification Step 1 (and intermediate steps for experimental <i>cis</i> P4H/ <i>trans</i> P4H).	342
5.21 Proline Hydroxylase (<i>cis</i> P4H/ <i>trans</i> P4H) and EFE Purification Buffers for Immobilised Metal (Ni ²⁺) Affinity Chromatography (IMAC). . . .	343
5.22 Size-Exclusion Chromatography Buffer Components - Proline Hydroxylases	343
5.23 Size-Exclusion Chromatography Buffer Components - Ethylene-Forming Enzyme	343
5.24 MonoQ Ion-Exchange Chromatography Buffer Components - Proline Hydroxylases	343

5.25 Phenyl Sepharose Hydrophobic Interaction Chromatography Buffer Components - Proline Hydroxylases	343
5.26 LC/MS ES-Source Settings	345
5.27 Proline hydroxylase LC/MS assay conditions for L-Pro-analogue de- tection.	354
5.28 EFE LC/MS assay conditions for L-P5C detection.	354
5.29 SiELC Primesep 100 HPLC Method for LC/MS analysis of assay sam- ples.	355
5.30 EFE GC assay conditions for ethylene detection.	356
5.31 SiELC Primesep 100 HPLC (30 min) method for LC/MS purification of 'polar' proline analogues.	357
5.32 SiELC Primesep 100 HPLC (30 min) method for LC/MS purification of 'very polar' proline analogues.	358
5.33 Proline hydroxylase NMR end-point assay conditions for starting ma- terial and product detection.	359
5.34 Proline hydroxylase NMR timecourse assay conditions for starting material and product detection.	359
5.35 Proline hydroxylase NMR timecourse assay conditions for starting material and product detection.	359
5.36 Proline hydroxylase NMR timecourse assay conditions for starting material and product detection.	360

LIST OF TABLES

Chapter 1

Introduction

1.1 C—H Bond Activation

The insertion of an oxygen atom into an unactivated, aliphatic C—H bond using conventional synthetic chemistry is challenging from both thermodynamic and kinetic perspectives. This is in part because of the strength of the sp^3 C—H single bond enthalpy (roughly 412 kJ mol^{-1})¹ and its low bond polarity (‘non-polar’). Expensive and toxic metal catalysts are often required to perform these transformations, which normally occur with limited stereoselectivity.² Thus, synthetic methods for the synthesis of complex compounds with many stereocentres often employ multiple steps with consequently low net yields. These low efficiency methods are thus often not very economically viable for producing desirable chemicals and pharmaceuticals.³ In contrast, enzymatic reactions can often efficiently produce many oxidised naturally occurring compounds with high regio- and stereo-specificity in one-pot, reagent mixtures⁴ and are, therefore, useful ‘biosynthetic’ tools. In particular, cytochrome P450 monooxygenases (CYP450 enzymes) have been well-studied, and some manifest useful oxidative capacities and have a broad range of substrates, reflecting their multiple biological roles (Figure 1.1).^{5–7}

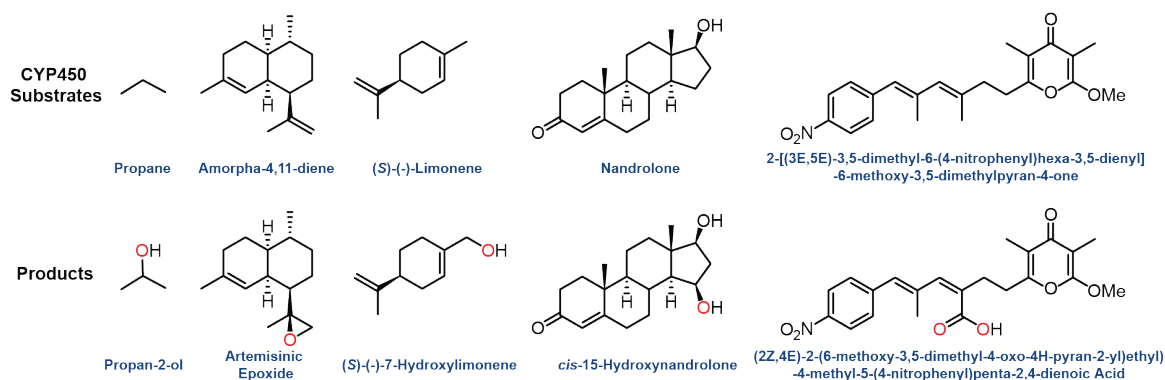


Figure 1.1: CYP450 enzymes have a wide variety of substrates with different structures; some of these products are displayed above, with the respective substrate names in blue.

1.1.1 Cytochrome P450 Monooxygenases

These cysteinato-haem CYP450 enzymes contain an iron-(III) protoporphyrin-IX prosthetic group covalently linked to the protein through the sulfur atom of a proximal cysteine ligand (Figure 1.2).⁸ The key reactive species in their catalytic cycle is a ferryl intermediate [a highly reactive iron(IV)-oxo species], which abstracts a hydrogen atom from the substrate, forming a carbon-based radical species and a hydroxyl radical that is datively bonded to the iron centre. The two radical species then combine to produce the hydroxylated product (Scheme 1.1).⁸⁻¹⁰

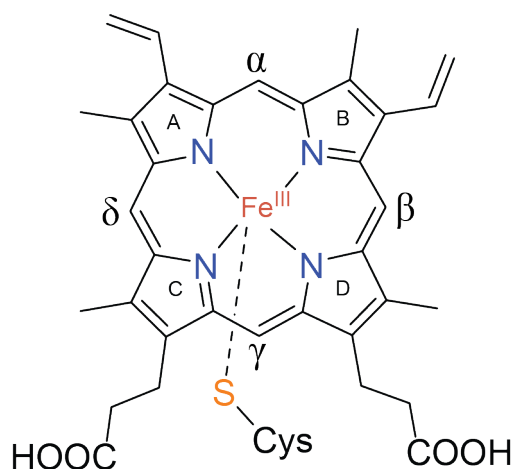
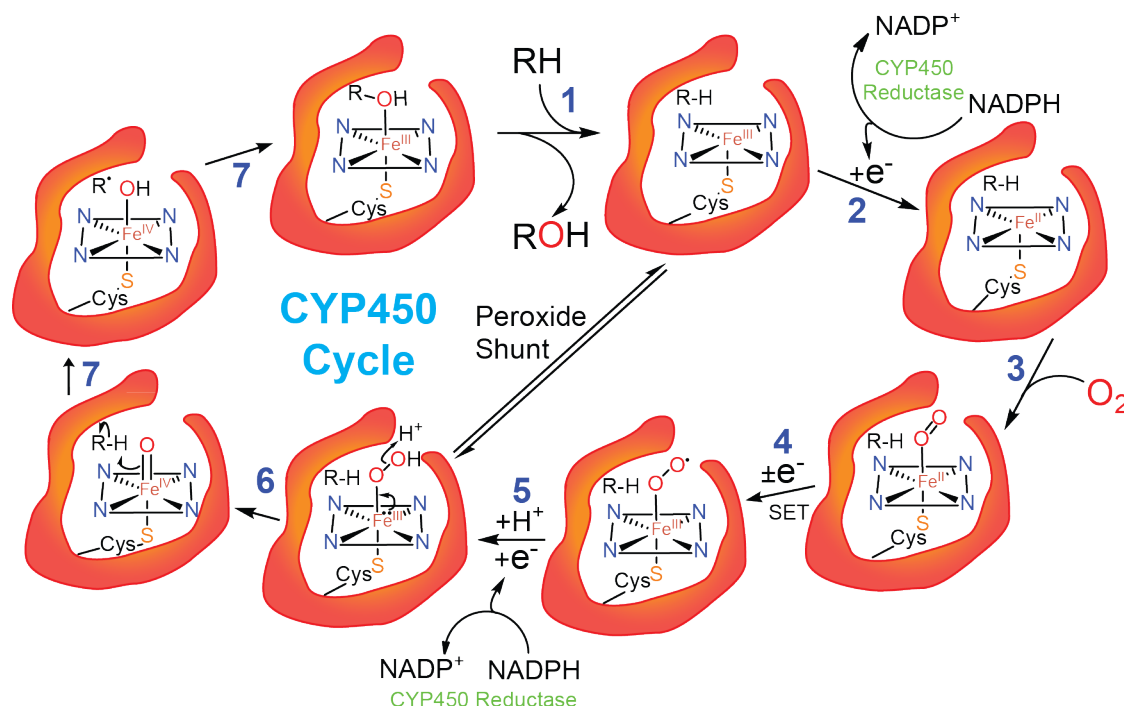


Figure 1.2: The cytochrome P450 (CYP450) iron(III) binding centre (haem site): it contains a protoporphyrin-IX prosthetic group with a proximal cysteine ligand.

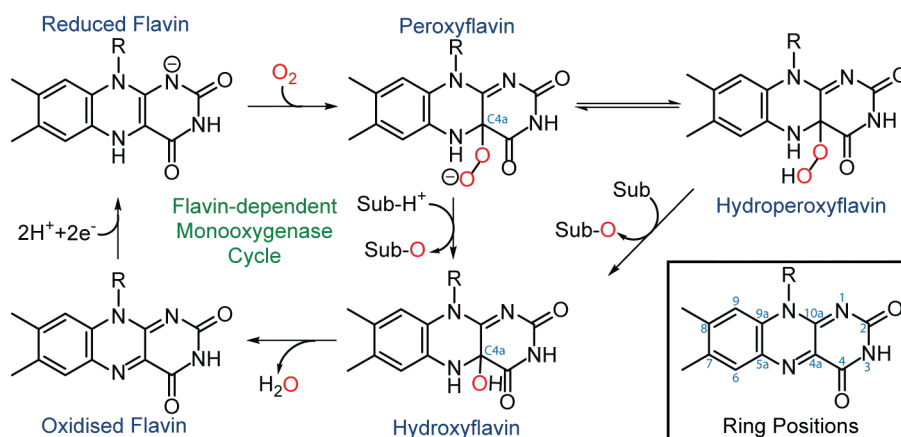


Scheme 1.1: The catalytic cycle of cytochrome P450 enzymes.

Molecular oxygen forms a dative bond to the iron centre and electron-transfer from the iron to oxygen form an iron-bound peroxide species, which dehydrates to form an Fe(IV)-oxo species. This ferryl intermediate then abstracts a hydrogen atom from the substrate to produce a free radical; the radical combines with the iron-bound hydroxyl radical to give an hydroxylated product.

1.1.2 Flavin-dependent Monooxygenases (FMOs)

Flavin-dependent monooxygenases are also versatile oxidising enzymes that catalyse multiple oxidation transformations, including hydroxylations, sulfoxidations, *N*-hydroxylations, epoxidations, halogenations, Baeyer-Villiger oxidations, and oxidative decarboxylations.¹¹ Oxygen activation in these redox enzymes typically involves the formation of a transiently-stable flavin C4a-oxygen adduct (peroxy species), which can react with nucleophilic or electrophilic substrate, depending of its protonation state, and break the O—O bond (Scheme 1.2).



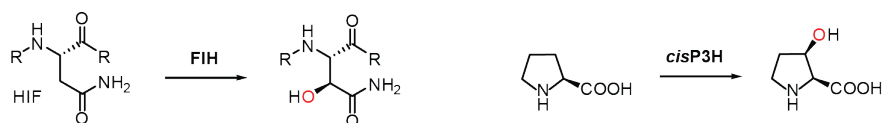
Scheme 1.2: The outline catalytic cycle of flavin-dependent monooxygenases (FMOs).

1.1.3 2-Oxoglutarate/Iron(II)-Dependent Dioxygenases

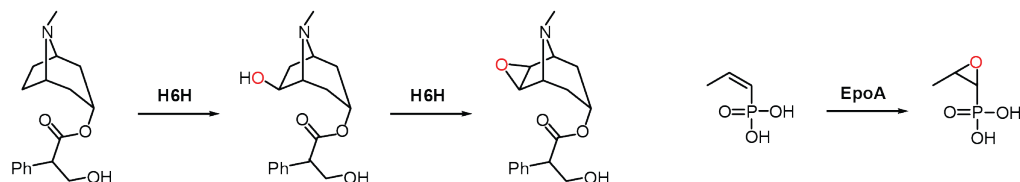
An important superfamily of oxidising enzymes with diverse and complex reactions are the 2-oxoglutarate/iron(II)-dependent dioxygenases (ODDs). The ODDs have been reported to catalyse numerous oxidative reactions which include hydroxylations, epoxidations, halogenations, desaturations, epimerizations, cyclisations, as well as ring formations, ring fragmentations, C-C bond cleavages, re-arrangements, demethylations, and demethylenations (Figure 1.3).^{12–17} The non-haem, ODD enzymes are ubiquitously distributed throughout nature, being found in bacteria, fungi, plants, and vertebrates.^{18,19}

By definition, the term ‘dioxygenase’ refers to enzymes which incorporate both (2) atoms of molecular oxygen into one or more substrates [‘di-’ being the prefix for ‘two’ or ‘twice’ from the Ancient Greek $\delta\iota\varsigma$ (dis)]. This definition can be slightly ambiguous, however, owing to the large number of dioxygenase types with different co-factor and co-substrate dependences; hence, the term ‘dioxygenase’ is sometimes used rather loosely. In the case of some ODDs, they do not catalyse reactions with complete incorporation of both oxygen atoms in their products.

Hydroxylation



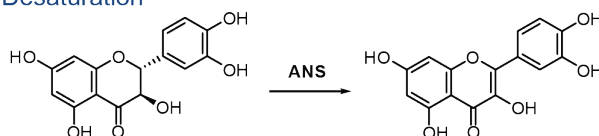
Epoxidation (+Hydroxylation)



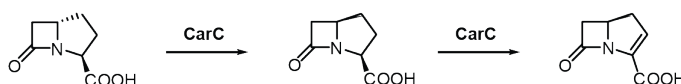
Halogenation



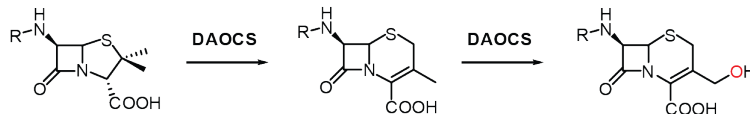
Desaturation



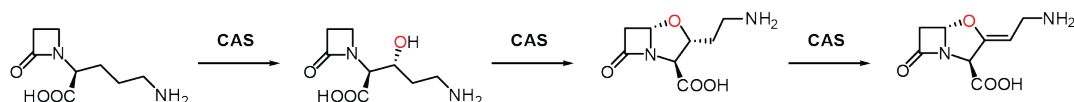
Epimerisation (+Desaturation)



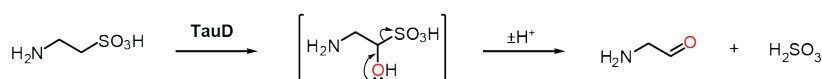
C-C Bond Cleavage, Ring Expansion/Rearrangement (+Desaturation & Hydroxylation)



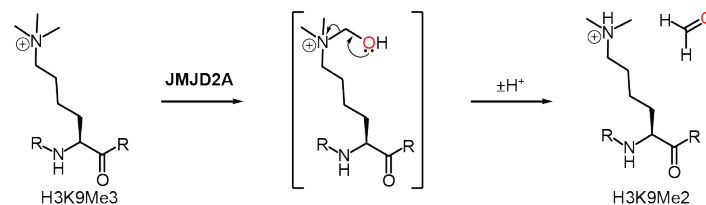
Ring Formation (+Hydroxylation +Desaturation)



Oxygenolytic Release (of sulfite)



Demethylation



Demethylation

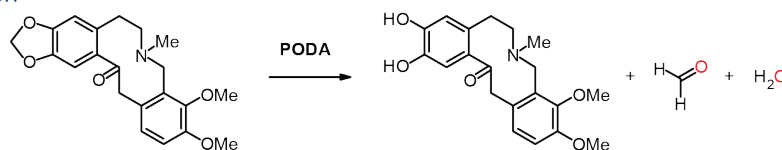


Figure 1.3: Examples of different reaction types catalysed by ODDs: illustrative enzymes are FIH,^{20,21} cisP3H,^{22,23} EpoA,²⁴ H6H,²⁵ SyrB2,²⁶ ANS,^{27,28} CarC,²⁹ DAOCS,^{30–32} CAS,^{33,34} TauD,³⁵ JMJD2A,^{36,37} PODA.¹⁶

2-Oxoglutarate/Fe(II)-dependent dioxygenases (ODDs), require ferrous iron and the citric acid cycle intermediate 2-oxoglutarate (2-OG, also known as ‘ α -ketoglutarate’, AKG) for catalysis^{13–15} (Figure 1.4 a). In the ODD active site, the ferrous iron co-factor binds a ‘catalytic triad’ of ligands comprising two histidine-residues and one aspartate- or glutamate- residue; the metal centre is site of the redox chemistry (Figure 1.4 b).

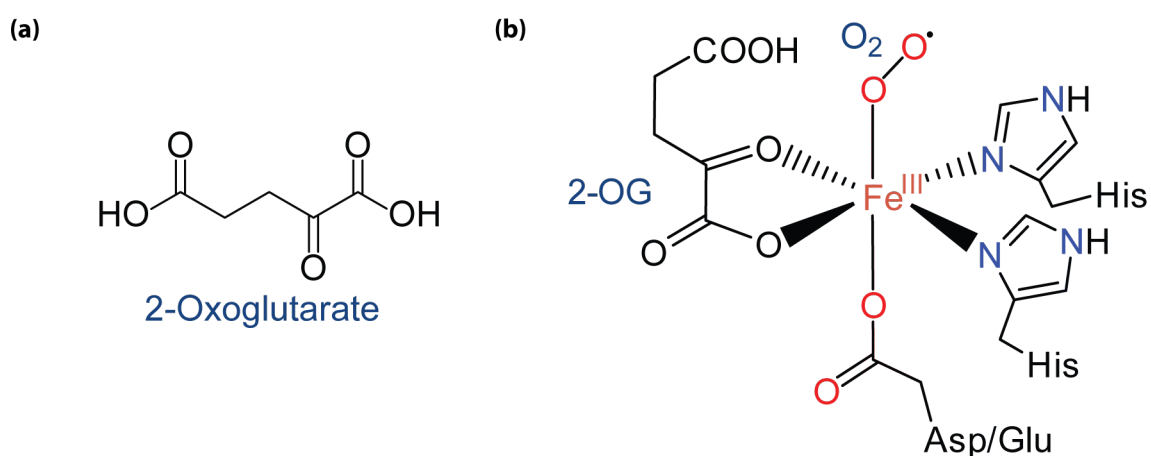
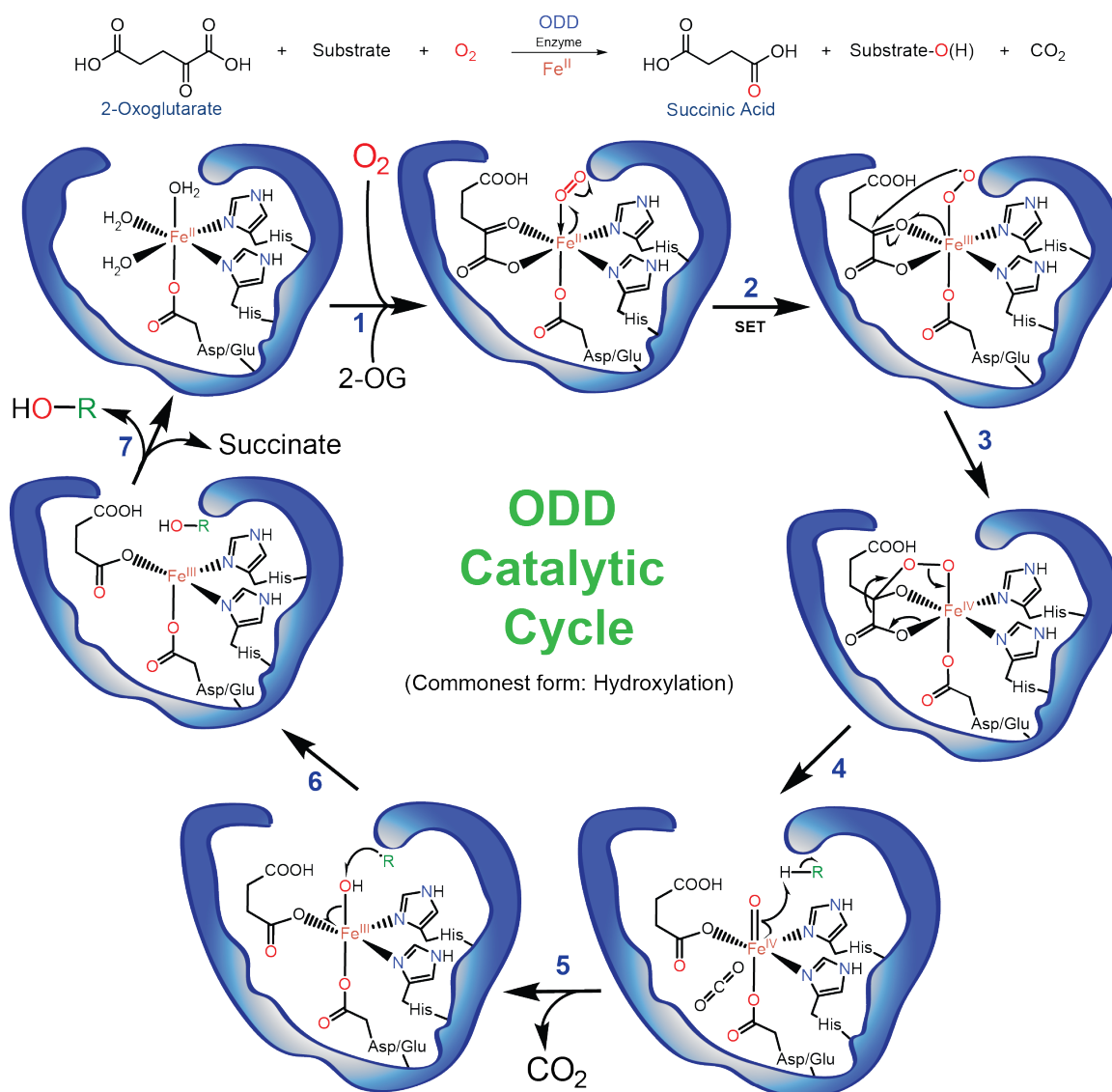


Figure 1.4: (a) Structure of 2-oxoglutarate and (b) a view of the ferrous iron binding site in ODDs; the metal binding site is formed by a triad (2 $\times$ His, 1 $\times$ Asp/Glu) of residues, which chelate to ferrous iron(II). A putative O₂-bound intermediate is shown.

Following binding of substrates and dioxygen, oxidative decarboxylation of 2-OG at the iron(II) centre leads to a highly-reactive ferryl intermediate [Fe(IV)-oxo species], which can perform a two-electron oxidation of the prime substrate *via* a radical mechanism (at least in some cases)(Scheme 1.3). Stopped-flow UV-visible spectroscopy techniques and Mössbauer Spectroscopy have been used to provide direct evidence for the high-valent iron intermediate in the case of Taurine/ α -Ketoglutarate Dioxygenase (TauD) from *E. coli*; this work provided the first formal characterisation of the high-spin³⁸ ODD ferryl intermediate. Succinate and carbon dioxide are produced as byproducts of the oxidative decarboxylation step of the catalytic cycle.^{13–15}



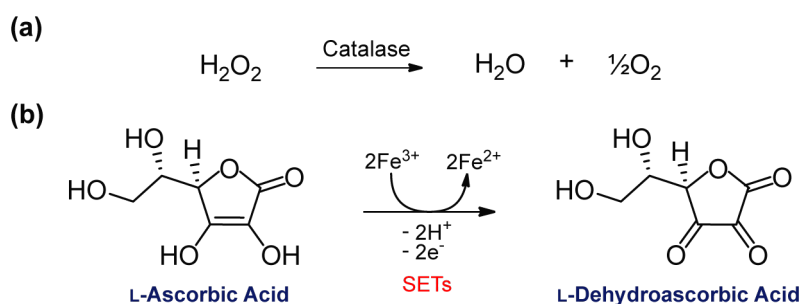
Scheme 1.3: Outline of the catalytic cycle of the 2-oxoglutarate/iron(II)-dependent dioxxygenases (ODDs); the most common reaction catalysed is hydroxylation.

Initially, redox reactions (steps 1-3) form a cyclic peroxy intermediate at the metallocentre by reaction of molecular oxygen and 2-OG;

(4) decarboxylation of this intermediate produces CO_2 and a ferryl intermediate.

(5) The $\text{Fe}(\text{IV})=\text{O}$ species then abstracts a H-atom from the substrate, producing a carbon-based free radical which can (6) recombine with the iron-bound $\text{HO}^\bullet$ radical to produce the hydroxylated product. In other instances, the radical may undergo a different type of terminating step (e.g. ring closure), but the majority of the cycle is conserved.

Various additives have been reported to enhance the activity of ODDs; in many cases catalytic turnover is increased by the presence of ascorbate ('Vitamin C') and, in fewer cases, by the presence of catalase. Catalase behaves as a protecting agent from damage by hydrogen peroxide (a powerful oxidising agent) by efficiently converting it into water and molecular oxygen³⁹ (Figure 1.4 a). Although not always essential for catalysis (if ever), ascorbate is a reducing agent that is thought to enzymatic turnover by completing substrate-uncoupled reactions (which may produce hydrogen peroxide);¹³ ascorbate may also be involved in maintaining ferrous iron (II) for catalytic turnover, though the full details of how this might be effected is not clear (Figure 1.4 b).^{18, 40}



Scheme 1.4: The reaction rates of ODDs can be improved, in some cases, by including: (a) the enzyme catalase, which removes oxidative hydrogen peroxide that may damage ODDs, and (b) ascorbic acid ('Vitamin C'), a known reducing reagent thought help maintain ferrous iron(II) levels.

1.2 L-Amino Acid Hydroxylases (AAHs)

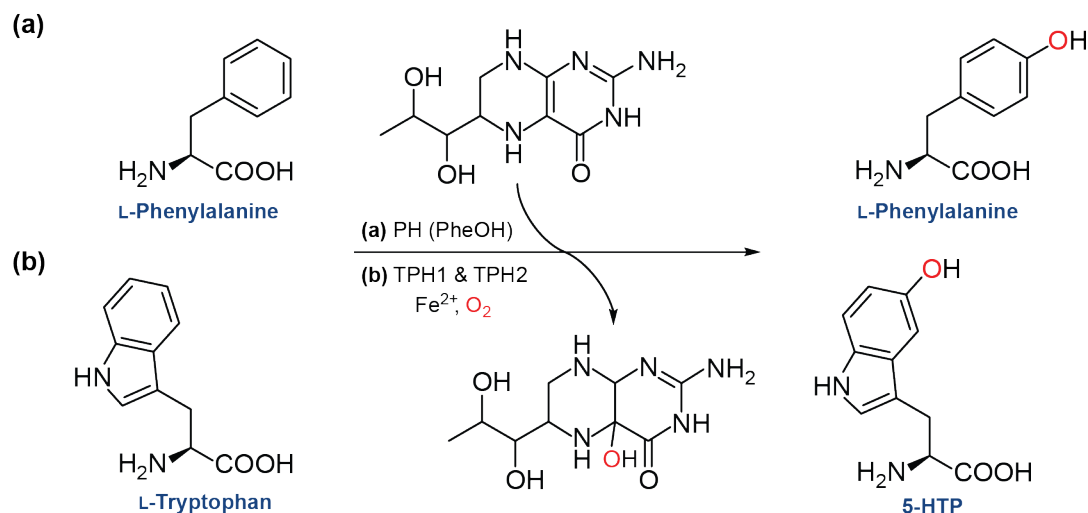
There are a few different types of amino acid hydroxylases found throughout nature, found in both mammalian and microbial organisms. Mechanisms involve the use of ferrous iron(II) and molecular oxygen to form a highly reactive Fe(IV)-Oxo species; some use tetrahydrobiopterin (BH₄) as a co-factor and others use 2-OG (as described above - see Section 1.1.3). In the remainder of this section, I summarise work reported on the different types of amino acid hydroxylase. Many of the func-

tional arguments in terms of the reactions catalysed were made during the course of my work.

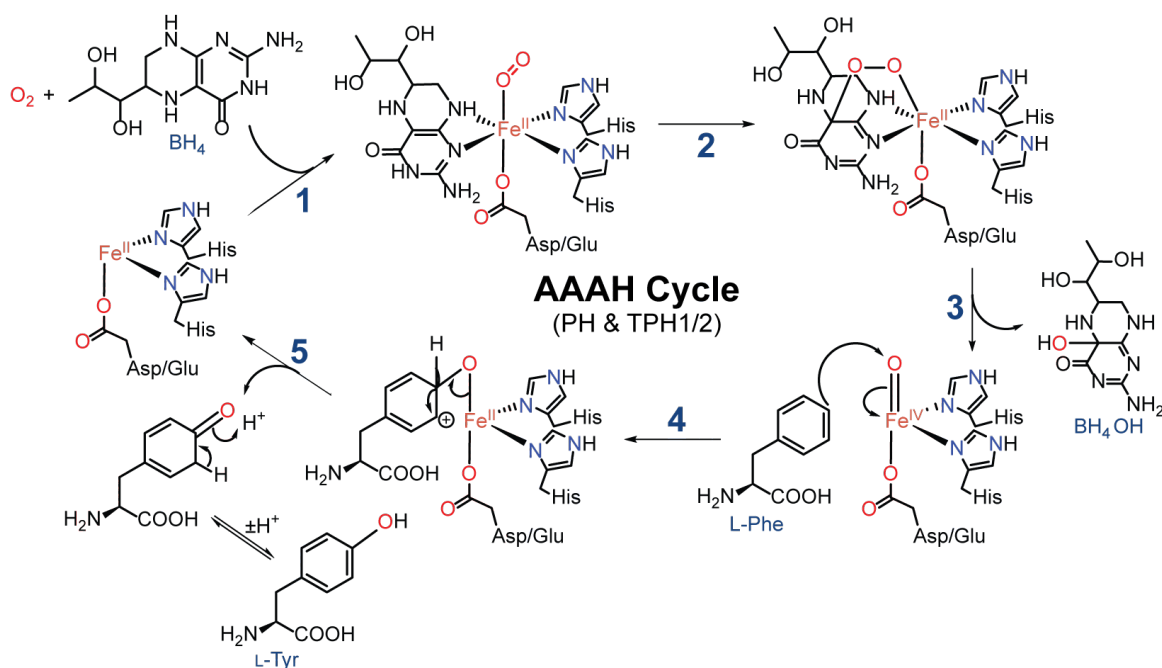
1.2.1 Aromatic Amino Acid Hydroxylases (AAAHs)

The aromatic amino acid hydroxylases (AAAHs) catalyse the hydroxylation of L-Phe, L-Tyr, and L-Trp.^{41,42} Dysfunction of animal AAAHs can result in serious disorders. Phenylalanine Hydroxylase (PAH) catalyses the hydroxylation of L-phenylalanine to L-tyrosine (Scheme 1.5 a); this is the rate-limiting step in phenylalanine catabolism. A deficiency of PAH, as may be caused by an autosomal recessive inborn error, results in a build-up of phenylalanine and is the basis for Phenylketonuria (PKU).⁴³ Tryptophan hydroxylases I and II (TPH1 and TPH2, respectively) catalyse the oxidation of L-tryptophan into 5-hydroxy-L-tryptophan (5-HTP)^{41,42} (Scheme 1.5 b). This is the rate-limiting step in the synthesis of serotonin (5-hydroxytryptamine, 5-HT),⁴⁴ which is an important monoamine neurotransmitter involved in numerous central nervous (CNS) activities; serotonin has several functions in peripheral tissues, including the maintenance of vascular tone and gut motility.⁴⁵ The mechanism of the AAAH monooxygenase enzymes involves the use of molecular oxygen, tetrahydrobiopterin (BH₄, a pteridine cofactor), and non-haem ferrous iron(II)⁴⁶ (Scheme 1.5). Molecular oxygen is heterolytically cleaved to form hydroxyhexahydrobiopterin and a ferryl [Fe(VI)=O] species which attacks the aromatic ring of the substrate, leading to the hydroxylated products (Scheme 1.6).

The AAH enzymes hydroxylate the aliphatic chains of L-amino acids by the outline mechanism in Section 1.1.3; they are largely found in microorganisms such as bacteria and fungi, often producing components of bioactive compounds as outlined below.



Scheme 1.5: The general reaction catalysed by aromatic amino acid hydroxylases.



Scheme 1.6: The catalytic cycle of the aromatic amino acid hydroxylases (AAAHs):

- (1) the oxygen and the BH_4 cofactor binds to the enzyme;
- (2) the BH_4 cofactor is oxidised forming an iron(II)-bound peroxy- BH_4 species which can (3) fragment, oxidising the iron(II) centre, to produce a highly-reactive Fe(IV)=O intermediate and BH_4OH ;
- (4) electrophilic attack of the aromatic ring system of amino acid leads to a carbocation intermediate which can (5) fragment to produce a hydroxylated product.

1.2.1.1 L-Asparagine Hydroxylase (AsnO)

Calcium-dependent antibiotic (CDA) is a non-ribosomally-produced, cyclic lipopeptide produced by *Streptomyces coelicolor* A3(2)⁴⁷ (Figure 1.5). CDA is an undecapeptide lactone that contains an *N*-terminal 2,3-epoxyhexanoyl side chain as well as several D-configured and non-proteinogenic amino acids.⁴⁸

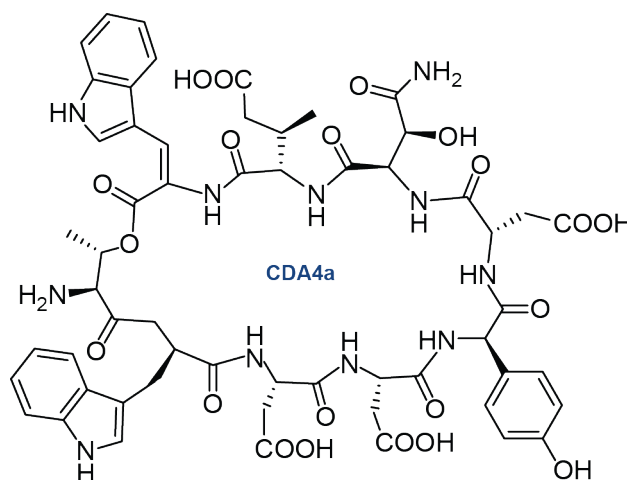
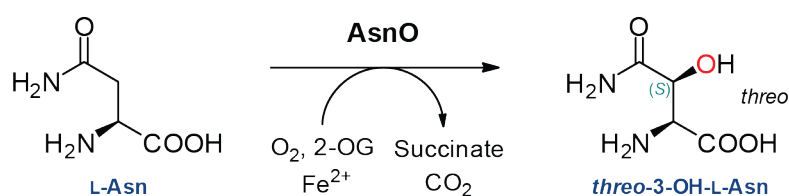


Figure 1.5: The structure of CDA4a, a calcium-dependent antibiotic.

The SCO3236 gene was found, by annotation of the CDA biosynthesis gene cluster,⁴⁹ to play a role in CDA biosynthesis and encode for a putative 2-OG/Fe(II)-dependent hydroxylase enzyme, AsnO.⁵⁰ Bioinformatic analysis⁵¹ revealed that AsnO shared high sequence identity (~33%) with clavamate synthase (CS/CAS - encoded for by *cas1* and *cas2*) from *Streptomyces clavuligerus*.^{33,34} Both AsnO and CAS are 2-OG/Fe(II)-dependent dioxygenases (ODDs).^{12,13} The hydroxylation catalysed by AsnO was thought to occur either post-synthetically on the asparaginyl residue of CDA⁴⁹ or on a precursor tethered to a single module [as is the case with SyrB2, a 2-OG/Fe(II)-dependent halogenase which chlorinates peptidyl carrier protein (PCP)-bound threonine²⁶]. Another possibility considered was hydroxylation of free amino acid to provide precursors for CDA biosynthesis, like that which occurs during the biosynthesis of etamycin,⁵² actinomycin,⁵³ and tuberactinomycin.^{54,55}

Experimental evidence, in the form of HPLC assay data and crystallographic studies, showed that AsnO acts on free L-asparagine alone; hydroxylation of L-asparagine found in CDA or tethered to PCP does not occur.⁵¹ A crystal structure of AsnO complexed with Fe(II), succinate, and hydroxy-L-asparagine (1.66 Å resolution - PDB 2OG7) is consistent with the stereoselective conversion of L-asparagine to give (2*S*,3*S*)-3-hydroxy-L-asparagine (Scheme 1.7); Fmoc-derivatised products also had the same HPLC retention times as derivatised (2*S*,3*S*)-3-hydroxy-L-asparagine standards, confirming the assigned selectivity.⁵¹



Scheme 1.7: The reaction catalysed by AsnO.

1.2.1.2 Aspartyl β -Hydroxylase (AspH)

The pseudomonal phytotoxin syringomycin E, and related nonribosomal peptides, contain a *threo*-3-hydroxy-L-aspartyl residue at the eighth position of the lipodepsipeptide backbone in a conserved, nonproteinogenic, tripeptide motif (Figure 1.6).^{56–60}

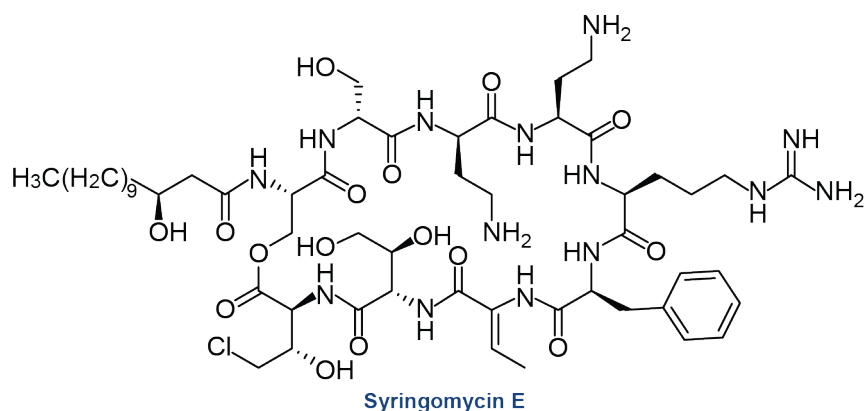
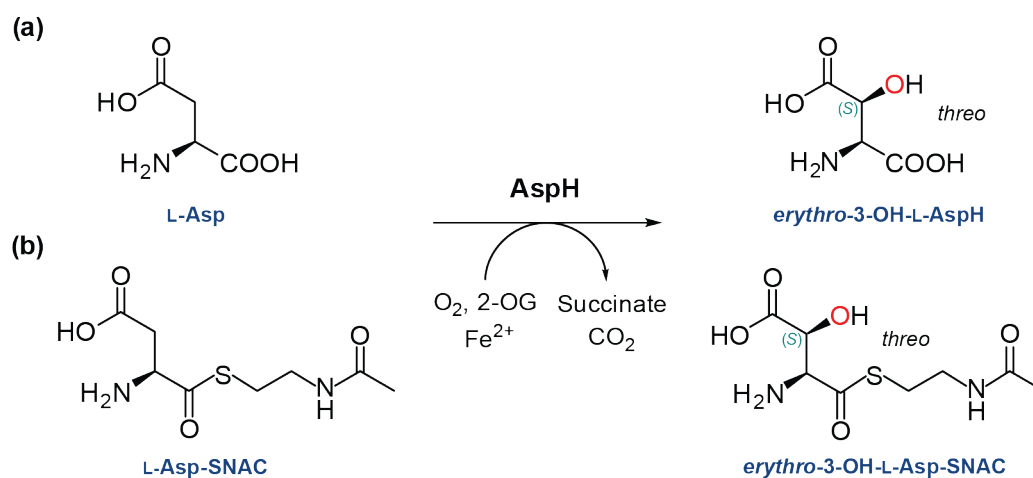


Figure 1.6: The structure of syringomycin E.

Bioinformatic analyses of the *P. syringae* pv *syringae* B728A genome revealed one ORF (primary locus Psyr_1584), *aspH*, that was predicted to be a non-haem, iron(II)-dependent aspartyl hydroxylase.⁶¹ Purified AspH, recombinantly expressed in *E. coli*, was then reported to catalyse the transformation of free L-aspartic acid into *threo*-3-hydroxy-L-aspartic acid (*threo*-3-OH-L-Asp) (Scheme 1.8 a); this result was particularly interesting as the expected stereochemistry was actually the ‘*erythro*’ form.⁶² If a thioester is formed with aspartate, forming *S*-L-aspartyl-*N*-acetyl-L-cysteine (L-Asp-SNAC), the L-aspartyl residue is still a substrate and the reaction occurs with the same stereospecificity (Scheme 1.8 b).⁶² A linear, *N*-butyryl nonapeptide mimic of a syringomycin lipononapeptide has also been shown to be hydroxylated by AspH, but the stereochemistry of this transformation is undetermined (Figure 1.7).



Scheme 1.8: The reactions catalysed by AspH.

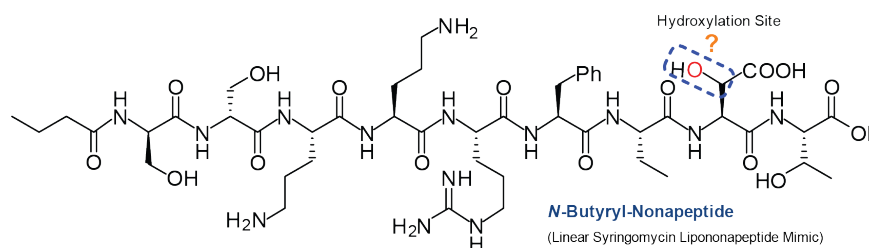
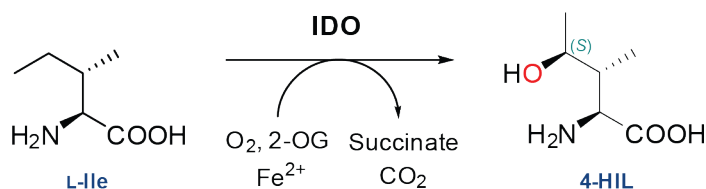


Figure 1.7: The structure of the *N*-butyryl nonapeptide mimic of a syringomycin lipononapeptide; it is a substrate of AspH.

1.2.1.3 Isoleucine Dioxygenase (IDO)

4-Hydroxyisoleucine (4-HIL) is a natural nonproteinogenic amino acid, which was initially extracted from fenugreek seeds (*Trigonella foenum-graecum*);⁶³ its absolute stereochemistry was later found to be (2*S*,3*R*,4*S*)-4-HIL.⁶⁴ 4-HIL is been reported to have insulintropic biological activity by increasing glucose-induced release of insulin.^{65–68} Unlike many pharmacological drugs used for the treatment of type II diabetes (*e.g.* sulfonyl ureas), 4-HIL-mediated insulin response is entirely dependent on glucose concentration, potentially allowing patients to avoid undesirable side-effects, such as hypoglycemia, when receiving therapy.^{69,70} 4-HIL is therefore a promising dietary supplement for the prevention and treatment of type II diabetes. There are several chemo-enzymatic methods for the synthesis of 4-HIL that have been published: an eight-step synthesis process (39% overall yield), involving the biotransformation of ethyl-2-methylacetoacetate to ethyl-(2*S*,3*S*)-2-methyl-3-hydroxybutanoate by *Geotrichum candidum*;⁷¹ a six-step chemo-enzymatic synthesis with complete stereochemical control, which involves the enzymatic resolution of an *N*-phenylacetyl lactone derivative, using immobilized penicillin acylase G (final step); and a 2-step, fully-enzymatic synthesis from acetaldehyde, α -ketobutyric acid, and L-glutamate using 4-hydroxy-3-methyl-2-keto-pentanoate aldolase (HPAL) and branched-chain amino acid aminotransferase (BCAT) (yields of 4-HIL are not sufficient for commercial production, and an abundant α -aminobutyric acid by-product is produced).^{72,73} The most effective method of producing 4-HIL was presumed to be *via* direct L-Ile hydroxylation. Further investigations showed that a component of the cell lysate from immature fenugreek seedlings exhibits L-Ile hydroxylation activity.⁷⁴ This hydroxylation reaction required 2-OG, ferrous iron(II), ascorbic acid, and oxygen, indicating that the enzyme carrying the activity is L-isoleucine dioxygenase (IDO). However, IDO from fenugreek has not been pu-

rified or subjected to detailed analysis. *Bacillus thuringiensis* strain 2e2 AKU 0251 was found to produce a small amount of 4-HIL in its culture broth in the exponential growth phase, though, with L-Ile hydroxylation activity (cell lysate) being maximal in the presence of 2-OG, ferrous iron(II), ascorbic acid, and oxygen, consistent with the presence of IDO.⁷⁵ IDO, from *B. thuringiensis* strain 2e2, was then purified and characterised, being shown to efficiently convert L-Ile to (2*S*,3*R*,4*S*)-4-HIL, a known anti-diabetic and anti-obesity drug,^{65–70} in a regioselective and stereoselective manner (Scheme 1.9).⁷⁵ IDO has now been thoroughly characterised⁷⁶ and is used in the biotechnological synthesis of 4-HIL.⁷⁷

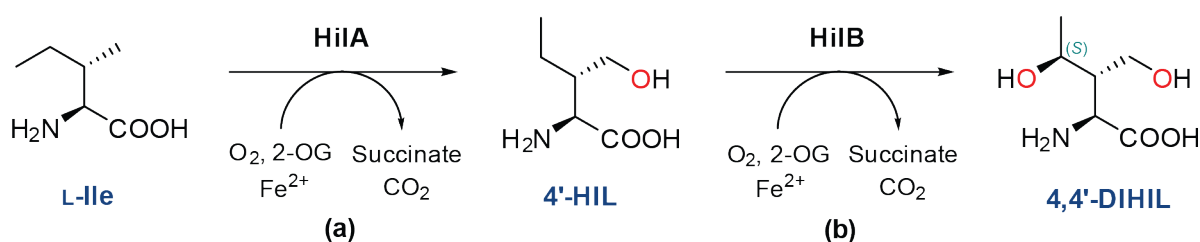


Scheme 1.9: The reaction catalysed by IDO: IDO hydroxylates L-Ile to produce (2*S*,3*R*,4*S*)-4-HIL.

1.2.1.4 L-Ile Dihydroxylation Cascade (HilA and HilB)

The identification of IDO^{75,77} (a member of the Pfam family, PF10014) from *Bacillus thuringiensis* triggered subsequent studies that revealed the existence of a novel sub-family of bacterial dioxygenases belonging to the PF10014 (BsmA) family.⁷⁸ They were reported to accept free L-amino acid substrates (L-leucine, L-isoleucine, and L-threonine) *in vivo*, producing 4-hydroxy-L-leucine, 4-hydroxy-L-isoleucine, and 4-hydroxy-L-threonine, respectively.⁷⁸ Detailed analysis of the genomes from several gram-negative plant pathogens belonging to the *Pantoea* and *Pseudomonas* genera uncovered a unique operon structure; it was designated the *Hil* operon.⁷⁹ This operon contained tandem *hilA* and *hilB* genes encoding putative dioxygenases belonging to the PF13640 and PF10014 Pfam families, respectively. HilB had been

previously found to catalyse C4-hydroxylation of L-isoleucine, but the K_m value for L-Ile was much higher than that measured for IDO,⁷⁶ suggesting that L-Ile might not be the natural substrate of HilB.⁷⁹ HilB is coexpressed with a putative dioxygenase, HilA, with these enzymes being translationally-coupled in some bacteria, suggesting that these enzymes form a complex that catalyses two consecutive reactions.^{79,80} It was thought that the natural substrate of HilB was actually the product of the reaction catalysed by HilA.^{79,80} Purification and characterisation of the HilA and HilB enzymes confirmed that these proteins catalyse consecutive reactions that transform L-Ile to 4'-hydroxy-L-isoleucine to 4,4'-dihydroxy-L-isoleucine (Scheme 1.10).



Scheme 1.10: The reactions catalysed by HilA and HilB; (a) HilA transforms L-Ile into 4'-hydroxy-L-isoleucine and (b) HilB transforms the product of this reaction into 4,4'-dihydroxy-L-isoleucine.

1.2.1.5 Leucine 5-Hydroxylase (LdoA)

(2*S*,4*S*)-5-HydroxyL-leucine is involved in the biosynthesis of (2*S*,4*S*)-4-methyl-L-proline, a known component of nostopeptolides of *Nostoc sp.* GSV224^{81,82} and *Nostoc sp.* XSPORK 13A (Figure 1.8), and nostocyclopeptides of *Nostoc sp.* ATCC 53789.⁸³

Bioinformatic analysis of the nostopeptolides biosynthetic cluster was used to predict that the product of the *orf5* gene, ORF5, would catalyse L-leucine hydroxylation.⁸² A homologue of ORF5 from *Nostoc punctiforme* PCC 73102, Npun F2185, was then reported to catalyse the transformation of L-leucine into (2S,4S)-5-hydroxy-

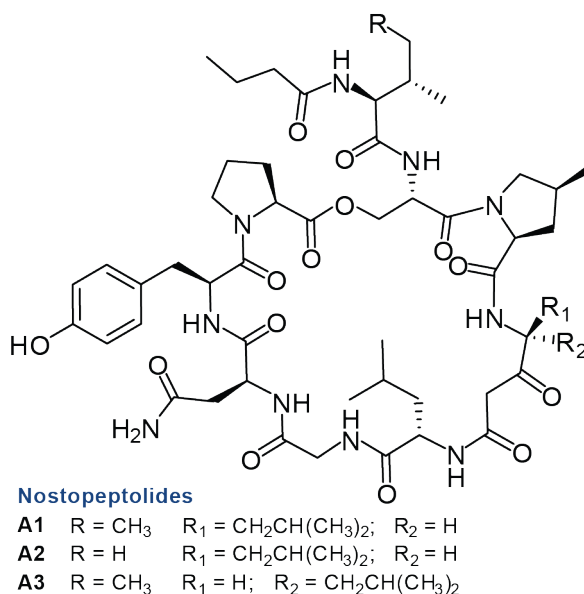
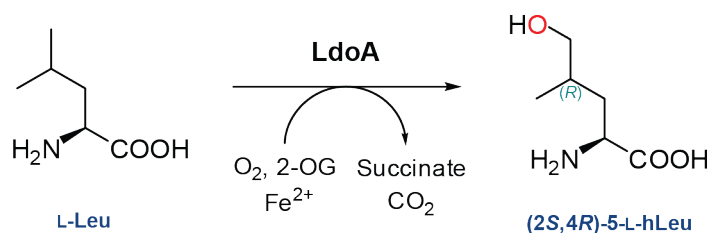


Figure 1.8: The structure of nostopeptolides A1, A2, and A3.

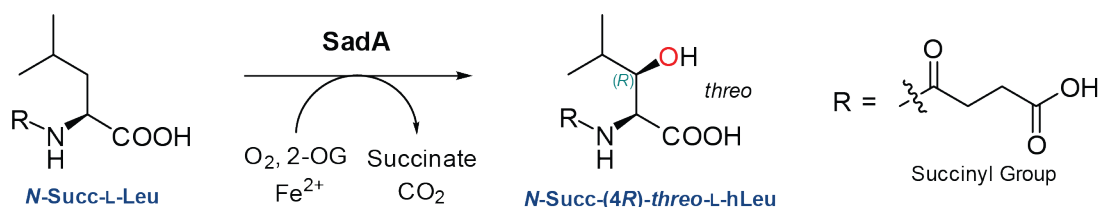
L-leucine in *Escherichia coli* cells heterologously expressing Npun F2185 as a biocatalyst⁸⁴ by an unknown mechanism. Further studies revealed that Npun F2185 is, in fact, a 2-OG/Fe(II)-dependent dioxygenase.⁸⁵ Purified recombinant Npun F2185 enzyme hydroxylated L-leucine and L-norleucine in the presence of 2-OG, ferrous iron, and L-ascorbate, producing (2*S*,4*S*)-5-hydroxyl-L-leucine and (2*S*)-5-hydroxy-L-norleucine, respectively (Scheme 1.11).⁸⁵ The enzyme was therefore named leucine 5-hydroxylase (LdoA).



Scheme 1.11: The reaction catalysed by LdoA: L-leucine is hydroxylated to produce (2*S*,4*S*)-5-hydroxyl-L-leucine.

1.2.1.6 *N*-Substituted L-Amino Acid Hydroxylase A

There are various methods for the chemical synthesis of *threo*-3-hydroxy-L-leucine (*threo*-3-OH-L-Leu),^{86,87} but they require complex reactions that provide low yields and high cost. Similarly, an early enzymatic method for preparing *threo*-3-OH-L-Leu using threonine aldolase resulted in very low product yields with low optical purity.⁸⁸ As a result, a search for a novel biocatalyst to produce *threo*-3-OH-L-Leu was carried out by screening putative ODDs (from the genome database) that possess homology with known amino acid hydroxylases; an activity assay was used to test enzymes for hydroxylation of L-leucine and its derivatives.⁸⁹ No enzyme in that particular study showed L-leucine hydroxylase activity, but one enzyme from *Burkholderia ambifaria* AMMD, now known as *N*-Substituted L-Amino Acid Hydroxylase A (SadA), reacted with several *N*-substituted L-leucines;⁸⁹ the primary substrate of SadA was reported to be *N*-succinyl L-leucine, which produces *N*-succinyl-*threo*-3-hydroxy-L-leucine with >99% diastereoselectivity (Scheme 1.12).⁸⁹



Scheme 1.12: The reaction catalysed by SadA: *N*-succinyl L-leucine is hydroxylated to produce *N*-succinyl-*threo*-3-hydroxy-L-leucine.

1.2.1.7 L-Arginine 3-Hydroxylase (VioC)

The tuberactinomycins (Tbms), such as viomycin (Figure 1.9), are non-ribosomal peptide antibiotics used for the treatment of mycobacterial infections;^{90–92} they act by binding RNA and, importantly, disturbing bacterial protein synthesis.^{93–95}

Initial analyses of the viomycin gene cluster from *Streptomyces vinaceus* ATCC

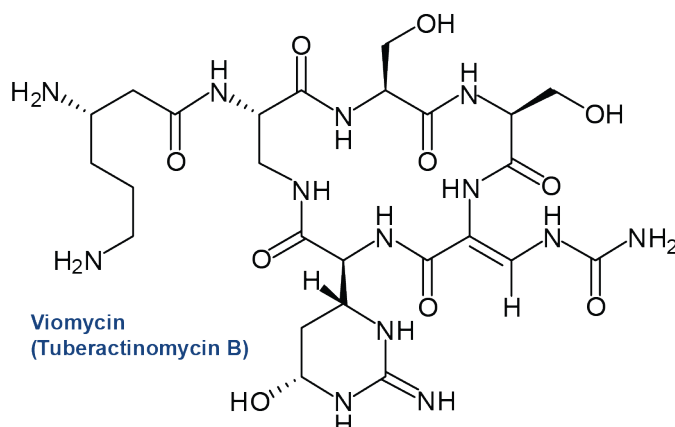
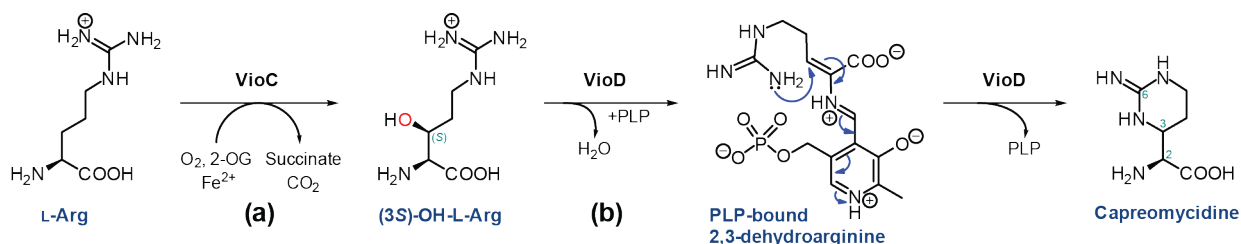


Figure 1.9: The structure of Viomycin B, a non-ribosomal peptide antibiotic.

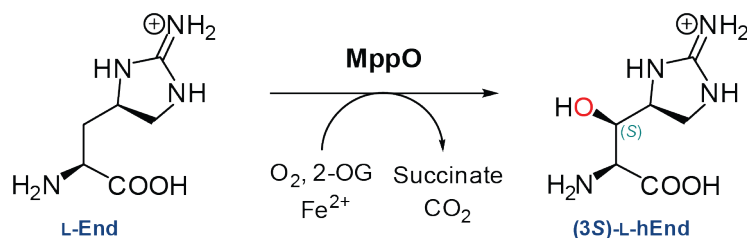
11861 identified an open reading frame, *vioC*, predicted to encode a non-haem iron/2-OG-dependent dioxygenase, VioC;⁵⁴ VioC was shown to be involved in viomycin biosynthesis with evidence collected in gene disruption experiments.⁵⁴ Sequence analysis then revealed close similarities with clavaminic synthase (CS), an enzyme known to perform β -hydroxylation, oxidative cyclisation, and dehydrogenation reactions;^{33,34} this prompted the hypothesis that VioC may catalyse one or more of the reactions that form capreomycin, a structural component of viomycin. HPLC assays revealed that VioC is in fact an ODD and catalyses the hydroxylation of L-arginine to produce (3*S*)-hydroxy-L-arginine (Scheme 1.13); HPLC chromatograms contained clear VioC-dependent product peaks which were positively identified as (3*S*)-OH-L-Arg using standards.^{55,96} The following step in viomycin biosynthesis involves a dehydrative elimination reaction, followed by a pyridoxal phosphate (PLP)-dependent cyclisation, both of which are catalysed by the enzyme VioD (*vioD* is the next gene in the MPP cluster).⁹⁶



Scheme 1.13: The reactions catalysed by VioC and VioD: (a) VioC converts L-Arg to (3S)-hydroxy-L-arginine and (b) VioD transforms (3S)-OH-L-Arg into capreomycinide using PLP.

1.2.1.8 Enduracididine β -Hydroxylase (MppO)

Bioinformatic analyses showed that the protein coded by the *mppO* gene in the MPP biosynthetic cluster,⁹⁷ MppO, has similarities with two ODDs, VioC (58% similarity, 42% identity)⁵⁵ (an L-arginine hydroxylase^{55,96}) and clavamate synthase (46% similarity, 31% identity)³³ (with known β -hydroxylase activity on L-arginine derivative, deoxyguanidine proclavaminic acid, as well its other functions). Based on these similarities, MppO was predicted to catalyse one or more of the reactions in the conversion of L-arginine to β -hydroxyenduracididine; this was confirmed when experimental evidence (*in vitro* and *in vivo* data) demonstrated that MppO acts primarily on L-enduracididine (L-End) to produce (3S)-hydroxy-L-End in a 2-OG/iron(II)-dependent manner (Scheme 1.14).⁹⁸ The product of this reaction is a structural component of the potent glycopeptide antibiotics, mannopeptimycins (MPPs), after which the MPP gene cluster is named (Figure 1.10).⁹⁹



Scheme 1.14: The reaction catalysed by MppO: L-enduracididine (L-End) is hydroxylated to produce (3S)-hydroxy-L-End.

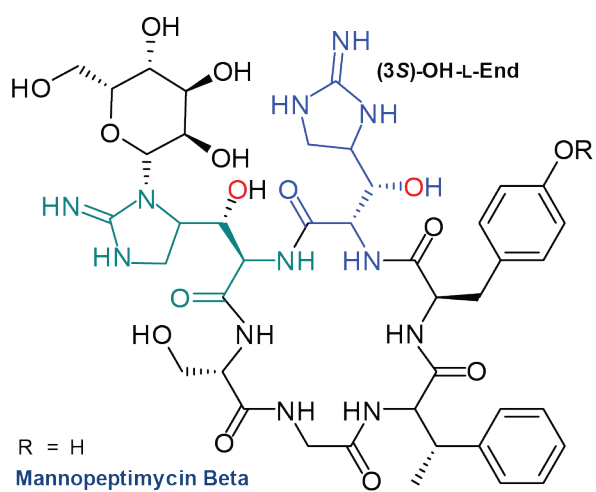


Figure 1.10: The structure of mannopectimycin β (MPP β).

1.2.2 Proline Hydroxylases

Proline hydroxylases are interesting amino acid hydroxylases, which catalyse hydroxylation of the L-proline aliphatic ring. The properties and functions of prolines (including hydroxyprolines and other derivatives) are outlined in the first part of this section; in the latter part of this section, a literature summary on L-proline hydroxylase enzymes is provided.

1.2.2.1 Proline and its Derivatives

Amino acids and their derivatives are valuable materials for organic synthesis¹⁰⁰ and biosynthesis,^{101–106} particularly L-proline, which has a relatively rigid ring conformations (Figure 1.11) which induce chirality; L-proline is therefore widely used for stereoselective synthesis as a catalyst and in stoichiometric quantities.^{107,108}

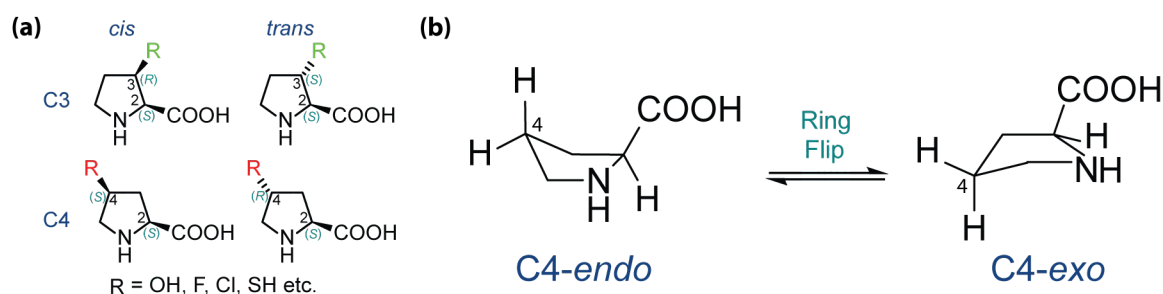


Figure 1.11: (a) Substitutions on the L-proline ring induce relatively fixed stereochemistry owing to the relatively rigid nature of the ring; (b) L-proline adopts 2 major, envelope ring conformations (C4-endo and C4-exo) which rapidly interconvert.

1.2.2.2 Hydroxyprolines in Nature

Hydroxyprolines are found throughout nature in both mammalian and microbial organisms; for example, three of the four diastereoisomers of the non-proteinogenic amino acid 4-hydroxyproline occur naturally.¹⁰⁹ The most abundant form of hydrox-

ypoline is *trans*-4-hydroxy-L-proline, a component of secondary metabolites such as echinocandins¹¹⁰ and actinomycins¹¹¹ (Figure 1.12). Echinocandins are antifungal drugs that inhibit the synthesis of glucan in fungal cell walls *via* noncompetitive inhibition of the enzyme 1,3- β -D-glucan synthase;^{112,113} actinomycins are a class of polypeptide antibiotics isolated from *Streptomyces* soil bacteria.¹¹⁴

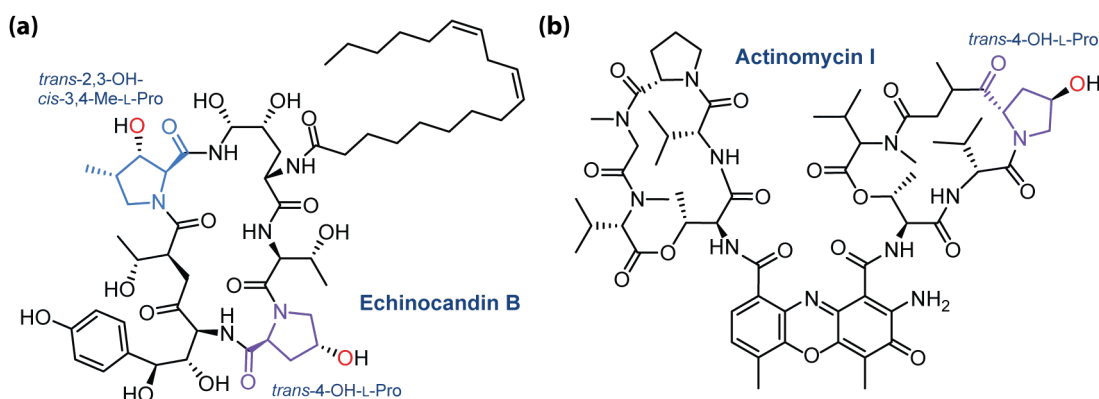


Figure 1.12: The structures of (a) echinocandin B and (b) actinomycin I, which are secondary metabolites, containing *trans*-4-hydroxy-L-proline. Echinocandin B also contains *trans*-2,3-hydroxy-*cis*-3,4-methyl-L-proline.

1.2.2.2.1 The Role of Prolyl Hydroxylation in the Hypoxic Response

In humans, *trans*-4-hydroxy-L-proline residues have been shown to play a key role in the hypoxic response. Hypoxia inducible factor (HIF)^{115–119} is a transcription factor, which regulates transcription of hundreds of genes encoding for proteins that help animals to cope with limiting oxygen conditions.^{120,121} Consequently, HIF is associated with many pathological and physiological conditions, including ischaemia, cardiovascular disorders, pulmonary hypertension, pregnancy disorders, and cancer.^{122,123} HIF is composed of 2 subunits, HIF- α and HIF- β ; both proteins contain a PAS domain and have a basic-helix-loop-helix architecture.¹²⁴ Under normal, ambient, oxygen partial pressures ('normoxia'), HIF- α forms a complex with the von-Hippel-Lindau tumour suppressor protein (pVHL)^{125,126} (which is bound to E3 ubiquitin ligase), resulting in ubiquitination and proteosomal degradation

of HIF- α (Figure 1.13).^{126–130} However, under ‘hypoxic’ conditions (low oxygen concentrations), HIF- α is stable to degradation, as pVHL does not bind it; HIF- α is thus free to interact with HIF- β , so inducing hypoxia responsive gene expression (Figure 1.13).¹²¹ It has been shown that *trans*-4-hydroxylation of 2 specific L-proline residues, Pro-402 and Pro-564, by HIF- α hydroxylases/prolyl hydroxylase domains (PHD1, PHD2, PHD3^{131,132} - particularly PHD2) is necessary for pVHL-mediated degradation;^{115–119} the 2 substrate motifs of HIF-1 α are known as NODD (residues 344-503) and CODD (residues 549-582) (*N*-terminal and *C*-terminal oxygen-dependent degradation domains, respectively).^{117,128,133,134} A crystal structure of the HIF- α -pVHL complex at Pro-564 (CODD sequence) indicates the structural basis for *trans*-4-hydroxy-L-proline recognition.^{135,136} In the hydrophobic core of pVHL, the structure showed binding pockets specific to *trans*-4-hydroxy-L-proline, which select for *trans*-4-hydroxy-L-proline over L-proline by forming hydrogen bonding interactions with the *trans*-4-hydroxyl group.^{135,136}

1.2.2.2.2 Collagen

In mammalian organisms, collagen fibres are a vital tensile element of skin, bones, tendons, cartilage, and teeth (Figure 1.14 c).¹³⁷ Consequently, collagen disfunctions are associated with diseases, including osteogenesis imperfecta, osteoporosis, osteoarthritis, arterial aneurysm, and Knobloch Syndrome.¹³⁸ It is well known that *trans*-4-hydroxy-L-proline (4Hyp) residues (along with some *trans*-3-hydroxy-L-proline-residues) play a prominent role in the structural integrity of triple-helical collagen.^{139–143} Hydroxyproline-residues are also present in over 10 other mammalian proteins containing collagenous domains.¹⁴⁰

The triple-helical structure (Figure 1.14) of collagen results, in part, from its very regular, repetitive amino acid sequence. Every third amino acid in the α -chain is L-glycine in a repeating ‘Gly-X-Y’ unit, where X and Y are often L-proline

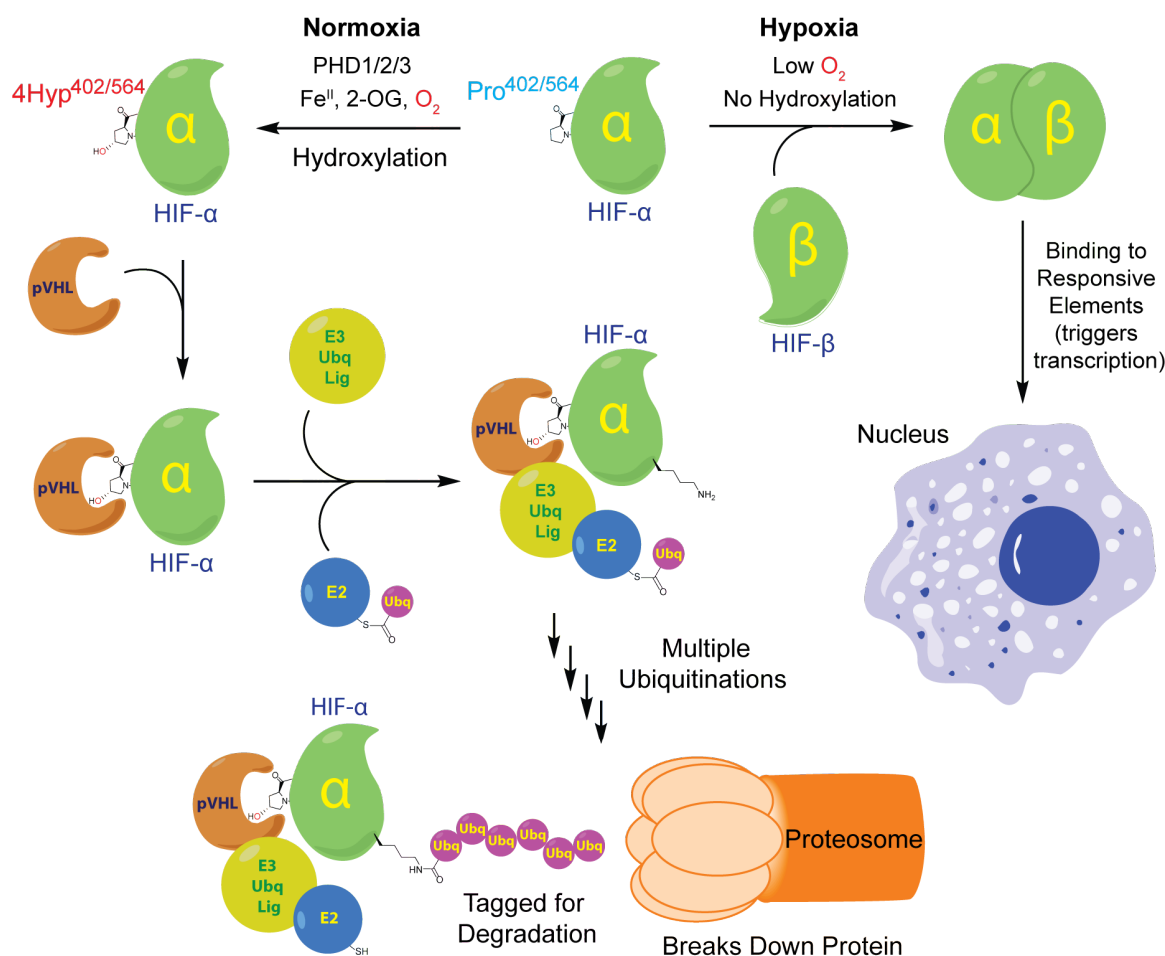


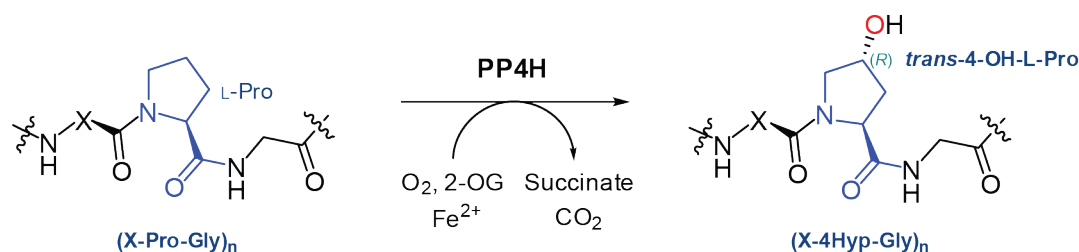
Figure 1.13: The chronic hypoxic response mechanism in animals: the HIF mechanism.

Under normoxic conditions (left), Pro-402 (NODD) or Pro-564 (CDD) of HIF-1 α (N-terminal and C-terminal oxygen-dependent degradation domains, respectively) is hydroxylated by a HIF-1 α prolyl 4-hydroxylase (PHD1/PHD2/PHD3) to produce a *trans*-4-hydroxy-L-Pro-402/564 (4Hyp) residue. The Von-Hippel-Lindau tumour suppressor protein (pVHL) then specifically targets the 4Hyp subunit, bringing with it a E3 ubiquitin ligase. The E3 ubiquitin ligase protein binds to a ubiquitin-conjugating (E2) enzyme, which catalyses the ubiquitination of hydroxylated HIF- α unit (usually at L-lysine residues). Repeated ubiquitination of HIF- α tags the protein for degradation by proteasomes.

Under hypoxic conditions (right), hydroxylation of Pro-402/564 is slowed, as the partial pressure of molecular oxygen is too low for optimal catalysis to occur. HIF- α is therefore not degraded and the transcription factor accumulates; HIF- α is now available to bind to HIF- β units and trigger the transcription of genes encoding for adaptive proteins.

and *trans*-4-hydroxy-L-proline (4Hyp), respectively (the ‘Gly-Pro-4Hyp’ triplet is the most common sequence repeat) (Figure 1.14 d and e).¹⁴⁴ The first high-resolution X-ray crystal structure¹⁴⁵ of collagen confirmed the triple-helix model previously proposed.^{146–148} It also indicated that collagen has a highly-ordered hydration network, involving hydrogen bonds with both peptide backbone carbonyl groups and hydroxyl groups from 4Hyp residues. It is thought this network stabilises the collagen triple helix, contributing to the rigidity of collagen fibres,¹⁴⁹ and the thermal stability of collagen has been reported to increase with an increase in the degree of L-proline hydroxylation.^{142, 143} It is also necessary for *trans*-4-hydroxy-L-proline to occupy the Y-position of collagen’s ‘Gly-X-Y’ repeating unit for the polypeptides to form the triple-helical structure (*n.b.* those sequences with 4Hyp in the X-position do not form a triple helix). Triple helix formation also does not occur with *cis*-4-hydroxy-L-proline (4S-Hyp) in either the X- or Y-positions.¹⁵⁰ A high resolution crystal structure of crystalline (Pro-Pro-Gly)₁₀ (*i.e.* X=Y=Pro) shows that L-proline residues in the X- and Y-positions adopts both the C4-*endo* and C4-*exo* conformations (see Section 1.2.2.1), respectively.^{151, 152}

The post-translational *trans*-4-hydroxylation of collagen L-proline residues (in the Y-position of the ‘Gly-X-Y’ triplet) is catalysed by procollagen prolyl 4-hydroxylase (PP4H), a well-studied ODD, prior to triple-helix formation (Scheme 1.15),^{155–157} vertebrate PP4Hs specifically recognise ‘X-Pro-Gly’ sequences as substrates.^{155–157}



Scheme 1.15: The reaction catalysed by PP4H: L-proline in Gly-X-Pro triplets (X is other L-amino acids other L-glycine) is hydroxylated to produce 4Hyp residues (giving Gly-X-(4Hyp) units).

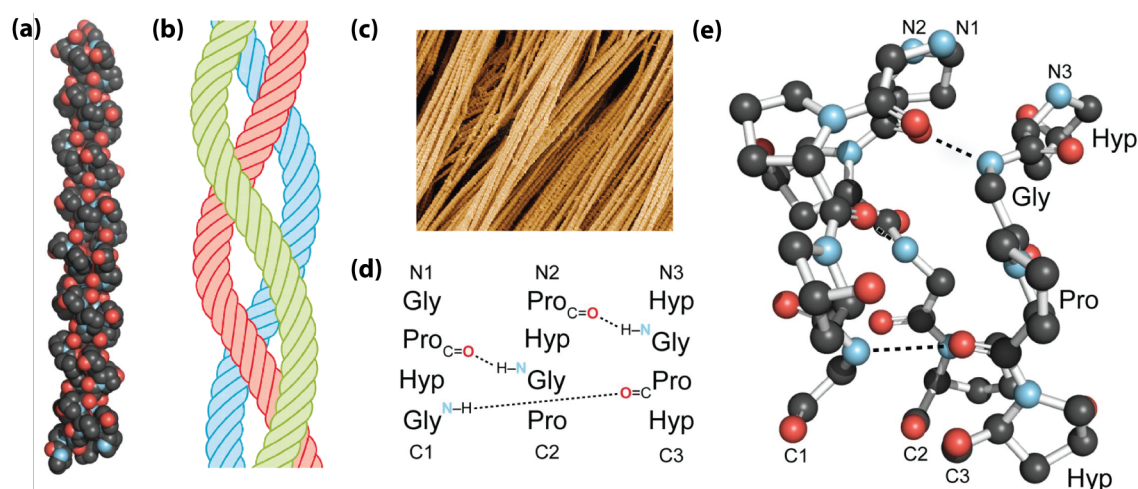


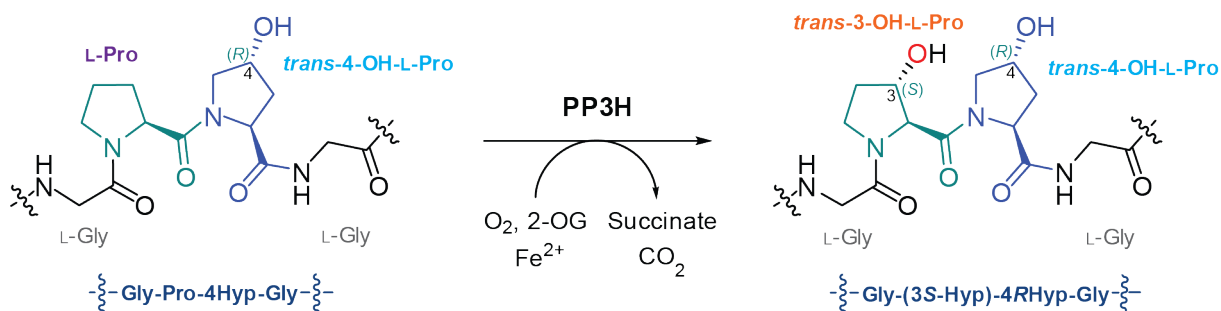
Figure 1.14: Molecular and structural aspects of collagen:

- (a) a space-filling model¹⁵³ from the first high-resolution crystal structure¹⁴⁵ of a collagen triple helix, formed from (ProHypGly)₄-(ProHypAla)-(ProHypGly)₅ [Protein Data Bank (PDB) entry 1CAG];
- (b) a schematic depicting the intertwining of strands in a collagen triple helix;
- (c) a scanning electron microscopy (SEM) image of collagen fibres ($\times 50,000$ magnification when printed 10cm wide);¹⁵⁴
- (d) the stagger of the three strands of a section of collagen (depicting hydrogen bond links);¹⁵³
- (e) a ball-and-stick image of a segment of collagen (PDB 1CAG),¹⁵³ showing interstrand hydrogen bonds.

Mammalian collagen PP4H activity is dependent of the peptide sequence (*i.e.* it varies depending on which amino acid is in the X-position)^{139, 140, 157} and has a maximal velocity when X is L-proline ('Pro-Pro-Gly' substrate).^{139, 157} The conformation of the collagen chains is important for turnover and triple-helical polypeptides are not accepted substrates of mammalian PP4Hs; triple-helix fibres require denaturation before they can be hydroxylated¹⁵⁸ and the β -turn motif of the 'Pro-Gly' fragment is also required.¹⁵⁹ The collagen polypeptide chain length was also shown to exhibit a significant effect on the substrate K_m for mammalian PP4H, decreasing with increasing chain length, suggesting a processive mechanism (*i.e.* moving along the chain and hydroxylating sites sequentially).^{139, 158, 160}

Despite the fact that *trans*-3-hydroxy-L-proline (3R-Hyp) is present in numerous plant proteins of *Fasciola hepatica*,^{161, 162} 3R-Hyp is far less abundant than 4Hyp in plant and animal collagens.^{139, 140} The ODD, procollagen prolyl 3-hydroxylase

(PP3H), catalyses the transformation of L-proline residues in collagen into *trans*-3-hydroxy-L-prolyl residues;¹⁶³ it was quickly found to specifically recognise Gly-Pro-4Hyp-Gly sequences as hydroxylation substrates (Scheme 1.16).¹⁶⁴ The requirement for the presence of *trans*-4-hydroxy-L-proline is strict and Gly-Pro-Pro-Gly sequences are not hydroxylated.¹⁶⁴



Scheme 1.16: The reaction catalysed by PP3H: L-proline in Gly-Pro-4Hyp-Gly sequences is hydroxylated to produce a *trans*-3-hydroxy-L-prolyl residue.

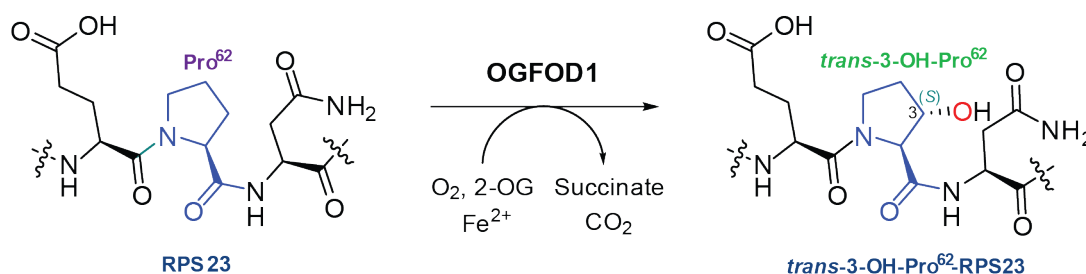
As in the case of collagen PP4H, PP3H activity is influenced by the substrate chain length, with longer substrates being better substrates.¹⁶³ The collagen triple-helix is not accepted as a PP3H substrate and must be denatured to separate the strands for turnover.¹⁶³

1.2.2.2.3 Ribosome Hydroxylation

2-OG- and Fe(II)-Dependent Oxygenase Domain-Containing Protein 1 (OGFOD1) was predicted to be a conserved 2-OG oxygenase with a catalytic domain related to the HIF prolyl hydroxylases (PHDs) (see Section 1.2.2.2.1 - page 23).¹⁶⁵ Yeast homologs of OGFOD1 are implicated in a range of cellular functions including the regulation of sterol response genes ('Ofd1' from *Schizosaccharomyces pombe*)¹⁶⁶ and translation termination/mRNA polyadenylation ('Tpa1p' from *Saccharomyces cerevisiae*).¹⁶⁷ Crystal structures of Tpa1p revealed that the protein contains a double-stranded β -helix (DSBH) fold domain, a motif characteristic of many 2-OG oxygenases, with typical Fe(II) and 2-OG binding residues;^{168,169} analysis of the

active site implies some distant similarity with the mammalian HIF prolyl hydroxylases (PHDs).^{168,169} Substitution of the predicted active site residues of Ofd1 and Tpa1p quells their corresponding functions,^{166,168} suggesting their functions are dependent on the oxidation of some form of regulatory substrate. The cellular roles of mammalian OGFOD1 are far from understood, however, human OGFOD1 was recently shown to be present in stress granules upon arsenite-induced stress, associating with haem-regulated kinase and eIF2 α to regulate eIF2 α phosphorylation.¹⁷⁰ Cells deficient in OGFOD1 have also been shown to possess a survival advantage under ischemic stress.¹⁷¹

Cellular and biochemical assays recently confirmed that OGFOD1 is, indeed, an ODD which modifies ribosomal prolyl residues.¹⁶⁵ OGFOD1 catalyses the *trans*-3-hydroxylation (Scheme 1.17) of Pro-62 of small ribosomal subunit protein, RPS23, a conserved prolyl residue in the ribosome-decoding center;¹⁶⁵ suppression or deletion of OGFOD1 is associated with the activation of translational stress pathways.



Scheme 1.17: The reaction catalysed by OGFOD1: the *trans*-3-hydroxylation of Pro⁶² of RPS23.

1.2.2.2.4 Cyclic Peptide Antibiotics

Etamycin

In 1954, a peptide antibiotic, etamycin, was isolated from culture broths of a *Streptomyces* strain.¹⁷² Simultaneously, another group reported an antibiotic termed Viridogrisein, obtained from *Streptomyces griseus*.¹⁷³ Analysis showed that etamycin and viridogrisein were in fact the same compound.^{174,175} Etamycin is particularly effective against gram-positive organisms and *Mycobacterium tuberculosis*, as well as causing a reversible leucopenia in dogs.^{172,176,177} Using degradation experiments, the structure of etamycin was shown to comprise a macrocyclic peptide lactone containing *cis*-4-hydroxy-D-proline (4S-D-Hyp) (Figure 1.15).^{175,178} Although etamycin is a potent antibiotic, its low water solubility limited its potential for medicinal applications.¹⁷²

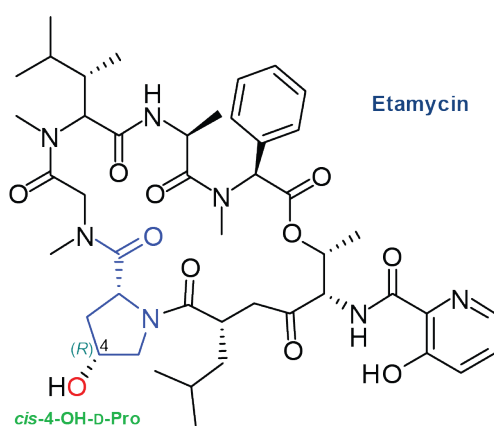


Figure 1.15: The structure of Etamycin.

Feeding studies indicated that each of the D-amino acid residues of etamycin were biosynthetically derived from their corresponding L-epimer;¹⁷⁹ it is thought that this epimerisation involves a dehydrogenation-hydrogenation reaction sequence.^{180,181} Both L-proline and *trans*-4-hydroxy-L-proline were shown to be precursors of *cis*-4-hydroxy-D-proline in etamycin biosynthesis.^{180,181} By 1979, evidence suggested that L-proline is enzymatically converted to *trans*-4-hydroxy-L-

proline prior to being incorporated into the peptide chain and the epimerisation reaction⁵² (a contrast to collagen biosynthesis which involves *trans*-4-hydroxylation of a peptidic substrate - see above).^{139,140} The *in vivo* reaction was found to be inhibited by Zn^{2+} and α,α' -dipyridyl¹⁸² and require ferrous iron.¹⁸² Similarly, *in vitro* reactions, using *Streptomyces griseoviridis* P8648 cell lysates, demonstrated a requirement for ferrous iron,⁵³ 2-OG, and L-ascorbate (for improved turnover),⁵³ implying the transformation is catalysed by an ODD (see Section 1.2.2.3).

Quite recently, Etamycin has been shown to exhibit antibiotic activity against methicillin-resistant *Staphylococcus aureus* (MRSA),¹⁸³ strains known to be highly virulent and easily transmitted in both community-associated and hospital-associated settings (known as CA-MRSA and HA-MRSA, respectively).^{184,185} Etamycin demonstrated potent activity against HA- and CA-MRSA in microbroth dilution assays, with minimum inhibitory concentrations (MIC) as low as $1\text{-}2\text{ mg l}^{-1}$ against HA- and CA-MRSA strains;¹⁸³ it was also reported to be active against other Gram-positive and several Gram-negative pathogens, as well as being non-cytotoxic at concentrations more than 20-fold above the minimum inhibitory concentration (MIC).

Telomycin

Telomycin, a cyclic peptide antibiotic, was first isolated from the culture broth of an unidentified strain of *Streptomyces* bacteria.¹⁸⁶ Telomycin is a colourless solid (amorphous or crystalline) which is sparingly soluble in water and relatively stable at room temperature over a wide range of pH values.¹⁸⁶ It is potently active against Gram-positive organisms, including strains of *Micrococcus pyogenes* var. *aureus* resistant to other antibiotics, but has little-to-no effect against Gram-negative bacteria.¹⁸⁷ The structure of telomycin was elucidated using data collected from degradation studies, potentiometric pK measurements, and infra-red and ultra-violet spectroscopy (Figure 1.16);¹⁸⁸ the antibiotics contains both *trans*-3-hydroxy-L-proline (3S-Hyp) and *cis*-3-hydroxy-L-proline (3R-Hyp).

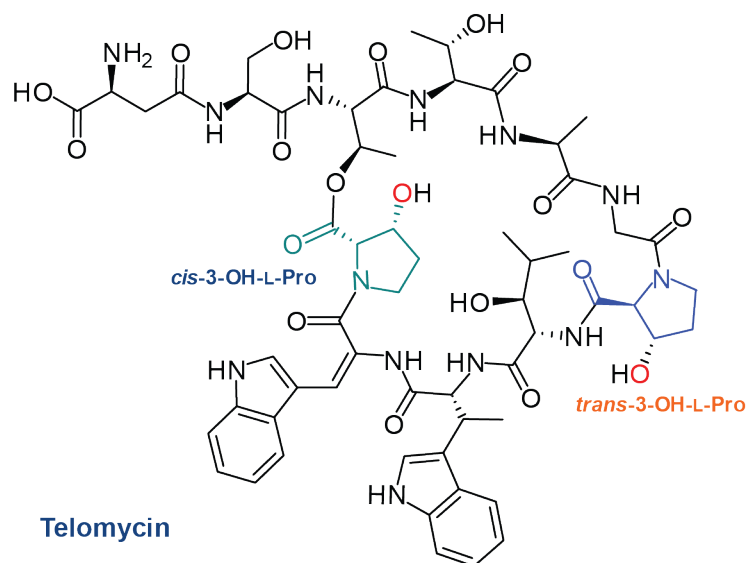


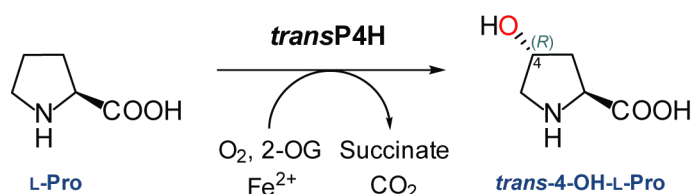
Figure 1.16: The structure of Telomycin.

The proton NMR spectra of telomycin,¹⁸⁹ together with conformational ¹H-NMR studies on the amide backbone protons,¹⁹⁰ revealed that the amide protons of L-threonine, L-hydroxyleucine, and β -methyltryptophan are not solvent accessible; this suggested that hydroxy-L-proline residues may be involved in β -like turns that enable internal hydrogen bonds between these protons.¹⁹⁰

1.2.2.3 L-Proline *trans*-4-Hydroxylase (*trans*P4H)

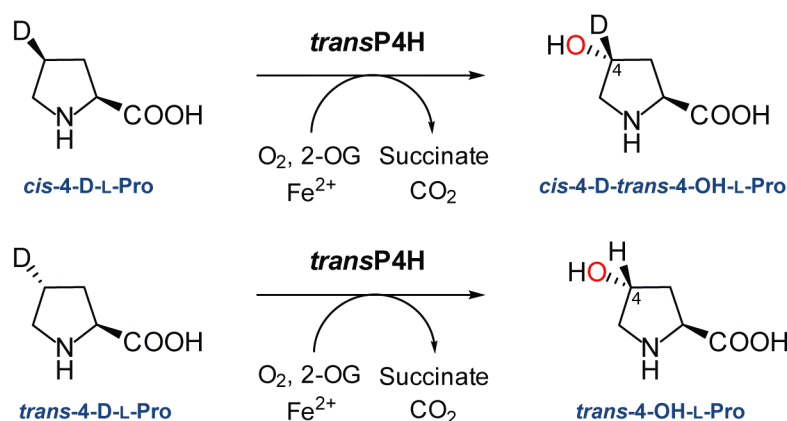
Mammalian systems are known to utilise post-translational hydroxylation of L-prolyl residues like, for example, the *trans*-4-hydroxylation of collagen polypeptides^{139, 140, 157} and the HIF-1 α transcription factor in the hypoxic response.^{131, 132} However, hydroxylation of free L-proline has yet to be demonstrated. In contrast, in microbial organisms, hydroxylation of free L-proline occurs, for example, in the biosynthesis of both actinomycin I (by *Streptomyces antibioticus*⁵²) and etamycin (by *Streptomyces griseoviridis* P8648)^{53, 182} (see Section 1.2.2.2); *trans*-4-hydroxy-L-proline is a precursor produced prior to incorporation.^{52, 53, 182} In the case of *Streptomyces griseoviridis* P8648, cell-free extract containing a L-proline hydroxy-

lating species was obtained, and was shown to require 2-OG and ferrous iron for activity, implying the enzyme responsible was a 2-OG oxygenase.^{53,182} Purification of the putative hydroxylase enzyme from *Streptomyces griseoviridis* P8648 (now known as *trans*P4H) revealed that it was, indeed, responsible for the regio- and stereospecific conversion of free L-proline into *trans*-4-hydroxy-L-proline in an 2-OG-/Fe(II)-dependent manner (Scheme 1.18).^{191–193}



Scheme 1.18: The reaction catalysed by *trans*P4H: free L-proline is hydroxylated to produce *trans*-4-hydroxy-L-proline with a retention of stereochemistry.

The stereochemical course of the *trans*P4H reaction was demonstrated by analysing the hydroxylation of deuterium-labelled L-proline substrates (Scheme 1.19);^{191,192} GC-MS analysis showed that hydroxylation of L-proline by *trans*P4H proceeds with a loss of the pro-*R* hydrogen atom at the C-4 position and with an overall retention of stereochemistry.^{191,192}

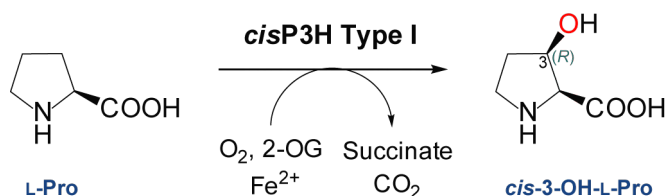


Scheme 1.19: The reaction catalysed by *trans*P4H proceeds with a retention of stereochemistry; deuterated substrates (*trans*-4- and *cis*-4-deutero-L-proline) were chemically synthesised and reactions analysed by GC/MS.

As well as in *Streptomyces griseoviridus* P8648, at least seven other bacterial strains containing *trans*P4Hs have been discovered by microbial screening;¹⁹⁴ one of particular interest is that from *Dactylosporangium* sp. RH1, which has been recombinantly expressed in *E. coli* to produce *trans*-4-hydroxy-L-proline using fermentation methods which could be valuable in industry.^{195–198}

1.2.2.4 L-Proline *cis*-3-Hydroxylase (*cis*P3H)

Whilst screening for L-proline 4-hydroxylases from more than 3,000 strains, Mori *et al.* found a novel enzyme, which hydroxylates free L-proline to produce *cis*-3-hydroxy-L-proline in a number of bacterial strains²² [as well as 8 new actinomycete strains producing *trans*P4Hs (unpublished results)]; these strains include *Streptomyces* sp. strain TH1, *Streptomyces canus* ATCC 12646 and ATCC 12647, and *Bacillus* sp. strains TH2 and TH3.²² These hydroxylating activities here dependent on the presence of 2-oxoglutarate and were thought to be catalysed by an ODD,²² as in the case of *S. griseoviridus* P8648.¹⁹³ The target gene was cloned from *Streptomyces* sp. strain TH1 using PCR-based methods and its product was purified; this was the first report of a purified proline 3-hydroxylase and its structural gene (hence the protein's current name, L-proline *cis*-3-hydroxylase Type I, *cis*P3H Type I).²³ The protein was assayed and proved to catalyse the conversion of free L-proline to produce *cis*-3-hydroxy-L-proline (Scheme 1.20).



Scheme 1.20: The reaction catalysed by L-proline *cis*-3-hydroxylase Type I (*cis*P3H Type I): *cis*P3H Type I converts L-Pro to *cis*-3-hydroxy-L-proline.

Kinetic characterisation of the protein using a Michaelis-Menten model^{199,200} revealed that the K_m s for L-proline and 2-oxoglutarate were 0.56 and 0.11 mM, respectively; the k_{cat} of *cis*-3-hydroxy-L-proline production was 3.2 s⁻¹. The effects of metal salts and organic acids in the 3-hydroxylation reaction were also tested. The addition of Co²⁺, Cu²⁺, or Zn²⁺ at a concentration of 1 mM strongly inhibited the reaction to less than 10% of the positive control rate. Citric acid (5 mM) also inhibited the reaction to about 30% of the positive control rate, but succinate (5 mM), the by-product of ODD reactions, had no effect.²³

An isozyme gene of *cis*P3H Type I was cloned from *Streptomyces* sp. strain TH1;²³ the isozyme gene (870 bp) encodes a protein (*cis*-3-hydroxylase Type II, *cis*P3H Type II) of molecular weight of 33,573 Da. Bioinformatic analysis revealed that the amino acid sequences of the two *cis*P3Hs are 76.2% identical,²³ sharing a conserved iron-binding, (2×His)-(1×Asp)-motif, common to ODDs (Figure 1.17). Although characteristics of both recombinant C3-hydroxylases are similar, the specific activities for turnover of L-proline and proline analogues are different (*cis*P3H Type II is generally more active than Type I); the specific activities of *cis*P3H Type I and II for L-proline turnover are 7,150 U mg⁻¹ and 12,700 U mg⁻¹, respectively.²⁰¹ Kinetic characterisation show that the K_m values for L-proline are 0.20 mM and 0.43 mM for *cis*P3H Type I and II,²⁰¹ respectively, and the K_m values for 2-OG are 0.083 mM and 0.047 mM, respectively.²⁰¹ Although the reactivity of these enzymes differ slightly, they share the same regio- and stereoselectivity.^{195,201} Recombinant *cis*P3H Type II was also shown to be more heat-tolerant than *cis*P3H type I, remaining above 85% optimal activity after the enzyme was incubated at pH 6.0 and below a temperature of 40 °C for 30 min;²⁰¹ the activity of *cis*P3H Type I only remained above 85% under the same conditions if the temperature was below 30 °C.²⁰¹ Both *cis*P3H Type I and Type II are inhibited by the same metal ion and chemicals such as Co²⁺, Ni²⁺, Cu²⁺, Zn²⁺, EDTA, and protocatechuate, with residual activities being

CHAPTER 1. INTRODUCTION

below 50% when each compound was at a concentration of 1 mM.²⁰¹

Amino Acid Conservation Colour Code

Unconserved 0 1 2 3 4 5 6 7 8 9 10 Conserved

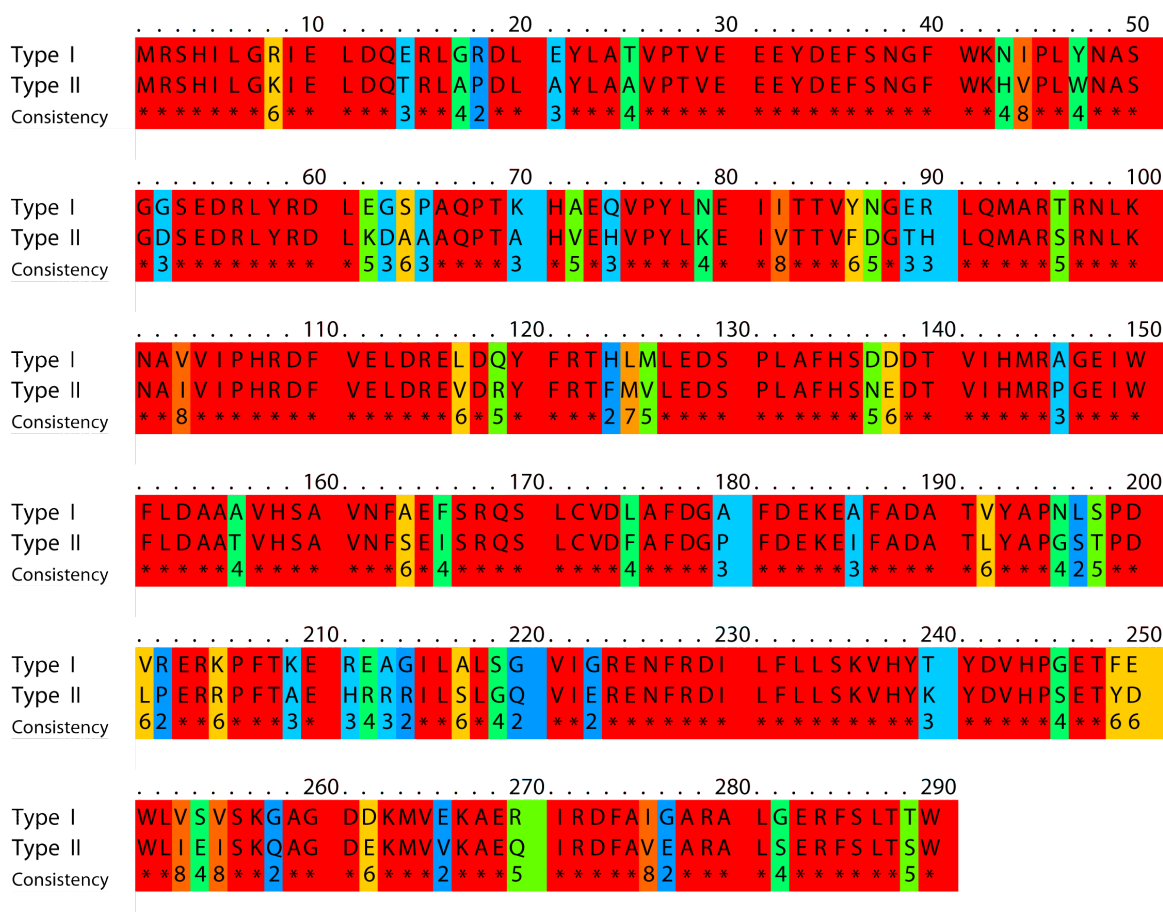


Figure 1.17: A comparison between the deduced amino acid sequences of *cisP3H* Type I and Type II. The sequence alignment and conservation scoring was performed by using PRALINE.^{202,203} The scoring scheme works from 0 for the least conserved alignment position, up to 10 for the most conserved alignment position.

Using recombinantly produced *cis*P3H enzymes (from *E. coli* hosts), *cis*-3-Hyp has been previously biosynthesised using cost-effective, fermentative processes,²⁰⁴ and it is likely that larger-scale, biotechnological methods could be valuable industrial processes.

In 2001, crystal structures were reported for *cis*P3H Type II²⁰⁵ in the absence of iron(II) (apo structure, PDB 1E5R - 2.3 Å resolution, $R_{\text{work}} = 20.2\%$, $R_{\text{free}} = 25.3\%$) and complexed to iron (II) (PDB 1E5S - 2.4 Å resolution, $R_{\text{work}} = 19.8\%$, $R_{\text{free}} = 22.6\%$); these were the first published structures of a free L-amino acid hydroxylase. The structures revealed conserved motifs also present in other ODDs, including the ‘jelly roll’ β -strand core and residues binding iron and 2-OG (Figure 1.18), consistent with divergent evolution within the extended family. The structure differs significantly from many other ODDs, possessing a discrete C-terminal helical domain. The structures suggest potential a model for L-proline binding and a mechanism for uncoupling of proline and 2-OG turnover.

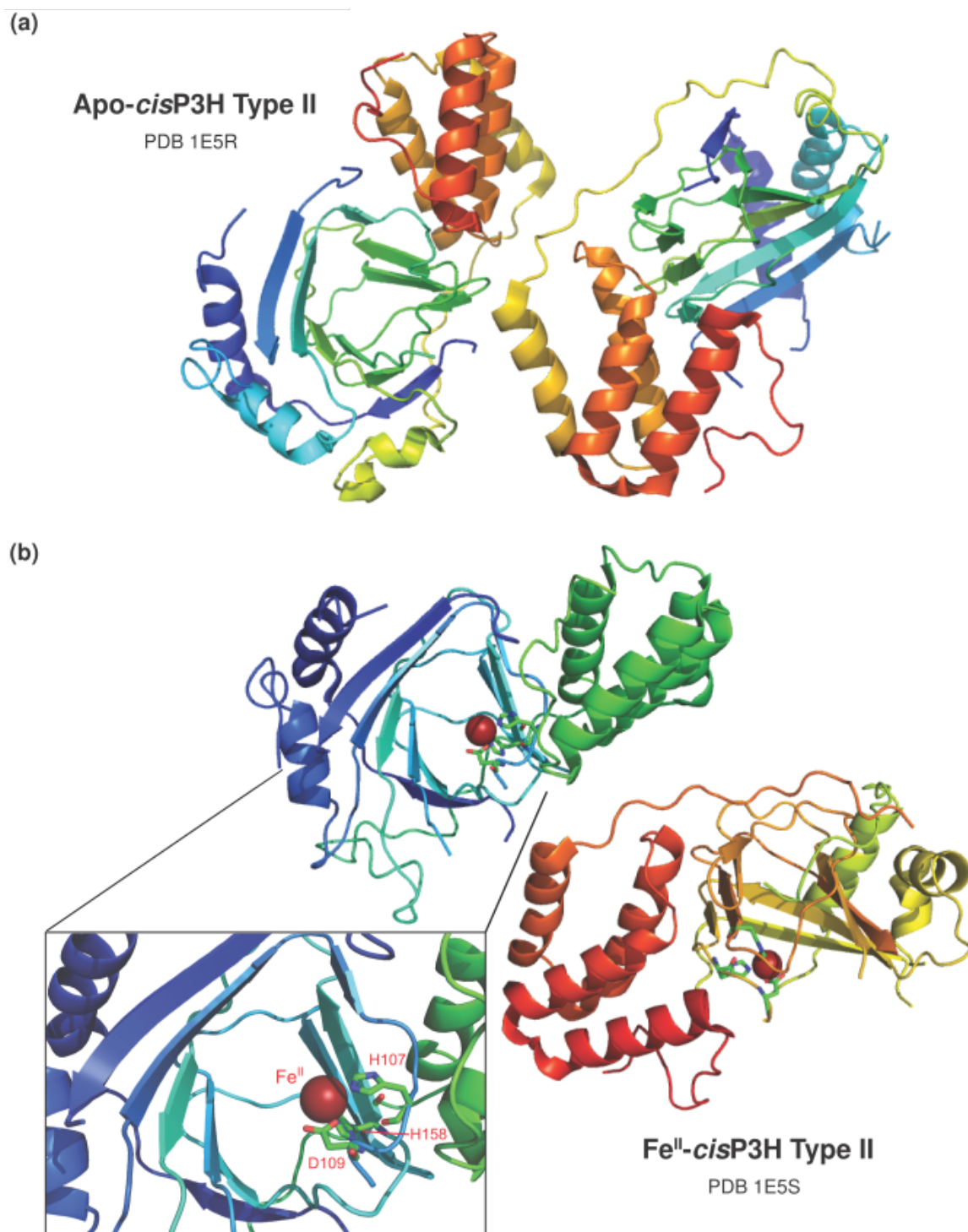
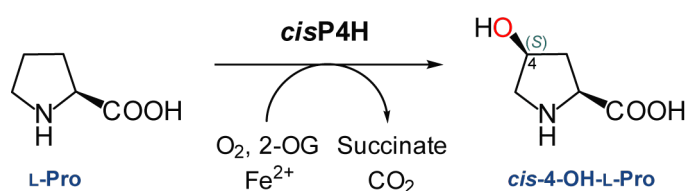


Figure 1.18: Views from X-ray crystal structures of *cis*P3H Type II (both are heterodimers): (a) the apo structure (PDB 1E5R) and (b) the iron(II)-bound form (PDB 1E5S), showing the active site.

1.2.2.5 L-Proline *cis*-4-Hydroxylase (*cis*P4H)

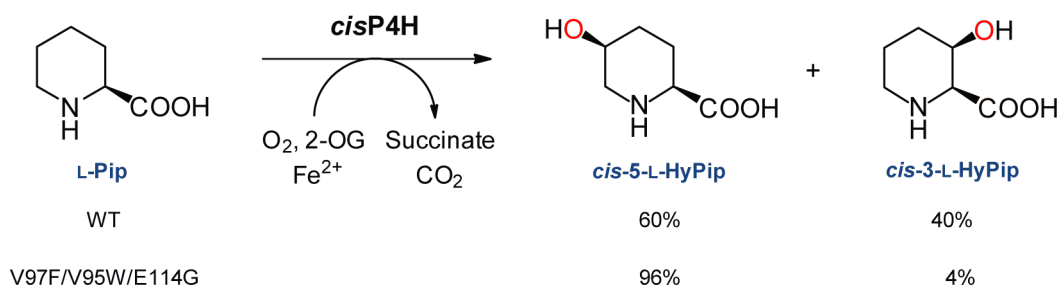
Despite the fact that a number of microorganisms exhibiting L-proline *cis*-4-hydroxylase activity had been identified,^{22,206} microorganisms producing *cis*-4-Hyp from free L-proline had never been reported until 2009.²⁰⁷ Two genes encoding previously uncharacterised proteins from *Mesorhizobium loti*²⁰⁸ and *Sinorhizobium meliloti*,²⁰⁹ nitrogen-fixing bacteria in root nodules of legumes,^{208,209} were cloned and overexpressed in *Escherichia coli*, respectively. Analysis of the functions of the 2 proteins revealed that possessed L-proline *cis*-4-hydroxylase activity;²⁰⁷ the catalytic activity was found to be Fe(II)- and 2-OG dependent, as well as regio- and stereoselective (Scheme 1.21).²⁰⁷ Although their reaction mechanisms were similar to other proline hydroxylases (*trans*P4H and *cis*P3H) and prolyl hydroxylases (PP4H and PHD1/2/3), these *cis*P4Hs (Ml *cis*P4H and Sm *cis*P4H, respectively) do not share much amino acid sequence homology with them (<40%).²⁰⁷ This was the first report of an enzyme that may be applicable to an industrial manufacturing process for *cis*-4-hydroxy-L-proline,²⁰⁷ a compound which was evaluated as an anticancer drug.²¹⁰



Scheme 1.21: The reaction catalysed by L-proline *cis*-4-hydroxylase (*cis*P4H): *cis*P4H converts L-proline to *cis*-4-hydroxy-L-proline.

In 2014, high-resolution, X-ray crystal structures of Ml *cis*P4H were solved with the enzyme in complex with Co²⁺, 2-OG, and substrates L-proline (PDB: 4P7W - 1.80 Å resolution) and L-pipecolic acid (PDB: 4P7X - 1.30 Å).² As well as hydroxylation of the natural L-proline substrate, both Ml*cis*P4H and Sm*cis*P4H hydroxylate of L-pipecolic acid (L-Pip), generating two regioisomers, *cis*-5-Hypip and *cis*-3-Hypip

(Scheme 1.22).²⁰⁷ To selectively produce *cis*-5-Hypip without simultaneous production of two isomers, protein engineering of the *cis*P4Hs was carried out using structural information as the basis for mutagenesis experiments (so-called ‘intelligent design’);[?] three rounds of directed evolution successfully produced a *cis*P4H triple-variant (V97F/V95W/E114G) protein which demonstrated the desired regioselectivity towards generating *cis*-5-Hypip (Scheme 1.22).[?]



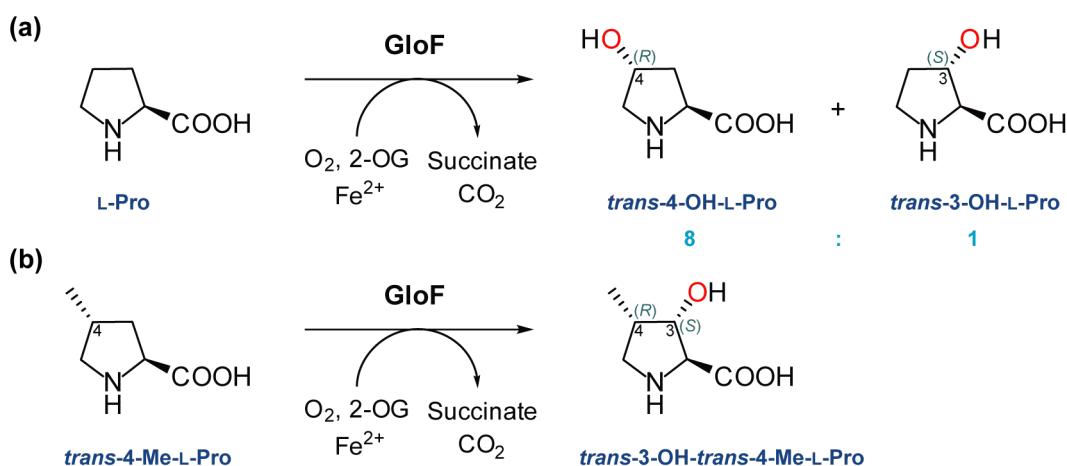
Scheme 1.22: Turnover of L-pipecolic acid by WT *cis*P4H and the regioselective production of *cis*-5-Hypip by a V97F/V95W/E114G *cis*P4H triple-variant.

1.2.2.6 L-Proline *trans*-3/4-Hydroxylase (GloF)

Echinocandins are cyclic, nonribosomal hexapeptides, based mostly on nonproteinogenic amino acids (see Section 1.2.2.2)¹¹⁰ and, despite previous studies on their biosynthesis by fungi, the origin of three amino acids, *trans*-3-hydroxy-4-methylproline, *trans*-4- and *trans*-3-hydroxyproline, are still unknown. However, a 2-OG-dependent L-proline *trans*-3-hydroxylase activity (beside a *trans*P4H activity) was identified in protein crude extracts of the pneumocandin-producing fungus, *Glarea lozoyensis*;²¹¹ the extract also required ascorbate and dithiothreitol for enzymatic turnover, and inclusion of ferrous iron was required for optimal activity,²¹¹ implying the enzyme(s) were 2-OG oxygenases. The nature of the enzymes involved have not been elucidated at the time of publication.²¹¹

Very recently, however, a proline hydroxylase protein, GloF, from the pneumocandin biosynthesis cluster of the fungus *Glarea lozoyensis* ATCC 74030, was iden-

tified;²¹² overexpression and characterisation of GloF revealed it was the first eukaryotic Fe(II)- and 2-OG-dependent proline hydroxylase to be discovered. GloF was shown to catalyse the simultaneous, *in vitro* generation of both *trans*-4- and *trans*-3-hydroxy-L-proline in a ratio of 8:1 from free L-proline (Scheme 1.23 a).²¹² This was the first enzyme with L-proline *trans*-3-hydroxylase (*trans*P3H) activity to be discovered.²¹² GloF also converts *trans*-4-methyl-L-proline into the corresponding *trans*-3-hydroxyproline, *trans*-3-hydroxy-*trans*-4-methyl-L-proline, given that the C4-position is blocked by a methyl group (Scheme 1.23 b).²¹²



Scheme 1.23: The reactions catalysed by GloF, a L-proline *trans*-3/4-hydroxylase: GloF also converts *trans*-4-methyl-L-proline into *trans*-3-hydroxy-*trans*-4-methyl-L-proline.

1.2.3 Ethylene Forming Enzymes

Increases in the demand for petroleum, its limited supply base, and the negative effects of it on the environment have prompted interest in developing alternative, renewable fuels. One potentially useful compound is ethylene (Figure 1.19).

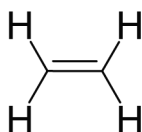


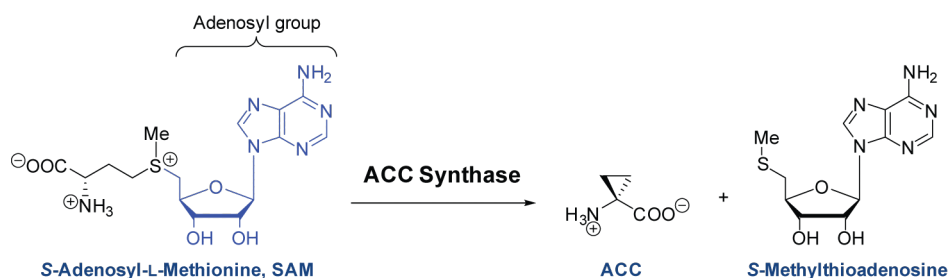
Figure 1.19: The structure of ethylene (ethene).

Worldwide, ethylene is the most produced organic compound, with global production being 142 million metric tonnes (2011). Production is forecast to increase to 165 million metric tonnes in 2015, with an economic impact of US\$200 billion per year.²¹³ Ethylene is also the most widely used feedstock in several industries including plastics, textiles, and solvents, and can be catalytically polymerised to gasoline-rich hydrocarbons in the C5-C10 range.^{214,215} Ethylene is currently produced from steam cracking of fossil fuels or from dehydrogenation of ethane, the largest CO₂-emitting process in the chemical industry, accounting for 1.5% of the United States' carbon footprint.²¹⁶ A renewable source of ethylene would therefore be incredibly useful for both chemical and energy industries, whilst also being environmentally beneficial.

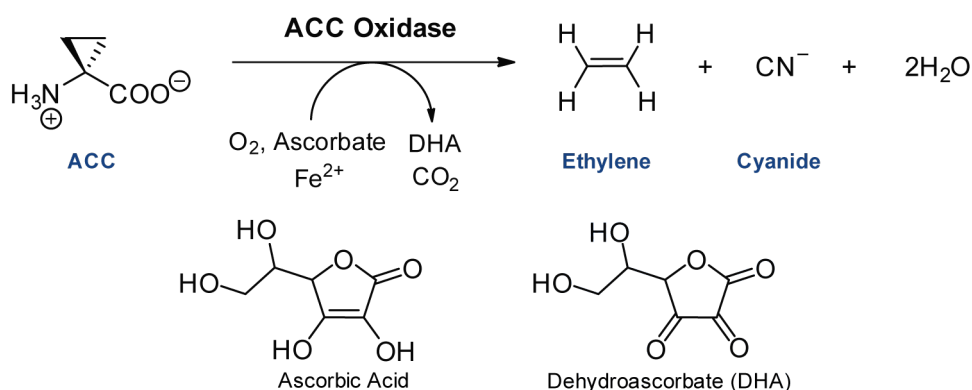
1.2.3.1 1-Aminocyclopropane-1-Carboxylate (ACC) Oxidase

Ethylene is produced biologically, acting as a plant hormone that modulates growth and development, and is involved in the defensive response to abiotic factors or stress (such as pathogen attack).^{217–219} Plants produce ethylene in a 2-step reaction from methionine *via* *S*-adenosyl-methionine (SAM). SAM is initially cyclised to 1-aminocyclopropane-1-carboxylic acid (ACC) by ACC synthase (Scheme 1.24).^{220–225} Subsequently, ACC oxidase catalyses the oxidative fragmentation of ACC to form ethylene and cyanide (CN[−]) (Scheme 1.25).^{220,226,227} As well as being related to 2-OG oxygenases by sequence, ACC oxidase also requires molecular oxygen and ferrous iron cofactors for activity; however, ACC oxidase uses ascorbate as a co-substrate rather than 2-OG and CO₂ is an activator (Scheme 1.25).^{227–230}

Although CN[−] is converted to β -cyanoalanine to avoid toxicity in plants,²³³ using this pathway for biotechnological ethylene production by other organisms is likely limited by the requirement for cyanide mitigation.



Scheme 1.24: The reaction catalysed by ACC Synthase: SAM is converted to ACC and S-thiomethylmethionine.

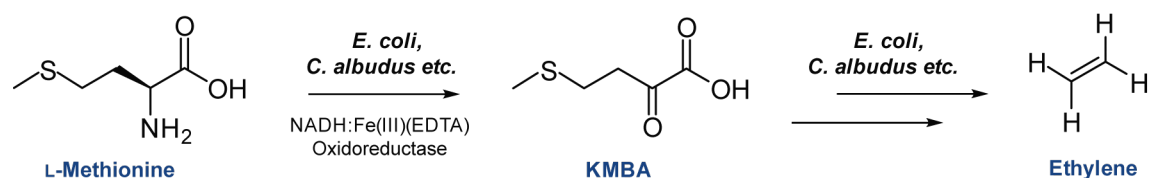


Scheme 1.25: The reaction catalysed by ACC Oxidase involves the oxidative, step-wise cleavage of the ACC ring^{231,232} to form ethylene and cyanide; ferrous iron is a cofactor and ascorbate is cosubstrate.

1.2.3.2 Ethylene-Forming Enzyme (EFE)

A variety of microbes, including bacteria and fungi, also produce ethylene, which probably acts as a casual agent in plant diseases.²³⁴ In a variety of bacteria, such as *E. coli* and *Cryptococcus albidus*, trace amounts of ethylene are spontaneously produced *via* oxidation of 2-keto-4-methylthiobutyric acid (KMBA), a transaminated derivative of methionine produced in an NADH:Fe(III)(EDTA) oxidoreductase-mediated reaction that is enhanced by limiting ammonia availability (C:N ratio = 20:1) (Scheme 1.26).^{235,236} It has been proposed that KMBA formation is a means of recovering amino nitrogen from methionine, resulting in the spontaneous generation of ethylene from KMBA.^{235,236}

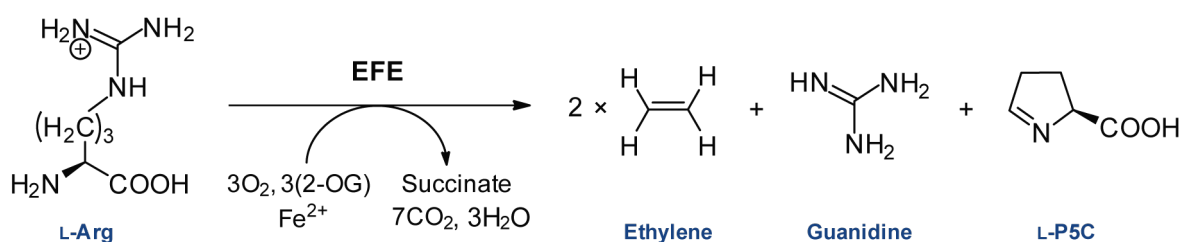
Another ethylene pathway found in *Pseudomonas syringae* and *Penicillium digita-*



Scheme 1.26: The transformation of 2-keto-4-methylthiobutyric acid (KMBA) to ethylene by *Escherichia coli*, *Cryptococcus albidus* etc.; KMBA is produced from L-methionine in NADH:Fe(III)(EDTA) oxidoreductase-mediated, transamination reaction.

tum uses molecular oxygen, 2-OG, and L-arginine as substrates in a reaction catalysed by ethylene-forming enzyme (EFE)^{237–241} Heterologous expression of a single *efe* gene from *P. syringae* resulted in ethylene production in a number of hosts including *E. coli*,^{242,243} *Saccharomyces cerevisiae*,²⁴⁴ *Pseudomonas putida*,²⁴⁵ *Trichoderma viride*,²⁴⁶ *Trichoderma reesei*,²⁴⁷ tobacco,²⁴⁸ and cyanobacteria;^{249–253} these hosts use a variety of carbon sources, including lignocellulose and CO₂, allowing for the potential use of a wide selection of feedstocks for bioethylene production. Ethylene is also non-toxic for these organisms and separates easily from the cultures (gas bubbles up), unlike alcohols, for example.

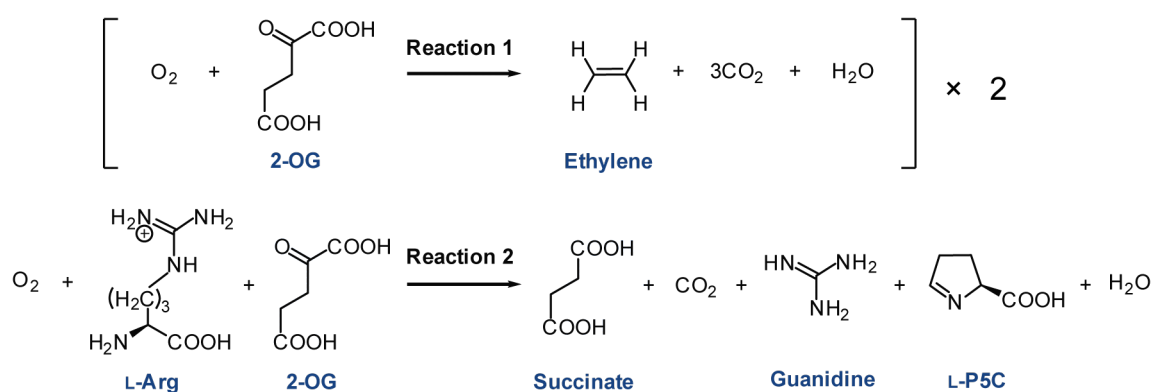
Analyses of cell-free extracts of the fungus *P. digitatum* IF0 9372²⁵⁴ and the Kudzu strain (*Pseudomonas syringae* pv. *phaseolicola* PK2)²⁵⁵ lead to the proposed overall reaction stoichiometry for EFE as shown below (Scheme 1.27).



Scheme 1.27: The overall reaction catalysed by EFE: O₂, 2-OG, and L-arginine are consumed to yield L-Δ¹-pyrroline-5-carboxylate (L-P5C), guanidine, succinate, and CO₂.

A ‘dual-circuit’ process for EFE was proposed involving two parallel reactions occurring in a 2:1 ratio;²⁵⁵ this was based on a substrate turnover ratio of 3:1 for

2-OG and L-arginine and a product ratio of 2:1:1 for ethylene, succinate, and L- Δ^1 -pyrroline-5-carboxylate (L-P5C) (Scheme 1.28 b).²⁵⁵ In the first reaction (major pathway - two cycles), L-arginine remains bound as a ‘cofactor’ while two 2-OG molecules are consumed to produce six CO₂ molecules, two water molecules, and two ethylene molecules (Scheme 1.28 - Reaction 1).²⁵⁵ In the second (minor pathway - 1 cycle) reaction, both 2-OG and L-arginine are consumed to yield L-P5C, guanidine, succinate, and CO₂.



Scheme 1.28: The two pathways catalysed by EFE: in reaction 1 (major), two reaction cycles produce 6CO₂, 2H₂O, and 2 ethylenes from two 2-OG molecules; in reaction 2 (minor), one cycle consumes both 2-OG and L-arginine molecules to generate L-P5C, guanidine, succinate, water, and CO₂.

Despite accounting for the material added and produced in the reactions in these *in vitro* studies,²⁵⁵ this reaction scheme only partly explains the mechanistic behaviour of other related and well-studied enzymes of the 2-OG- and Fe(II)-dependent hydroxylases.¹² In most 2-OG hydroxylases, oxidative decomposition of 2-OG into CO₂ and succinate occurs with coupling to the hydroxylation of a co-substrate (e.g. L-arginine to L-hydroxyarginine).¹² Modern analytical techniques are required to address these inconsistencies in the proposed EFE mechanism;²⁵⁵ it is unclear whether the ethylene-producing reaction is coupled with the succinate producing pathway, or whether EFE is actually promiscuous enough to catalyse two, distinct, parallel reactions.

1.3 Project Aims

The overall aims of the work detailed in this thesis were to study the substrate selectivity of two types of 2-OG oxygenase. Firstly, I aimed to state the limit of substrate selectivity of the proline hydroxylases of different natural regio- and stereoselectivity (from different organisms). Secondly, I aimed to characterise the highly unusual EFE from *Pseudomonas syringae* pv. *phaseolicola* PK2. One aim of this work was to determine which classes of compounds the EFE enzyme could accept as substrates, in order to reveal which chemical motifs are necessary for EFE catalysis to occur.

Chapter 2

Cloning and Protein Expression

This chapter outlines the strategies used to produce the enzymes used in the characterisation experiments described in the subsequent chapters of this thesis; various expression and purification methods were tested and developed to optimise enzyme yields and purity. In order to carry out the desired substrate analogue and biophysical methods, substantial quantities of protein (≥ 10 mg, if possible) are desired.

2.1 Proline Hydroxylase Production

Two proline hydroxylase enzymes had been previously produced within the Schofield Laboratory, comprising *trans*P4H (from *Streptomyces griseoviridis* P8648^{191, 193, 256, 257} and *cis*P3H Type I (from *Streptomyces sp.*²⁵⁸). Methods used to produce these proline hydroxylases were evaluated and used as initial starting points for my own approaches.

2.1.1 *cis*P3H Type I Production

A number of *cis*P3H Type I production methods had been trialled in past work carried out by Küllchen;²⁵⁸ various *E. coli* expression plasmids with different fusion-

peptide sequences were considered. Methods for *cis*P3H Type I production were pursued, as described below, until an acceptable option was found.

2.1.1.1 *cis*P3H Type I Expression Systems

Previous work from the Schofield Laboratory explored the production of recombinant *cis*P3H Type I using various *E. coli* expression vectors (Dr Felix Küllchen's DPhil thesis).²⁵⁸ The previous work is summarised below.

A plasmid encoding for a full-length *Streptomyces sp.* *cis*P3H Type I construct (5'-*HindIII* and 3'-*KpnI* cloning sites) was obtained from collaborators (Kyowa Hakko Kogyo Co. Ltd, Japan, 1999); the vector was a pBR322-based,^{259,260} multicopy expression plasmid, pWTH12, designed to be used in cells grown on a minimal medium.²⁶¹ This pWTH12 plasmid contains an ampicillin-resistance gene (coding for a class A serine β -lactamase) and a *P_{trpx2}* operator sequence (a tandem *Trp* promoter composed of two *E. coli*-derived tryptophan operons^{262,263} connected in series).^{264,265} Expression tests using the pWTH12 vector in *E. coli* JM109 host cells produced yields of *cis*P3H Type I that were deemed unsatisfactory; other expression plasmids were therefore considered to improve protein yields.²⁵⁸

The *cis*P3H Type I construct was then subcloned²⁵⁸ into two different NEB pTYB vectors (derived from the pBR322^{259,260} *E. coli* plasmid),²⁶⁶ namely pTYB1²⁶⁶⁻²⁶⁸ and pTYB11^{266,269} (using 5'-*Nde I* and 3'-*Sap I* sites and 5'-*Sap I* and 3'-*EcoR I* sites, respectively²⁵⁸). Proteins produced using pTYB vectors use the IMPACTTM-CN system (Intein Mediated Purification with an Affinity Chitin-binding Tag), which involves fusing the target protein with a tag consisting of intein and a chitin-binding domain; a solution containing overexpressed protein can then be passed through a chitin column and affinity purified (see Appendix A.1).²⁶⁷⁻²⁶⁹ In the presence of thiols such as DTT, β -mercaptoethanol, or cysteine, the intein tag of the fusion protein undergoes specific self-cleavage, releasing the target protein from the chitin-bound

intein tag; proteins produced using the IMPACT-CN system are therefore purified very simply in one column-step.^{267–269} Expression trials using the pTYB1 construct (*N*-terminal tag) produced very poor *cisP3H* biosynthesis yields, and work using this vector was not pursued.²⁵⁸ In contrast, overexpression of *C*-terminally tagged *cisP3H* protein using the pTYB11 construct gave good yields; however, close inspection of the data revealed that the majority of the *cisP3H* produced was water-insoluble and formed inclusion bodies (*i.e.* inactive).²⁵⁸ A trial purification using the soluble fraction of the *E. coli* cell extract was nevertheless attempted, but, unfortunately, next-to-no *cisP3H* protein was generated.²⁵⁸ Once again, another expression system was required.

2.1.1.2 The pET Expression System

A commonly used and often effective expression system for producing recombinant proteins in *E. coli* is the Novagen pET system;²⁷⁰ expression of target genes in pET plasmids is under strict control of the bacteriophage T7 promoter^{270–273} and makes use of the bacterial hosts chromosomal copy of T7 RNA polymerase gene (these bacteria are termed λ DE3 lysogens).^{270–273} The T7 RNA polymerase is so selective and active that most of an *E. coli* cell's resources are used for the target protein production and the apparently 'overexpressed' protein can, in some cases, comprise over 50% of the total cell protein after a few hours of induction.^{270–273} The expression of the host λ DE3 lysogen's T7 RNA polymerase is under control of the *lacUV5* promoter, which normally enables stringent control of target gene expression.²⁷² Under normal growth, the *lac* repressor (coded by the host cell *lacI* gene) binds the *lacUV5* promoter,²⁷⁴ blocking T7 RNA polymerase from transcribing the target gene; the basal expression of the target gene is therefore under stringent control (*i.e.* 'leaky' expression of the target protein is inhibited). However, the addition of a *lac* repressor-binding substrate to the cell culture induces a conformational change

in the *lac* repressor protein, inactivating it; the *lac* repressor is therefore released from the *lacUV5* promoter site,²⁷⁴ allowing T7 RNA polymerase to bind the promoter and initiate target gene transcription.^{270–273} The ‘natural’ binding substrate of the *lac* repressor protein is allolactose (a disaccharide - Figure 2.1 a), but a structural mimic of allolactose, isopropyl β -D-1-thiogalactopyranoside (IPTG)(Figure 2.1 b), is often used to induce target protein expression in practical applications.^{270–273} Varying the IPTG concentration used to induce protein expression can be used to tune protein expression levels.^{270–273} IPTG is more stable to β -galactosidase hydrolysis than allolactose, as IPTG contains a thiol linkage in the place of an O-glycosidic linkage, and so the induction concentration of IPTG remains more consistent in the cell culture as the target protein is expressed.^{270–273} The most widely used *E. coli* hosts for pET expression systems are derivatives of BL21 strains;^{270,272} BL21 strains are deficient in *lon* and *ompT* proteases, which improves the yields of the recombinant proteins they produce, as well as improving the stability of these overexpressed proteins towards degradation during their purification.

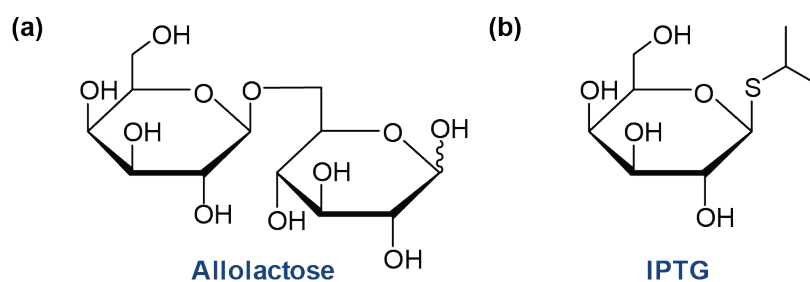


Figure 2.1: The structures of (a) allolactose and (b) isopropyl β -D-1-thiogalactopyranoside (IPTG).

Given the effectiveness of the pET expression system, and the fact that a *transP4H* protein had been previously produced in the Schofield Laboratory²⁵⁷ using a pET-24-a(+) vector, Küllchen *et al.* decided to try to produce recombinant *cisP3H* Type I using a pET-24-a(+) expression plasmid (Figure 2.2).²⁵⁸ The gene encoding for full length *Streptomyces sp. cisP3H* Type I (for sequences, see Appendix A.2.1) was

subcloned into pET-24-a(+) using 5'-*NdeI* and 3'-*HindIII* cloning sites to express tag-less, wild-type *cisP3H* protein.²⁵⁸

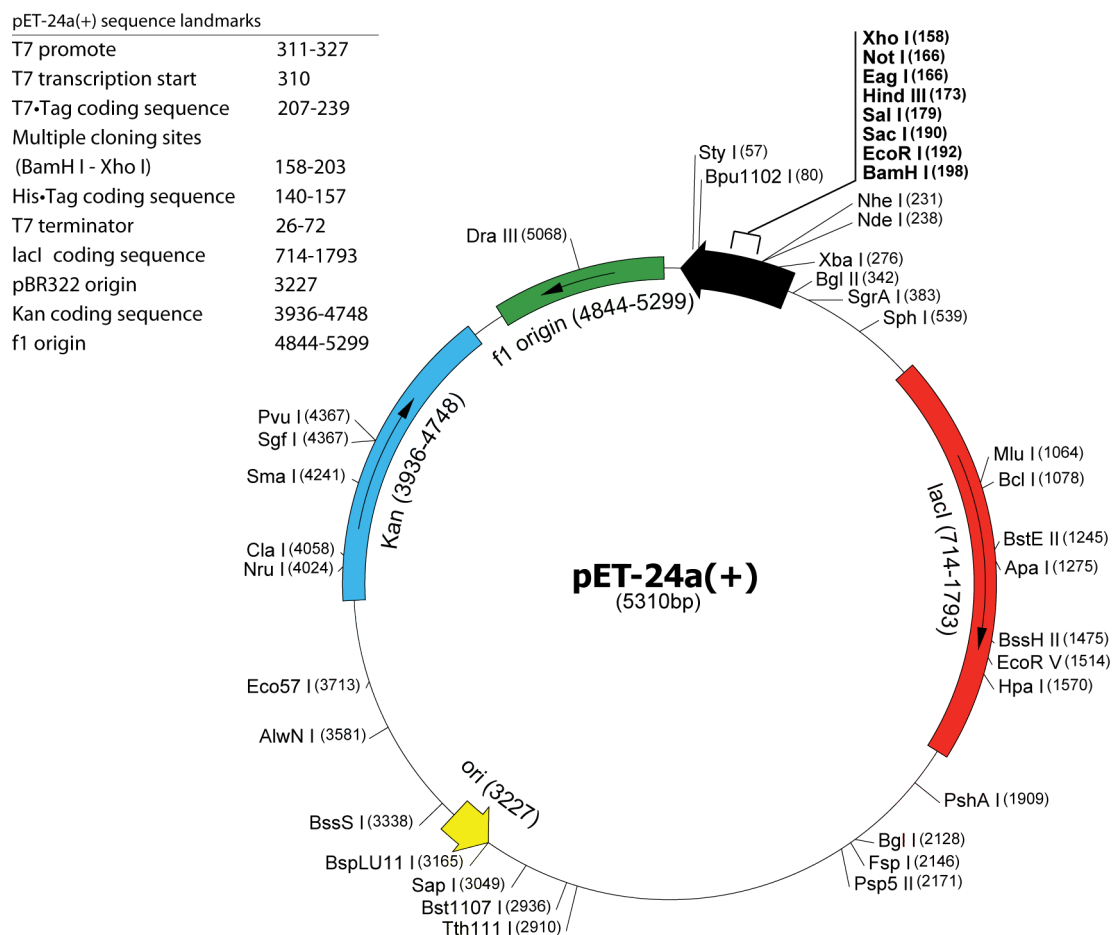


Figure 2.2: The Novagen pET-24-a(+) expression vector map.

The pET-24a-d(+) vectors carry an *N*-terminal T7-Tag sequence plus an optional *C*-terminal His₆-Tag sequence. These plasmids possess a kanamycin-resistance gene for selection and a *lac* operon for (IPTG) expression control. Unique restriction sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. This figure is based on that provided in the Novagen manual.

Küllchen *et al.* carried out expression tests for *cisP3H* Type I in a Novagen pET-24-a(+) vector in BL21 (DE3) host cells grown in 2-YT medium at 27 °C and 37 °C²⁵⁸ (a moderately rich medium commonly used for *E. coli* cell culture²⁶¹ - see Section 5.1.2.1); SDS-PAGE analysis revealed that expression at the lower temperature (27 °C) produced more soluble protein and in sufficient amounts for characterisation.²⁵⁸

The best yields of soluble *cisP3H* Type I were also produced when the culture was induced at a cell density with an OD₆₀₀ (optical density - using 600 nm-wavelength light) between 0.5 and 1.0 (*i.e.* in the middle of the exponential growth phase);²⁵⁸ cell cultures were therefore induced at OD₆₀₀ ~0.8 (the middle of the optimal range) in all further growths.²⁵⁸ The effect of induction length (*i.e.* time between induction and cell harvesting) on yields of *cisP3H* Type I was also explored; cells were cultured at 27 °C and harvested after 3, 4, 5, and 6 hr. Production of *cisP3H* Type I appeared to largely plateau after 5 hr when sufficient protein was produced; a 5 hr induction time was selected for further growths as a compromise between time efficiency and protein yields.²⁵⁸ IPTG induction concentrations were tested in the range of 5.0×10^{-4} mM-1.0 mM; *cisP3H* Type I production was found to be optimal using IPTG concentrations between 0.1 mM and 1.0 mM, and so a concentration of 0.5 mM IPTG was chosen for further work.²⁵⁸ Using 0.5 mM IPTG is for induction a sufficient excess for optimal protein expression, but not so high that the IPTG becomes toxic to the cells or is wasted unnecessarily (IPTG is expensive to purchase).

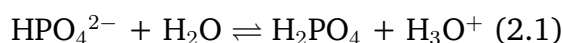
2.1.2 *cisP3H* Type I Expression and Purification

Given that the method using a pET-24-a(+) construct to produce wild-type *cisP3H* Type I was successful,²⁵⁸ the method was developed further in attempt to improve target protein yields and purity.

Initially, the *cisP3H* Type I pET-24-a(+) vector was transformed into a highly competent strain of *E. coli*, BL21-Gold (DE3), which is an improved BL21 competent cell,^{275,276} providing increased transformation efficiency and high- quality miniprep DNA. BL21-Gold-expression strains feature the *Hte* phenotype^{275,276} from Agilent's highest-transformation-efficiency competent cells, XL10-Gold,²⁷⁷ which increases the transformation efficiency of BL21-Gold cells to $>1 \times 10^8$ cfu/μg for a pUC18 standard plasmid.^{275,276} The gene coding for endonuclease I (*endA*), an enzyme

which rapidly degrades plasmid DNA, is also inactivated in BL21-Gold cells.^{275,276} BL21-Gold (DE3) cells can therefore be used for both high-yield, target protein expression and direct cloning of many protein expression constructs.^{275,276}

In an attempt to produce higher yields of *cisP3H* Type I, more nutritious media were tested for improved protein production; the rationale behind this approach was that healthier cells might produce more overexpressed recombinant protein and grow to higher cell densities before the stationary growth phase was reached. A highly nutritious medium known as ‘terrific broth’ (TB)^{261,278} was selected for cells producing *cisP3H* Type I. In terrific broth medium, the yeast extract and tryptone concentrations are increased compared with more conventional media, such as LB or 2-YT, which helps recombinant strains of *E. coli* maintain an extended growth phase and accommodates elevated cell yields.^{261,278} Potassium phosphates are also added to form a phosphate buffer system (Equation 2.1), which maintains the pH of the cell culture broth at relatively neutral for a significant period during cell growth, working to prevent an increase in culture medium acidity with time (organic acids, such as lactic acid, are produced during nutrient metabolism); this buffering effect allows cell cultures to be grown for longer and to higher populations (*i.e.* higher cell density/OD₆₀₀), since cells stay healthy for longer, which delays the onset of the stationary growth phase.^{261,278} Potassium ions are also used by cellular systems, so a useful supplementation. Glycerol is also added as a carbon and energy source.



Equation 2.1: The equilibrium reaction of a phosphate buffer system; from Le Châtelier’s Principle,^{279,280} the position of the equilibrium will shift towards the reactants (left reaction) on the addition of a protic acid in attempt to minimise the effect of the acid on the system.

In attempt to improve the proportion of soluble *cisP3H* Type I produced, a lower expression temperature was tested with an extended induction length; the cell cultures were cooled to ~20 °C (27 °C had been used previously²⁵⁸), induced for *cisP3H*

Type I production with 0.5 mM IPTG (at a cell density of (~ 0.8 after ~ 4 h of growth at 37°C), and then grown overnight (~ 16 h).

2.1.2.1 Small-Scale *cisP3H* Type I Test Purification

A small-scale, 1 L growth of *cisP3H*-producing *E. coli* was first cultured; fresh transformants were used to inoculate a 100 mL 2-YT medium starter cultures (grown overnight to $\text{OD}_{600} \sim 2$ - 10 mL was used to initiate the main culture). The cells were healthy and grew to a high density ($\text{OD}_{600} \sim 7$). A small-scale, protein purification method for *cisP3H* Type I was designed based, in part, on a previous successful method. Cell lysis was carried out in a MES buffer solution (pH 6.0) containing concentrated glycerol [20% (v/v)], and a low concentration of Polysorbate 20 (Tween-20) detergent [0.1% (w/v)], to promote the effective solubilisation of overexpressed *cisP3H* Type I. Protease inhibitor cocktail tablets and a mild reducing agent, tris(2-carboxyethyl)phosphine (TCEP) (0.5 mM), were also added to the lysis buffers to protect the *cisP3H* from proteolytic degradation and oxidation, respectively. A low concentration of the ligand, ethylenediaminetetraacetic acid (0.2 mM EDTA), was also added to the buffers to inhibit metalloproteases (by removing the active-site metal ions from these enzymes) and reducing proteolytic degradation. The method used for cell disruption was sonication; the cell suspension was kept on ice during sonication to protect the *cisP3H* from the damaging heating effects of the sonication probe. Cold (4°C), centrifuge-clarified cell lysate was then treated with a [5% (w/v)] branched-chain polyethylenimine (PEI) solution (Figure 2.3); the PEI solution was prepared in cell lysis buffer from a $\sim 50\%$ (w/v) stock and carefully added dropwise to the magnetically-stirred protein solution until the final PEI concentration reached 0.1% (w/v). Since PEI molecules often have protonated amines in solution (Figure 2.3), they have a net positive charge, which attracts negatively charged DNA molecules and nucleoproteins,²⁸¹ causing them to precip-

itate. The precipitate-containing lysate solution was then re-centrifuged to ‘pellet’ the solid contaminants, and then syringe-filtered for use in subsequent chromatography steps.

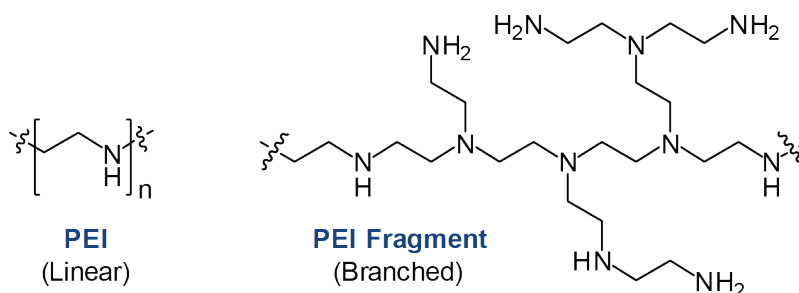


Figure 2.3: The structures of polyethylenimine (PEI) isoforms; the simplest form of PEI is a linear chain, but PEI can also be a branched-chain polymer.

The filtered cell lysate was applied to an equilibrated 30 mL DEAE Sepharose FF anion exchange column, followed by more lysis/binding buffer to wash the column of excess impurities. A step gradient of 45% elution buffer (225 mM NaCl final concentration) was applied to the column to remove the majority of the impurities bound to the DEAE Sepharose FF anion exchange (Figure 2.4 a); Kullchen *et al.*²⁵⁸ found that *cis*P3H remained bound to the DEAE Sepharose resin under these conditions (~225 mM NaCl, pH 6.0), whereas most contaminants eluted. After the DEAE Sepharose resin had been thoroughly washed (*i.e.* when the UV_{280nm}-absorbance trace returned to baseline levels), a linear salt gradient increasing from 45%-100% elution buffer (225 mM-500 mM NaCl) was applied to the column over ~80 mL (~2 $\frac{2}{3}$ column volumes), and the *cis*P3H was eluted from the column (Figure 2.4 a). SDS-PAGE analysis revealed the presence of the protein in the elution fractions (Figure 2.4 b). The fractions containing *cis*P3H Type I (52-60) were pooled, spin concentrated, and frozen (*cis*P3H Type I yield ~20 mg); the hydroxylase activity of the *cis*P3H was later tested and confirmed using an *in vitro* NMR assay (see Section 3.3.1.1).

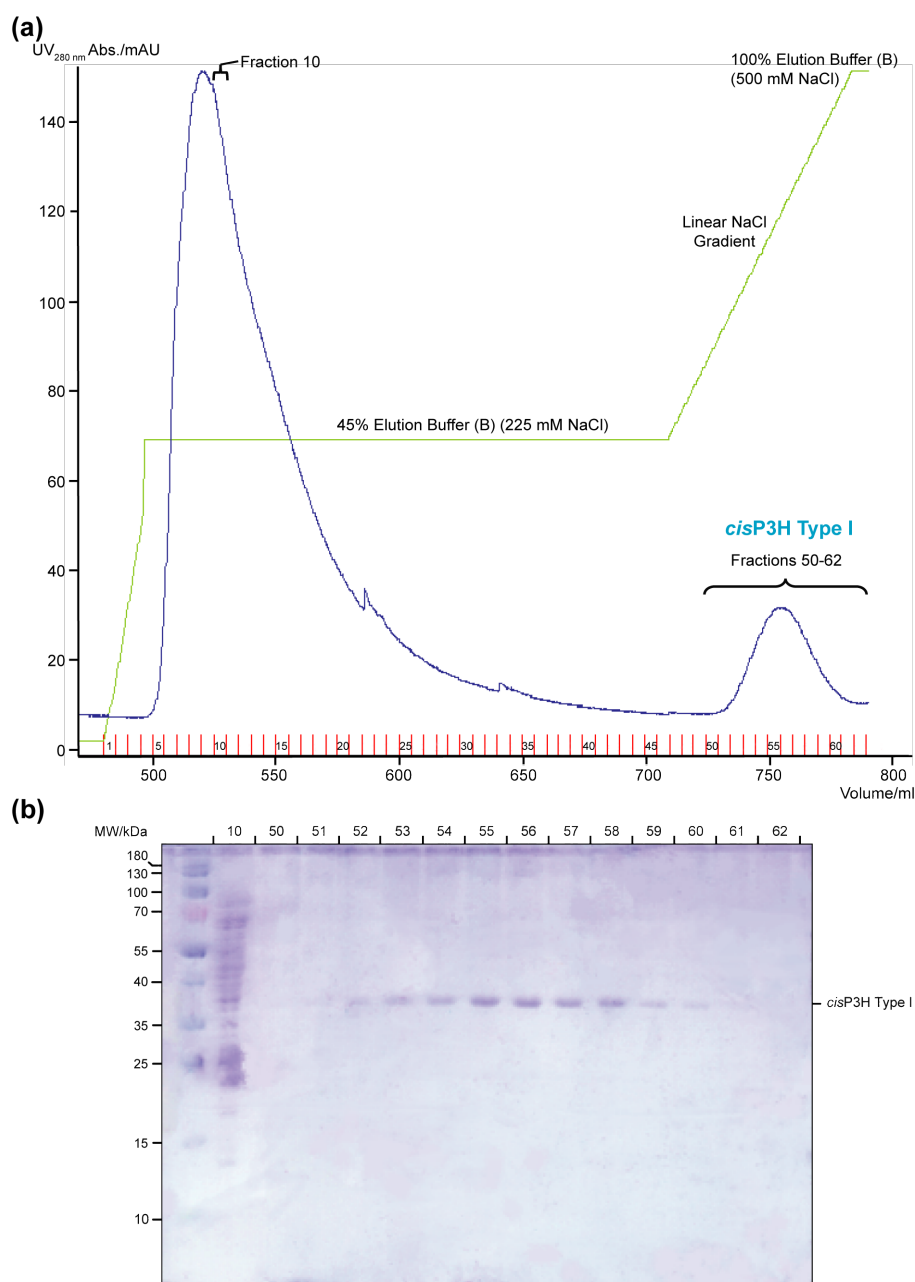


Figure 2.4: Small-scale, anion-exchange-chromatography (AEC) purification of *cisP3H* Type I.

(a) The UV_{280nm}-absorbance FPLC chromatogram for the *cisP3H* Type I AEC.

(b) The fractions were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] to assess protein purity. The marker denotes the target protein; the bands ran at a mass of the correct order of magnitude.

Column: 30 mL Pharmacia DEAE Sepharose Fast Flow

Buffer A: 50 mM MES-NaOH, 0.2 mM EDTA, 20% (v/v) glycerol, 0.1% (w/v) Tween 20, 0.5 mM TCEP, pH 6.0.

Buffer B: 50 mM MES-NaOH, 500 mM NaCl, 0.2 mM EDTA, 10% (v/v) glycerol, pH 6.0.

2.1.2.2 Large-Scale *cisP3H* Type I Test Purification

Given the success of the small-scale purification of active *cisP3H* Type I (Section 2.1.2.1), the method was scaled up 5-fold and a test purification attempted. The first few steps were based on the initial, small-scale test: the same buffers were used; cell lysis was achieved by sonication; lysates were treated with PEI to remove DNA and other contaminants; and the first chromatographic step was anion-exchange chromatography. The same 30 mL DEAE Sepharose FF AEC column was also used, since the binding capacity of the resin was predicted to be sufficient (based on the properties quoted by the resin manufacturer²⁸²); the same buffer gradients used in the small-scale purification could therefore be employed in this larger-scale purification [*i.e.* a step-gradient-wash step: 0%-45% elution buffer (B)], followed by a linear salt gradient [45%-100% elution buffer (B)] to elute *cisP3H* from the column (Figure 2.5). In addition, some buffer volumes were scaled up accordingly; for example, the *E. coli* cell pellets (containing *cisP3H* Type I) were resuspended in roughly 5 times the volume of lysis buffer used in the small-scale purification, in an attempt to keep the column-binding protein concentrations the same; the volumes of buffer used to wash the column were also increased slightly to account for increased quantities of contaminants with greater scale.

The yields of *cisP3H* Type I using this procedure were good (~100 mg from a single preparation; the concentration was ~55 mg/mL) and the level of purity seemed similar to that achieved on the smaller scale (*i.e.* it appeared to be only slightly less pure than in the small-scale purification by SDS-PAGE analysis). The fractions containing *cisP3H* Type I (fractions 4-22) were pooled and spin-concentrated (to ~1 mL).

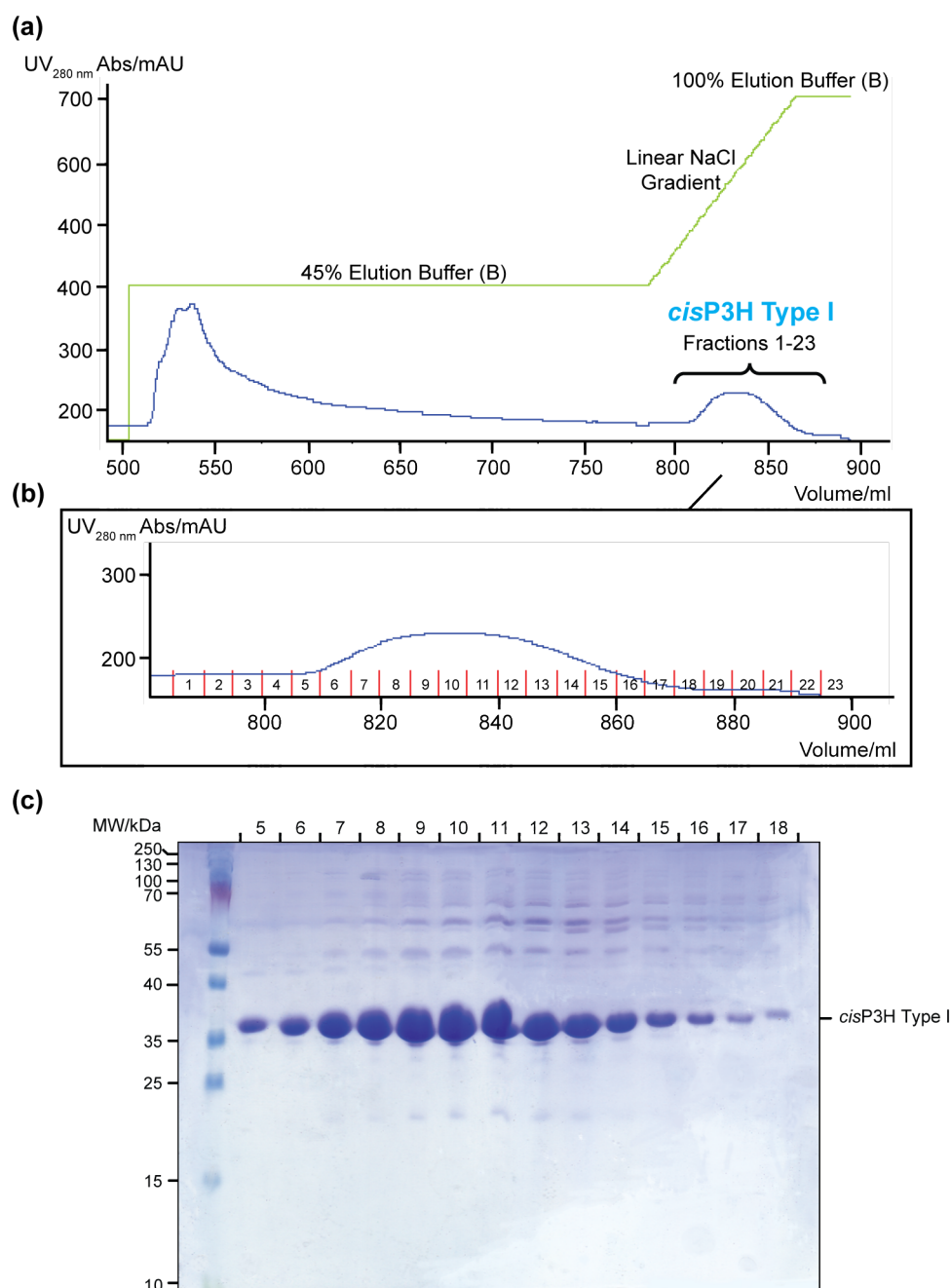


Figure 2.5: Large-scale, anion-exchange-chromatography (AEC) purification of *cisP3H* Type I.

(a) The UV_{280nm}-absorbance FPLC chromatogram for the *cisP3H* Type I AEC.

(b) The fractions were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] to assess protein purity. The marker denotes the target protein; the bands ran at a mass of the correct order of magnitude.

Column: 30 mL Pharmacia DEAE Sepharose Fast Flow

Buffer A: 50 mM MES-NaOH, 0.2 mM EDTA, 20% (v/v) glycerol, 0.1% (w/v) Tween 20, 0.5 mM TCEP, pH 6.0.

Buffer B: 50 mM MES-NaOH, 500 mM NaCl, 0.2 mM EDTA, 10% (v/v) glycerol, pH 6.0.

SDS-PAGE analysis of the DEAE-Sepharose AEC purification fractions (Figure 2.5) revealed that most contaminants were either substantially heavier (possible aggregates) or much lighter than *cis*P3H Type I; it seemed like it may be possible to separate *cis*P3H Type I from its contaminants based on protein size. Size-exclusion chromatography was therefore chosen as a second column-step. A gel filtration step should also remove excess glycerol and detergent from the protein, chemicals which may interfere with subsequent experiments; in particular, assays and crystallisation experiments are often sensitive to conditions.

The concentrated *cis*P3H fractions were loaded onto a preparative-scale (~320 mL) Pharmacia HiLoad 26/60 Superdex 75 size-exclusion column. The size exclusion step did indeed improve the purity of the *cis*P3H Type I, largely separating it from high-molecular-weight impurities, which eluted from the column earlier than *cis*P3H Type I (Figure 2.6); the fractions containing *cis*P3H Type I only contained faint degradation bands of low molecular weight (MW) (Figure 2.6). The purest fractions (4-11) were pooled, spin-concentrated, and frozen; protein purity was estimated at >85% by SDS-PAGE analysis (Figure 2.6).

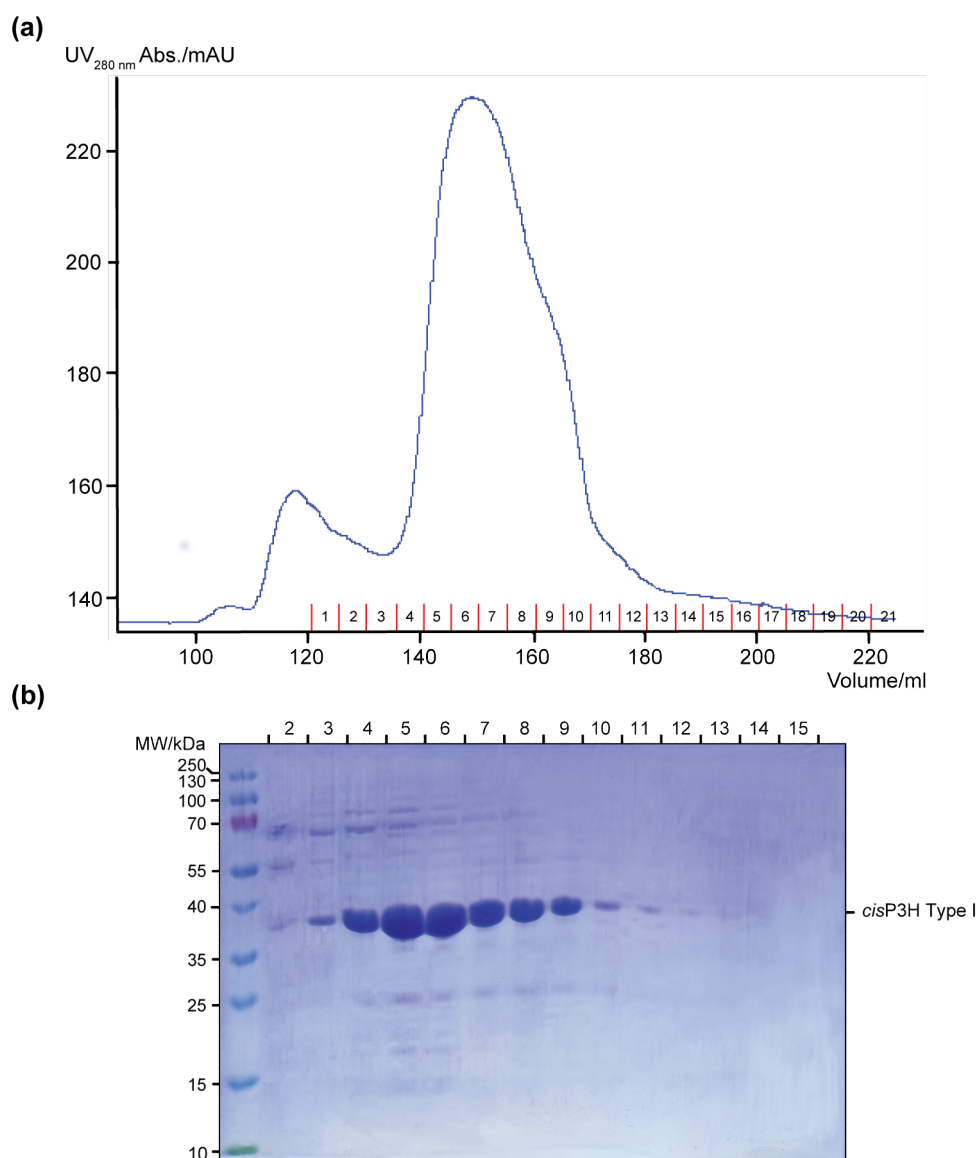


Figure 2.6: Preparative-scale size-exclusion chromatography of *cisP3H* Type I fractions.

(a) The UV_{280nm}-absorbance FPLC chromatogram for the *cisP3H* Type I gel filtration fractions; (b) The fractions were analysed by tris-glycine SDS-PAGE [12.5% (w/v) acrylamide] to assess protein purity. The marker denotes the target protein; the bands ran at a mass of the correct order of magnitude. High-MW impurities are eluted before the major *cisP3H* Type I peak, which has a slight shoulder (right side) produced by a small amount of low-MW contamination.

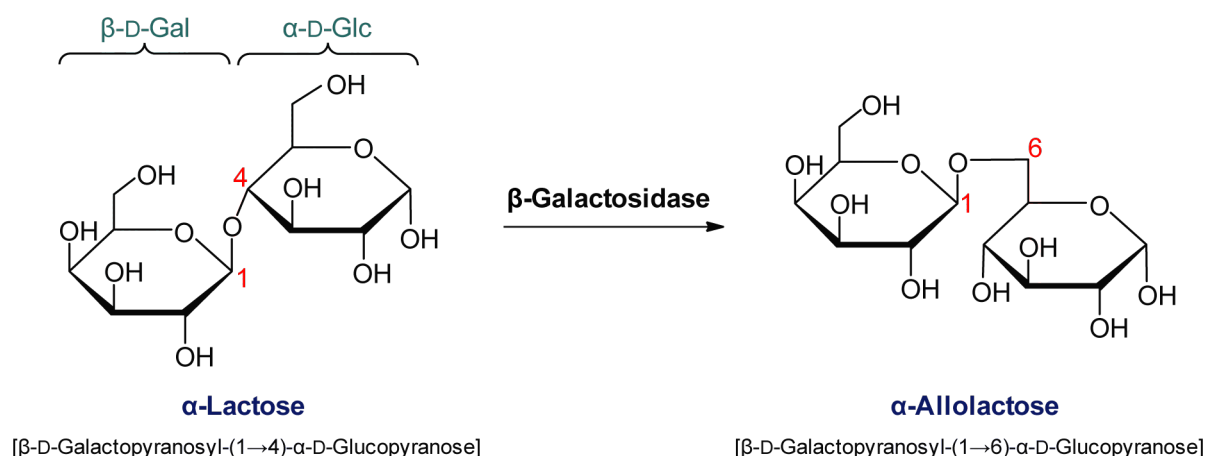
Column: Pharmacia (320 mL) HiLoad 26/60 Superdex 75.

Buffer: 150 mM Tris-HCl, pH 7.5.

2.1.2.3 *cis*P3H Type I Production by Auto-Induction

To increase yields of *cis*P3H Type I for biochemical studies, other cell growth media were then explored. A method using a highly-nutritious, auto-induction *E. coli* growth medium, ZYM-5052,²⁸³ was tested for *cis*P3H Type I production; this particular medium was chosen because of its similarity with terrific broth (TB),^{261,278} which had been identified as a successful culture medium for *cis*P3H Type I production. Both terrific broth and ZYM-5052 contain similar quantities of yeast extract, tryptone, and glycerol, as well as a potassium phosphate buffer system to help maintain cell culture pH-neutrality.^{261,283} Auto-induction media do not require the addition of an inducer, such as IPTG, to cell cultures for target protein expression; instead, the chemical components of an auto-induction medium regulate the expression of recombinant proteins.²⁸³ The ZYM-5052 system uses a mixture of D-glucose (minor component) and α -lactose (major component)²⁸³ to control *lac*-operon-regulated target protein expression, forming a so-called glucose-lactose 'diauxie'; this a system in which batch-cultured cells sequentially metabolise two available sugars in two separate growth phases rather than simultaneously.^{283–286} In the initial stages of a ZYM-5052 batch growth, *E. coli* cells will preferentially use D-glucose as a carbon source and α -lactose will not be significantly metabolised; the cell density therefore rises rapidly and cell growth reaches the exponential phase before target protein expression has been induced.^{283–287} D-Glucose was shown to have an inhibitory effect on the formation of β -galactosidase by *E. coli*,^{285,287} the enzyme which catalyses the hydrolysis of β -galactosides, into monosaccharides. In the case of lactose [β -D-galactose-(1 $\rightarrow$ 4)- α -D-glucose], β -galactosidase can catalyse the conversion of lactose into allolactose [β -D-galactose-(1 $\rightarrow$ 6)- α -D-glucose] (Scheme 2.1), the natural inducer of the *lac* operon.²⁸⁸ Inhibition of β -galactosidase synthesis by *E. coli* therefore prevents the expression of *lac*-operon-controlled target proteins.

By the time a ZYM-5052 *E. coli* batch culture reaches its exponential growth phase, almost all of the D-glucose present has been metabolised (only a small amount is added) and the cells then begin to metabolise lactose as their primary carbon source instead (*i.e.* a ‘diauxic’ shift occurs).^{283,285–287} Lactose metabolism by the *E. coli* produces readily-available allolactose (converted from lactose), triggering the expression of *lac*-operon-controlled target genes.²⁸³ Since target protein production is automatically induced at the optimal point during the batch growth, cultures only need to be inoculated and grown to saturation; yields of target protein are typically several-fold higher than obtained by conventional IPTG induction.²⁸³



Scheme 2.1: The conversion of lactose to allolactose in a transglycosylation reaction catalysed by β -galactosidase (coded by *lacZ* gene in *lac* operon).

A batch of 4×1 L ZYM-5052²⁸³ cultures was initially grown for large-scale production of *cisP3H* Type I; an incubation method similar that used to produce *cisP3H* Type I using terrific broth was implemented (*i.e.* cells in shaking flasks are grown for ~4 h at 37 °C and then at 20 °C overnight for ~16 h). Cells from these test cultures were then resuspended in lysis buffer and *cisP3H* Type I was initially purified (Figures 2.7 and 2.8) by the same method used previously (*i.e.* AEC followed by a size exclusion step - see Section 2.1.2.2). The yields of *cisP3H* Type I were extremely good using the ZYM-5052 autoinduction medium (~200 mg at ~35 mg/mL), yielding roughly twice the quantity of *cisP3H* Type I produced using terrific broth as a growth medium and IPTG induction; the *cisP3H* Type I produced was also of similar purity (>85%). The ZYM-5052 auto-induction medium was therefore used to produce sufficient *cisP3H* in the majority of the subsequent experiments.

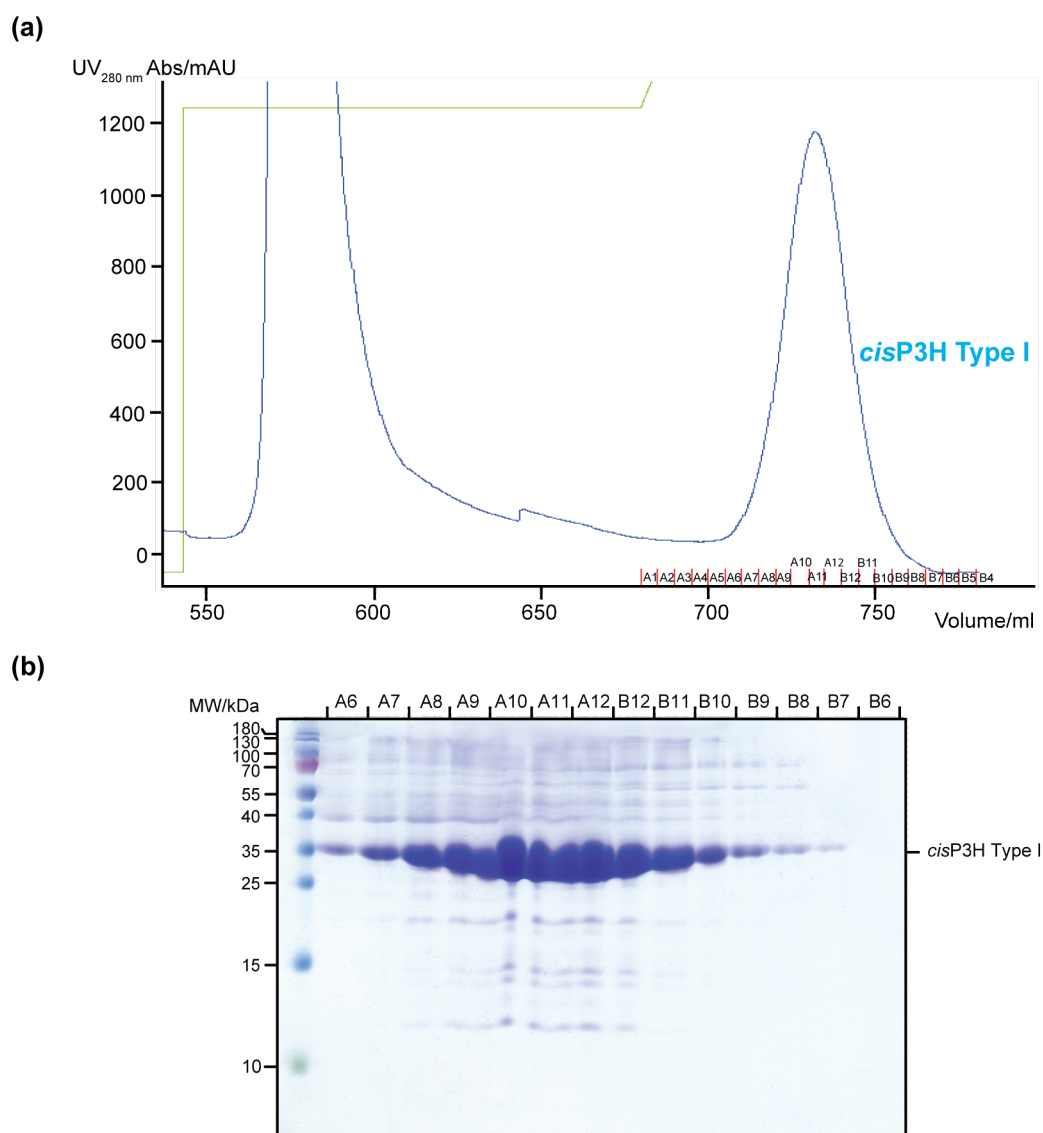


Figure 2.7: Anion-exchange-chromatography (AEC) purification of *cisP3H* Type I produced by *E. coli* cells grown on ZYM-5052 auto-induction media.

(a) The UV_{280nm} -absorbance FPLC chromatogram for the *cisP3H* Type I gel filtration fractions.

(b) The fractions were analysed by tris-glycine SDS-PAGE [15%(w/v) acrylamide] to assess protein purity. 5 μ L samples of each fraction were run in each lane. The marker denotes the target protein; the bands ran at a mass of the correct order of magnitude.

Column: 30 mL Pharmacia DEAE Sepharose Fast Flow

Buffer A: 50 mM MES-NaOH, 0.2 mM EDTA, 20% (v/v) glycerol, 0.1% (w/v) Tween 20, 0.5 mM TCEP, pH 6.0.

Buffer B: 50 mM MES-NaOH, 500 mM NaCl, 0.2 mM EDTA, 10% (v/v) glycerol, pH 6.0.

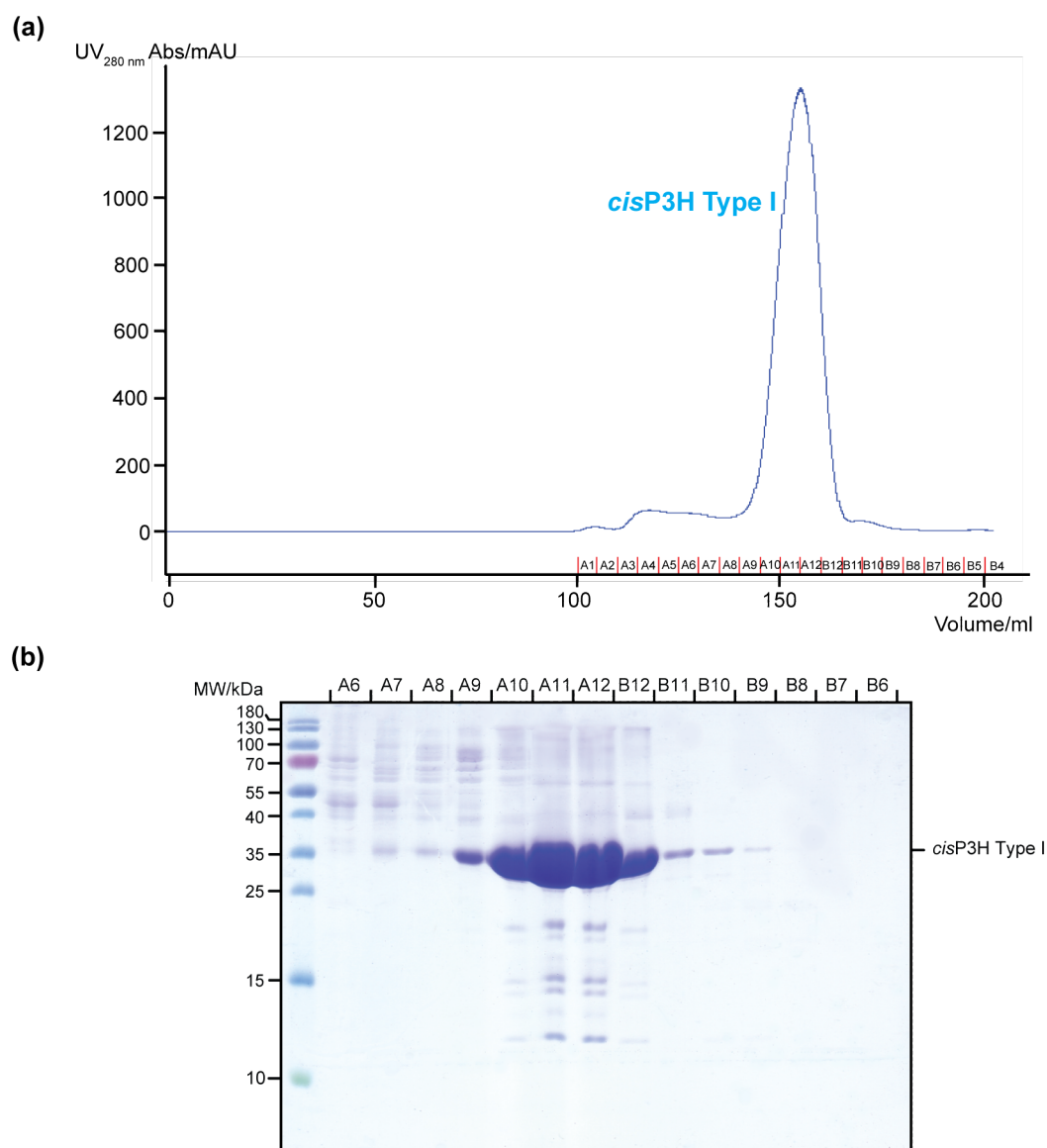


Figure 2.8: Size-exclusion chromatography of *cisP3H* Type I produced by *E. coli* cells grown on ZYM-5052 auto-induction media.

(a) The UV_{280nm} -absorbance FPLC chromatogram for the *cisP3H* Type I gel filtration fractions; (b) The fractions were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] to assess protein purity. 2 μ L samples of each fraction were run in each lane. The marker denotes the target protein; the bands ran at a mass of the correct order of magnitude.

Column: Pharmacia (320 mL) HiLoad 26/60 Superdex 75

Buffer: 50 mM MES-NaOH, 200 mM NaCl, pH 6.5.

To test whether it would be possible to improve the purity of the *cis*P3H protein batch, the purest *cis*P3H fractions from the size exclusion step were concentrated and buffer-exchanged (into 20 mM Tris-HCl, pH 7.6) using a 10-kDa-cut-off spin concentrator, loaded onto an 8 mL Mono Q anion-exchange column. This purification step was carried out at a different pH (7.6) from that used in the initial DEAE Sepharose (weak) anion-exchange step (pH 6.0). The resin of a Mono Q column is strong ion exchanger often used in protein ‘polishing’ steps; it was hoped that impurities would interact with the Mono Q resin in manner different from *cis*P3H, so that the *cis*P3H could be separated from the contaminants. The bound protein was eluted from the Mono Q resin using a linear gradient of salt (0 mM-500 mM NaCl) based on typical procedures suggested by the manufacturer of the resin²⁸⁹ (Figure 2.9). However, the MonoQ-based purification step did not result in a significant increase in purity. Although the Mono Q column separated the *cis*P3H Type I slightly from high-MW contaminants, with most of the *cis*P3H Type I eluting slightly earlier than these contaminants, the *cis*P3H was not sufficiently resolved from the contaminants; the major *cis*P3H peak in the UV_{280nm}-absorbance chromatogram has a noticeable shoulder on its right side. In addition, the increased length of the purification (and perhaps the change of buffer pH) resulted in slightly more *cis*P3H degradation products; inspection of the Mono Q purification SDS-PAGE gel (Figure 2.9 b) shows an increase in the number and intensity of lower MW bands compared with the previous gel filtration step. Unless more effective buffer conditions and/or improved salt gradients are possible, a method using an additional Mono Q anion-exchange column step is probably not appropriate to improve *cis*P3H purity.

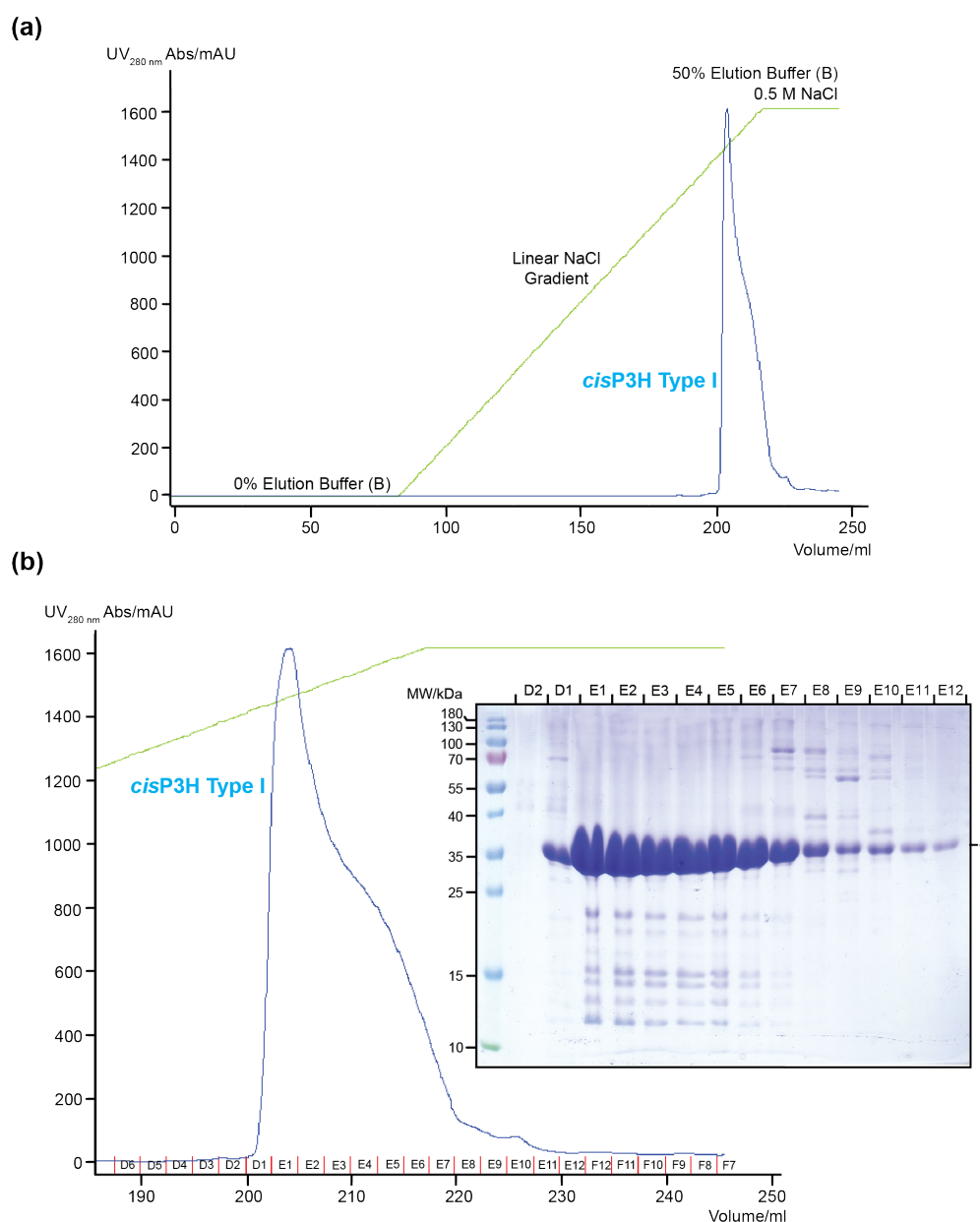


Figure 2.9: Mono Q anion-exchange chromatography of *cisP3H* Type I produced by *E. coli* cells grown on ZYM-5052 auto-induction media.

(a) The UV_{280nm} -absorbance FPLC chromatogram for the *cisP3H* Type I gel filtration fractions.

(b) The fractions were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] to assess protein purity. 5 μ L samples of each fraction were run in each lane. The * marker denotes the target protein; the bands ran at a mass of the correct order of magnitude.

Column: Pharmacia (8 mL) Mono Q Anion-Exchange

Buffer A: 20 mM Tris-HCl, pH 7.6.

Buffer B: 20 mM Tris-HCl, 1 M NaCl, pH 7.6.

2.1.2.4 *cisP3H* Type I Purification Development

Other combinations of chromatography steps for the purification of *cisP3H* Type I were attempted; a method using a larger DEAE Sepharose FF column (with a resin volume of 250 mL compared to 30 mL) was tested to investigate if using the larger DEAE Sepharose FF column would improve the separation of *cisP3H* Type I from its contaminants (Figure 2.10). We had hoped that the increase in resin volume and volumes of buffers flowing through the column may magnify the differences in the resin-binding characteristics between *cisP3H* Type I and its contaminants (for superior peak resolution). The improvement in separation from using a larger column, however, was probably not significant enough to improve *cisP3H* Type I purity over using a smaller one (Figure 2.10); the UV_{280nm} chromatogram for the purification shows slight resolving of the contaminant peaks from the major *cisP3H* Type I elution peak, but there is still significant overlap (Figure 2.10).

The purification also takes longer with a larger column, since more buffer has to flow through the column to complete the purification (litres as opposed to a few hundred mL); the *cisP3H* Type I protein is also eluted into a larger volume of buffer and this needs to be spin-concentrated and buffer-exchanged to carry out the next purification step. This extra preparation time may result in an increase in proteolytic degradation of the *cisP3H* Type I product if one is not careful. The original 30 mL DEAE Sepharose FF anion-exchange column chosen is therefore more still appropriate for the first column step of the *cisP3H* Type I purification.

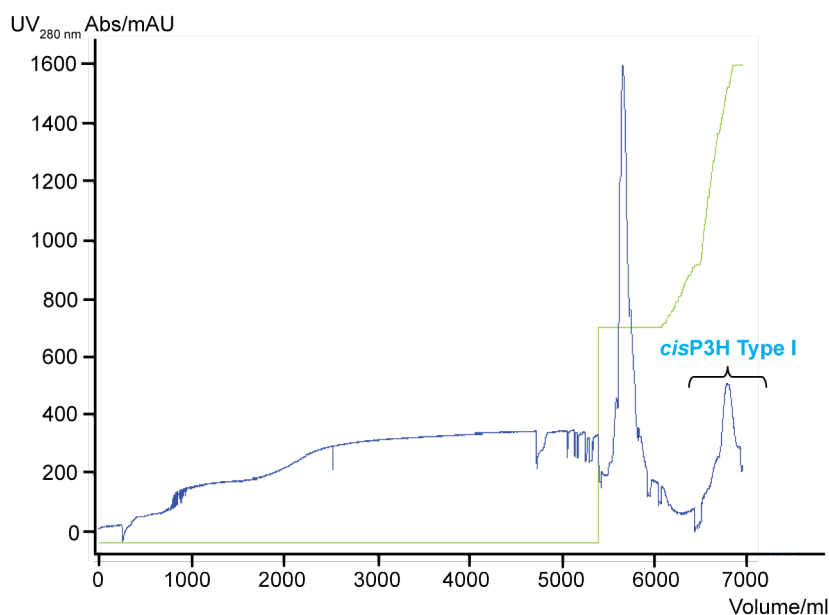


Figure 2.10: Anion-exchange-chromatography (AEC) purification of *cisP3H* Type I produced by *E. coli* cells grown on ZYM-5052 auto-induction media: the UV_{280nm}-absorbance FPLC chromatogram for the *cisP3H* Type I purification fractions.

Column: 250 mL Pharmacia DEAE Sepharose Fast Flow

Buffer A: 50 mM MES-NaOH, 0.2 mM EDTA, 20% (v/v) glycerol, 0.1% (w/v) Tween 20, 0.5 mM TCEP, pH 6.0.

Buffer B: 50 mM MES-NaOH, 500 mM NaCl, 0.2 mM EDTA, 10% (v/v) glycerol, pH 6.0.

Previous work showed that, under some conditions, the aqueous solubility of *cisP3H* Type I was poor²⁵⁸ owing to the protein's high hydrophobic amino-acid composition. Given the hydrophobic properties of *cisP3H* Type I, an intermediate, hydrophobic-interaction chromatography step was explored to improve *cisP3H* Type I purity; a Phenyl Sepharose HP (HiLoad 26/10) was used to purify the *cisP3H* Type I fractions from the previous DEAE Sepharose FF AEC step (Figure 2.11 a) before performing a final Superdex 75 size-exclusion step (Figure 2.11 b). Although the *cisP3H* Type I protein did bind to the phenyl sepharose resin quite well, eluting sharply at low salt concentrations, it appeared that the contaminants also bound well (some probably being degradation products or oligomers). Since a small amount of the *cisP3H* did not bind the column in the HIC loading stages (*i.e.* was

present in the flow-through), and most of the contaminants did bind, the net result was a slight decrease in the ratio of *cis*P3H to contaminants in the HIC elution fractions (*i.e.* the quantity of *cis*P3H Type I fell whilst the level of contamination remained the same, causing a slight drop in purity). Analysis of the chromatogram from the size-exclusion step that followed the HIC step reveals a more prominent shoulder on the right side of the *cis*P3H Type I peak than seen for purifications that did not use a HIC step (Figure 2.11 b); this is consistent with the *cis*P3H peak being less intense, owing to purification losses of *cis*P3H, whilst the contamination peak remained roughly the same intensity. Unless more appropriate conditions are possible, this particular HIC column is probably not effective at increasing *cis*P3H Type I purity.

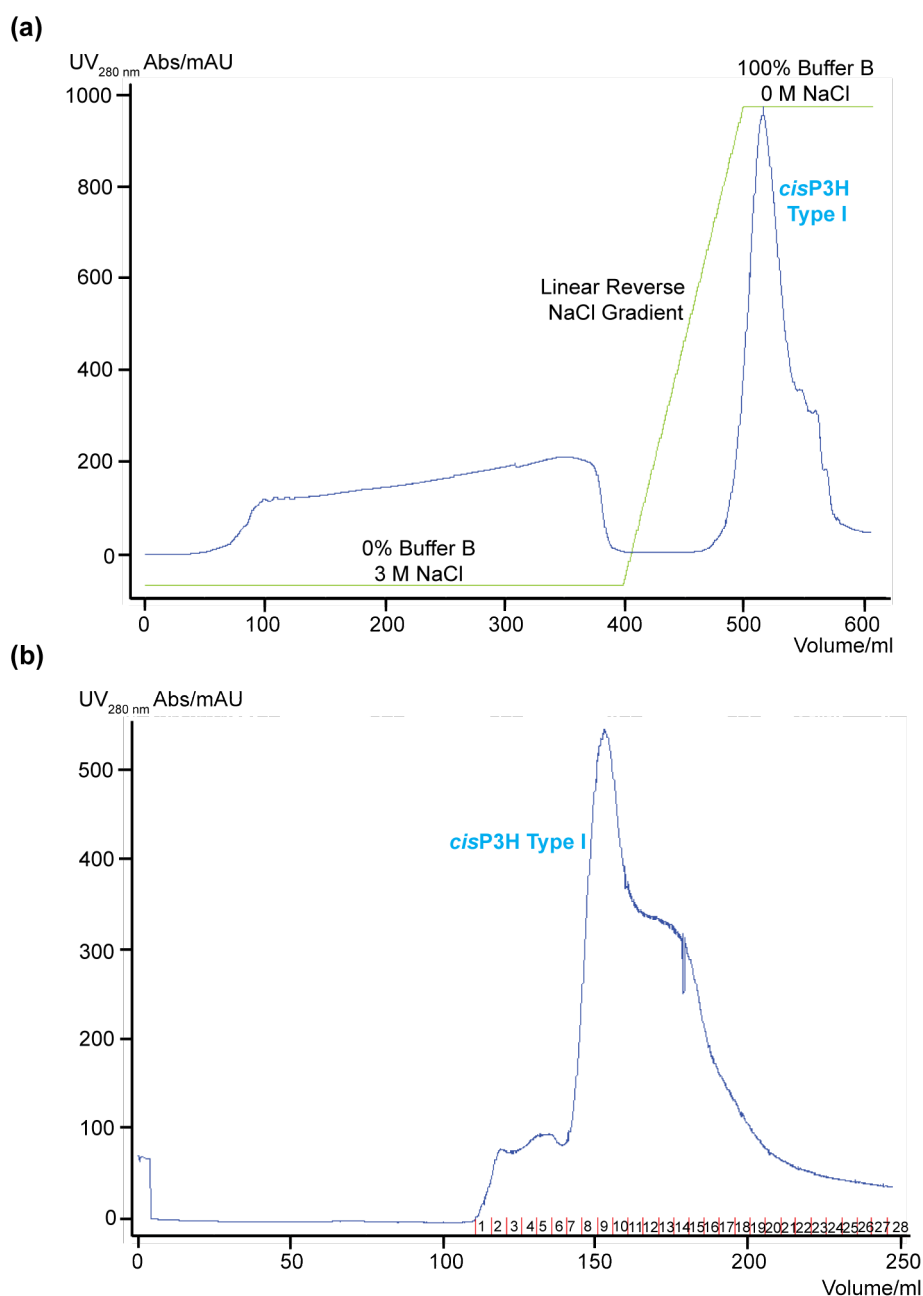


Figure 2.11: The UV_{280nm}-absorbance FPLC chromatograms for the purification of *cisP3H Type I* by (a) hydrophobic-interaction-chromatography (HIC) and (b) size-exclusion chromatography (SEC).

HIC Column: 250 mL Pharmacia DEAE Sepharose Fast Flow
HIC Buffer A: 50 mM MES-NaOH, 3 M NaCl, 5% (v/v) glycerol, pH 6.0.
HIC Buffer B: 50 mM MES-NaOH, pH 6.0.

SEC Column: Pharmacia (320 mL) HiLoad 26/60 Superdex 75
Buffer: 50 mM MES-NaOH, 200 mM NaCl, pH 6.5.

2.1.2.4.1 *cis*P3H Type I Affinity Purification

In an attempt to improve the purity of *cis*P3H Type I produced, alternative purification methods were considered using fusion peptides. Increasing protein sample purity was especially desirable for crystallisation trials; these experiments often require high-purity, homogenous protein solutions to grow protein crystals that diffract X-ray beams and produce useful structural data.^{290–293} Adding a hexahistidine tag to a protein allows it to be purified using an effective immobilised-metal-affinity chromatography (IMAC) method, which makes use of a resin containing immobilised metal ions (Nickel (II) is a common choice).²⁹⁴ The nitrogen lone pairs of histidine side-chains (imidazole rings) in a His₆-tag form dative, covalent bonds with the immobilised metal ions (forming complex ions); the His₆-tagged fusion protein is therefore strongly anchored to the column resin, allowing it to be selectively 'fished' out of a cell lysate solution and whilst the impurities are washed away (Figure 2.12).²⁹⁴

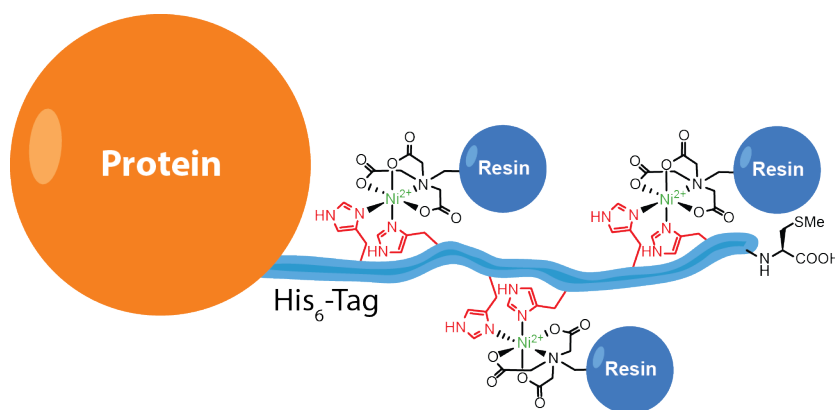


Figure 2.12: The binding of a *N*-terminally hexahistidine-tagged protein to NTA-resin-immobilised Ni²⁺ ions of an IMAC column: histidine residues chelate the Ni²⁺ centres to form octahedral complex ions. Using an IMAC method, fusion proteins are selectively purified from cell extracts.

Proline hydroxylase enzymes, including *cis*P3H Type I, have been effectively produced with an *N*-terminal His₆-tag^{198,207,212,295} and therefore a vector was chosen that encodes for an *N*-terminal fusion peptide sequence.

A pET-based expression vector, pNIC-28-Bsa4 (Figure 2.13),^{296,297} was selected as the protein had already been successfully produced with a Novagen pET-24-a(+) vector;²⁵⁸ this vector also makes use of ligation-independent cloning (LIC), which is a simple method with a high probability of success. In addition to providing antibiotic (kanamycin) selection, the pNIC-28-Bsa4 vector also makes use of the *Bacillus subtilis* *sacB* gene for negative selection of colonies containing unmodified, parent vector pNIC-28-Bsa4 DNA; this massively reduces the number of background colonies on transformation plates containing 5% sucrose.^{296,297} The *sacB* gene encodes for the sucrose-inducible enzyme, levansucrase (EC 2.4.1.10), which catalyses the synthesis of lethal amounts of levan (Scheme 2.2);^{298–302} an accumulation of levan in the periplasm of gram-negative bacteria causes cell lysis.^{303,304}

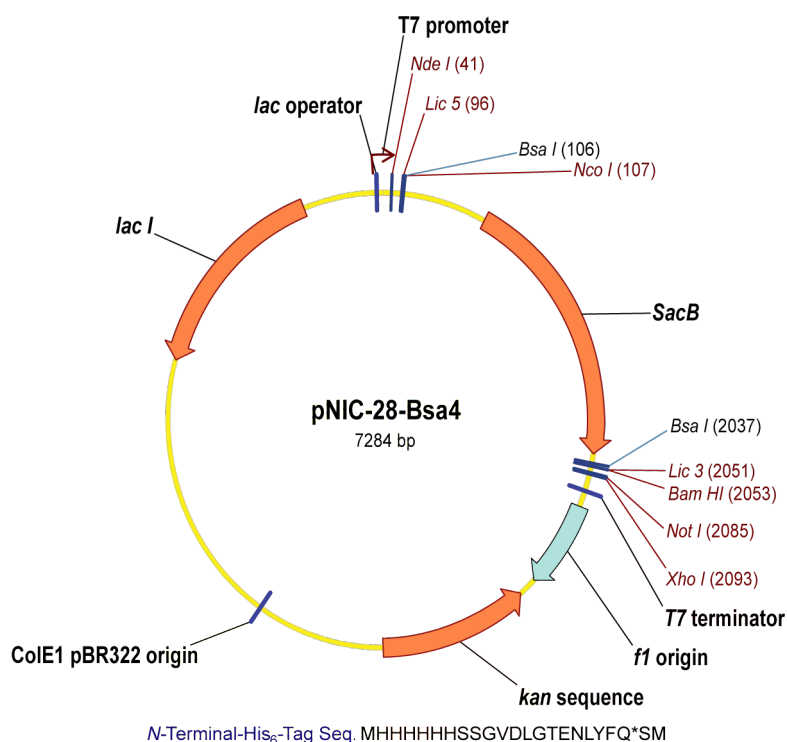
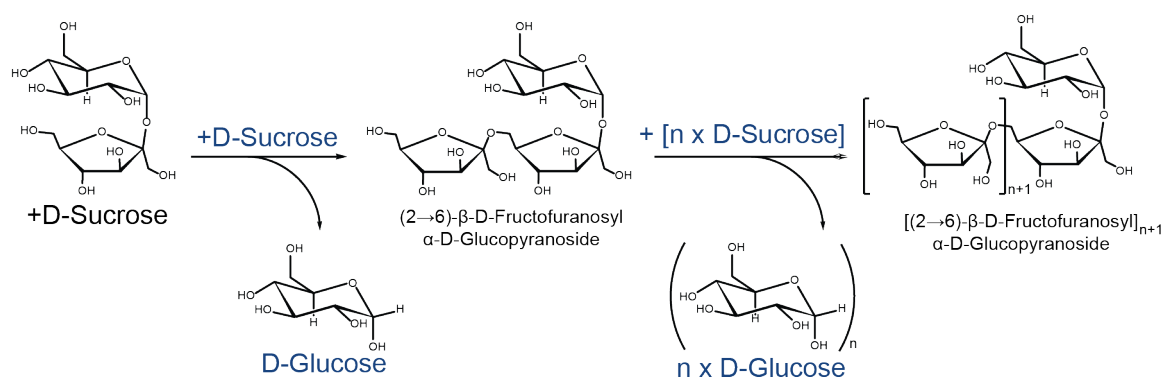


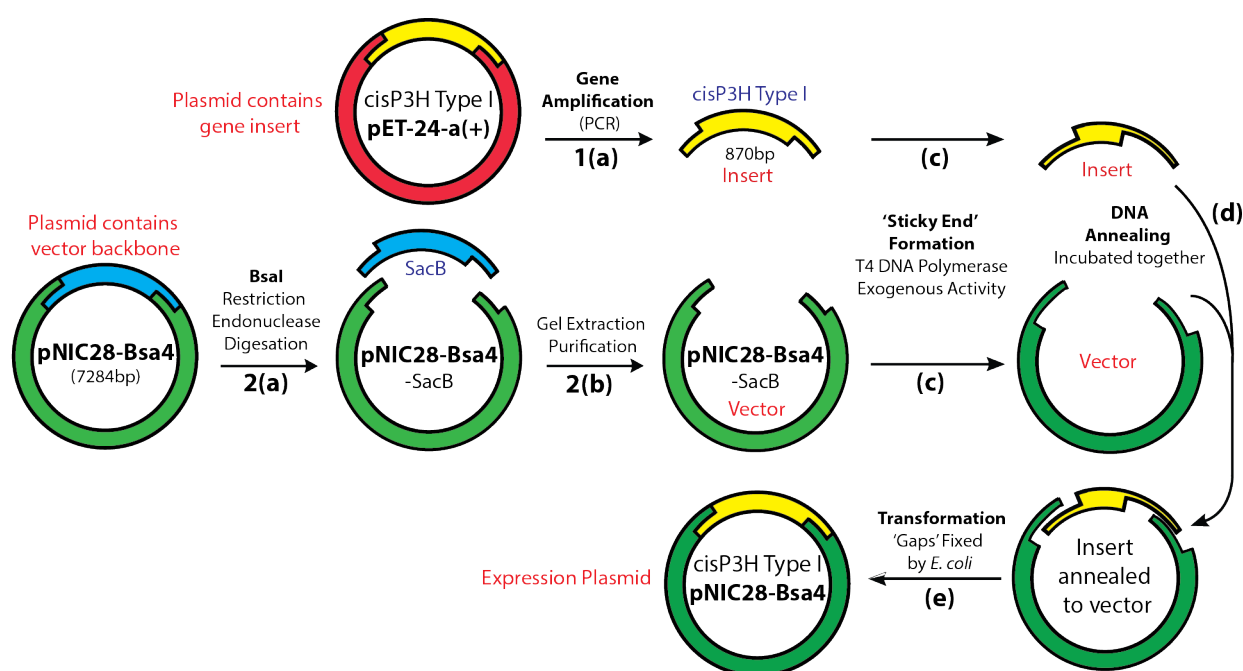
Figure 2.13: The pNIC-28-Bsa4 *E. coli* Expression Vector.²⁹⁶

This vector is a pET-based expression vector [analogous to pET-28a-c(+)] with a sequence coding for a 22-amino-acid, N-terminal His₆-tag fusion peptide, with a TEV-protease cleavage site. It includes sites for LIC cloning, and a 'stuffer' fragment that includes the *sacB* gene, which negatively selects for bacteria containing unaltered pNIC-28-Bsa4 (*i.e.* unsuccessful clones) when they are grown on 5% sucrose.



Scheme 2.2: The levansucrase-catalysed polymerisation of D-sucrose molecules to form levans. In each cycle, a D-glucose leaving group at the 2-position of the sucrose terminus is displaced by a new sucrose unit in a (2→6)-glycosylation reaction; the product is an extended poly-β-fructofuranosyl chain with another D-sucrose unit on the end.

A cloning strategy was devised to insert the full-length *cisP3H* Type I construct into a pNIC-28-Bsa4 vector (Scheme 2.3); pure samples of the *cisP3H* Type I insert and linear pNIC-28-Bsa4 vector were prepared with complementary 3'-sticky ends and combined to anneal the vector and insert fragments together (forming a loosely formed ring). The chain-breaks/gaps in the *cisP3H*-pNIC-28-Bsa4 rings are then fixed by DNA repair mechanisms when these DNA complexes are transformed into *E. coli* competent cells.



Scheme 2.3: An overview of the steps involved in subcloning the full-length *cisP3H* Type I construct into a pNIC-28-Bsa4 expression vector.

The method involves producing pure samples of the *cisP3H* Type I insert [steps 1(a) and (c)] and the empty, linear pNIC-28-Bsa4 vector [steps 2(a), 2(b), and (c)] (both fragments having complementary 3'-sticky ends); the insert and vector can then be combined, so that they selectively anneal to each other sufficiently strong to form a loose DNA ring. These rings can then be transformed into *E. coli* competent cells, which covalently fix the insert and vector fragments together using DNA repair mechanisms. No DNA ligation steps are therefore required [hence, the term 'ligation-independent cloning' (LIC)].

The full-length *cisP3H* Type I gene was first copied and amplified using the polymerase chain reaction (PCR) (Figure 2.14)^{305–307} using the pET-24-a(+) *cisP3H* Type I construct²⁵⁸ as a template. PCR is a cyclical technique which uses DNA polymerase enzymes [*e.g.* *Taq* Polymerase, *Pfu* (Turbo) DNA polymerases] and short oligonucleotide primers to copy specific gene sequences, such that the number of copies of the gene increases exponentially with more completed cycles (Figure 2.14).^{305–307}

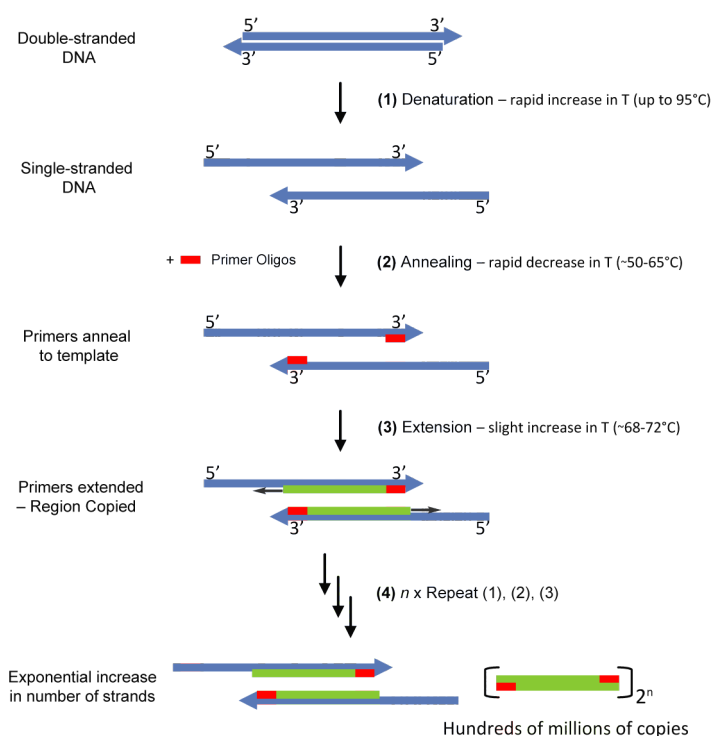


Figure 2.14: The polymerase chain reaction (PCR) cycle for gene amplification:^{305–307} in each cycle of successful PCRs, two DNA strands produce two new strands after each cycle resulting in exponential growth of the population of DNA coding the amplified region.

In the first step (1), a thermal cycler heats the reaction mixture to 95 °C to denature the double-stranded DNA strands to produce single-stranded DNA.

In the second step (2), the mixture is cooled rapidly to a temperature just below the melting temperatures of complementary primer sequences (typically 50 °C - 65 °C), annealing the primers to the single-stranded, template DNA chains at the 5'- and 3'-ends of gene being amplified.

After the primers have annealed to the template DNA, (3) the temperature of the reaction mixture is raised slightly to a temperature optimal for DNA extension by DNA polymerases, promoting the synthesis of the gene being copied. These steps are repeated for several cycles (usually 30-38 cycles) producing several thousands-to-millions of copies of the target gene.

Since the *cisP3H* Type I gene has a high GC content, its melting temperature is predicted to be high (guanosine and cytosine form 3 hydrogen bonds when they base pair), the *cisP3H* gene is likely to undergo secondary structure formation (e.g. hairpins) and mispriming in PCRs, causing premature termination of polymerase gene extension; the primers required for amplification of the *cisP3H* Type I gene were also GC-rich, further complicating matters. To improve the yields of amplification of the *cisP3H* Type I gene by PCR, (1 M) betaine³⁰⁸ and [10% (v/v)] dimethyl sulfoxide (DMSO)³⁰⁹ (Figure 2.15) were added to the PCR mixture; these additives have been shown to improve PCR gene amplification efficiency.^{308, 309}

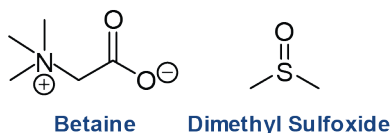


Figure 2.15: The structures of betaine and DMSO.

The subsequent PCR amplification of the *cisP3H* Type I construct was successful; agarose-gel-electrophoresis analysis of the PCR products showed an intense DNA band at roughly 0.9 kb in size, consistent with the proper amplification of the *cisP3H* Type I gene (Figure 2.16). The *cisP3H* Type I gene was then purified using a microcentrifuge spin column.

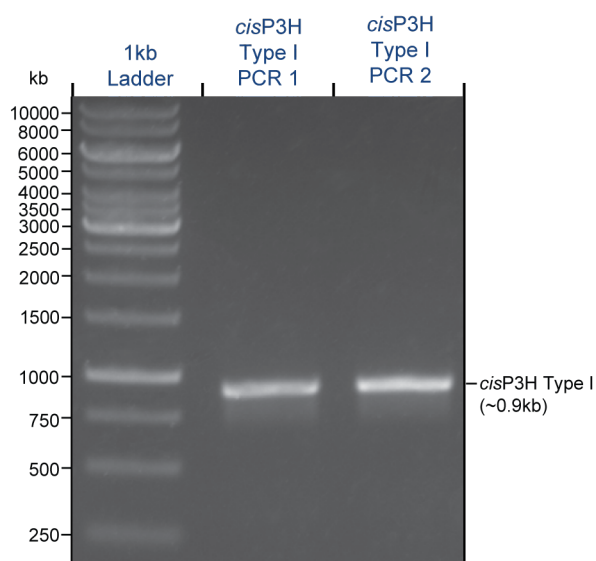


Figure 2.16: Visualisation of the *cisP3H* Type I PCR product DNA: 5 μ L of the reaction mixture was run on a 0.8% (w/v) agarose-TAE gel containing SYBR Safe DNA stain. The gel was run with a potential difference (p.d.) of $\sim$ 100 V.

To enable the pNIC-28-Bsa4 vector to accommodate the *cisP3H* Type I gene insert, the pNIC-28-Bsa4 vector was digested with the endonuclease, BsaI, in order to linearise the pNIC-28-Bsa4 plasmid and remove the *sacB* gene (a counter-selection marker^{303,304}) from it. The products of the BsaI restriction digestion were then analysed by Agarose-gel electrophoresis and purified by gel extraction (Figure 2.17).

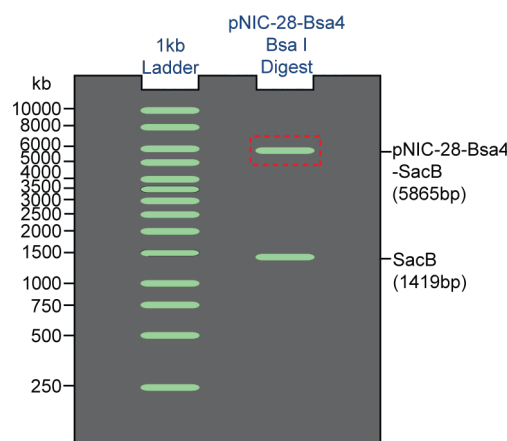


Figure 2.17: Depiction of the agarose-gel-electrophoresis analysis of the pNIC-28-Bsa4 BsaI-digestion products and the gel extraction of the bands corresponding to linear, digested pNIC-28-Bsa4 (-*sacB*). The complete reaction mixture was run on a 0.8% (w/v) agarose-TAE gel containing SYBR Safe DNA stain; the DNA bands were visualised with a UV-light source for a limited time period (to protect the DNA from damage) and the linear pNIC-28-Bsa4 bands were quickly excised from the gel. The linear pNIC-28-Bsa4 was then purified from the TAE-agarose using a commercial kit using microcentrifuge spin columns.

Both the *cisP3H* Type I insert and the BsaI-digested pNIC-28-Bsa4 vector were treated with T4 DNA Polymerase in the presence of dCTP and dGTP, respectively [step (c) of Scheme 2.3]. This step exploits the T4 Polymerase's high 3'-to-5'-exonuclease activity in the absence of all 4 dNTPs,³¹⁰ producing long, overhanging sections of DNA ('sticky ends'). The insert and vector were then combined in a 1:1 ratio and the two DNA fragments annealed to each other sufficiently strong enough to be transformed into *E. coli* competent cells; BL21-Gold (DE3) cells were used for their high transformation efficiency^{275,276} and their high recombinant protein expression yields^{275,276} (*i.e.* the new *cisP3H*-containing plasmid could be transformed directly into an *E. coli* expression host without having to use cloning-specific competent cells first). The colonies from the transformation were screened for successful *cisP3H* Type I clones by carrying out PCR reactions using template plasmids isolated from colony growths; PCRs using templates containing a *cisP3H* gene would produce DNA fragments roughly 0.9 kb in size. Agarose-gel-electrophoresis experiments showed that PCRs using templates from 5 colonies produced DNA fragments

corresponding to *cisP3H* Type I genes (~0.9 kb) (Figure 2.18), consistent with those plasmids containing a to *cisP3H* Type I insert. Colony 1 was sent for sequencing, which confirmed that colony 1 contained a plasmid with a correct *cisP3H* Type I sequence.



Figure 2.18: Agarose-gel-electrophoresis visualisation of products from PCRs using colony plasmids as templates; templates from 5 of the colonies screened contained 0.9 kb DNA fragments, corresponding to *cisP3H* Type I inserts. 5 μ L of sample was loaded into each well and the gel was run using a potential difference of 100 V.

The expression of *N*-terminally hexahistidine-tagged *cisP3H* Type I (*N*-His₆-*cisP3H* Type I) was tested for fresh transformants carrying a *cisP3H* Type I pNIC-28-Bsa4 expression plasmids; BL21-Gold (DE3) *E. coli* cells were grown using both ZYM-5052 auto-induction medium and TB medium (cells being induced for *cisP3H* Type I expression with 0.5 mM IPTG in the case of TB). A typical nickel-affinity purification was carried out using buffers containing an abundance of glycerol [20% (v/v) and 10% (v/v)], salt (500 mM NaCl), and a small amount of detergent (in the lysis buffer only) to promote effective solubilisation of *N*-His₆-*cisP3H* Type I and prevent aggregation/precipitation (Figure 2.19); these IMAC conditions mirrored those used before for proline hydroxylases (see Section 2.2.2) and fitted within the criteria recommended by the IMAC column (GE Healthcare 5 mL HisTrap FF) manufac-

turer. Unfortunately, the yields of *cis*P3H Type I were poor using both TB (with 0.5 mM IPTG induction) and ZYM-5052 auto-induction medium. SDS-PAGE analysis of the nickel-affinity *N*-His₆-*cis*P3H purification fractions showed that cells grown on ZYM-5052 produced a small amount (~10 mg spin-concentrated to ~35 mg/mL) of purifiable *N*-His₆-*cis*P3H Type I (Figure 2.19), but the cells grown on TB (with 0.5 mM IPTG induction) produced almost no *N*-His₆-*cis*P3H (*n.b.* SDS-PAGE gel not shown - bands were too faint to scan). A method using the pNIC-28-Bsa4 was deemed ineffective and other methods were considered.

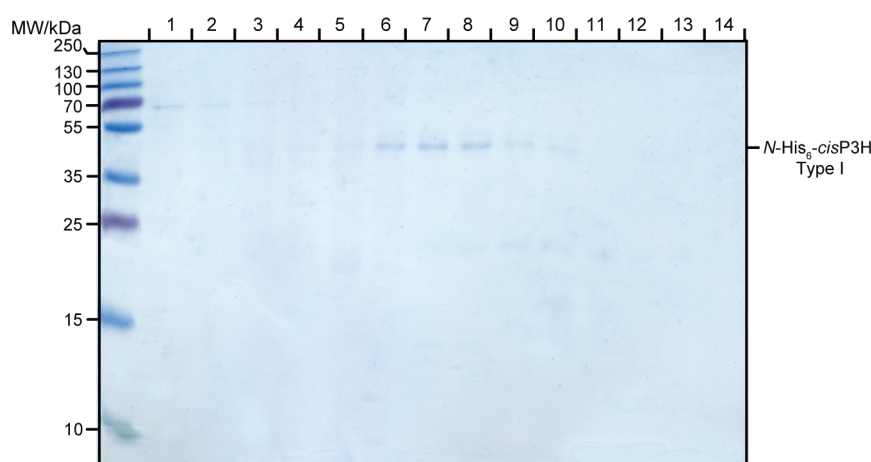


Figure 2.19: SDS-PAGE analysis of the fractions from the nickel-affinity purification of *N*-His₆-*cis*P3H Type I produced by *E. coli* cells grown on ZYM-5052 auto-induction medium.

5 μ L samples of each purification fraction were run on a 15%(w/v) acrylamide SDS-PAGE gel (run with a p.d. of 220 V). The marker denotes the band corresponding to *N*-His₆-*cis*P3H Type I.

2.2 *cis*P4H and *trans*P4H Production

Full-length constructs coding for *Sinorhizobium meliloti cis*P4H and *Dactylosporangium sp.* (strain RH1) *trans*P4H were designed with appropriate 5'- and 3'-restriction cloning sites (*Nde*I and *Bam*H, respectively); the codon-optimised genes (for *E. coli* hosts) were then commercially synthesised using the GeneArt Gene Synthesis service (Life Technologies, Inc.).

2.2.1 Takara pCOLD I Vector Cloning

Some recombinantly-produced proline hydroxylases have been reported to be largely present in the insoluble (*i.e.* inactive) fraction of *E. coli* cell lysates;¹⁹⁸ an *E. coli* expression vector was therefore chosen to maximise the proportion of active, soluble enzyme produced.

Takara's 'pCOLD' series of expression vectors offer 'cold shock' expression technology for high yield, high purity protein production.³¹¹ These pCOLD vectors contain the cold shock protein A (*cspA*) promoter, which selectively induces high, target-protein synthesis (60% intracellular protein) at low temperatures (~15 °C - 'cold shock' conditions);³¹¹ overexpression of proteins in *E. coli* at low temperature also improves their solubility and stability,^{312,313} as well as both suppressing the expression of host proteins and decreasing protease activity (reducing target degradation).^{314,315} Immediately upstream of the *cspA* transcription initiation site, the pCOLD vectors also contain a *lac* operon for controlled expression of the target protein;³¹¹ constitutive expression of the *lacI* gene prevents leaky expression of the cloned genes at 37 °C.³¹¹ Cold-shock induction of gene expression in pCOLD vectors is carried out by simultaneous addition of isopropyl- β -thiogalactopyranoside (IPTG) upon a downshift in temperature (37 °C $\rightarrow$ ~15 °C).³¹¹ In addition, four of the pCOLD vectors also contain a five-codon sequence, 5'-ATGAATCACAAAGTG-3' (amino acid sequence - MNHKV), directly after the *cspA* 5'-UTR (Untranslated Region), which has been shown to enhance translation initiations;³¹⁶ consequently, this sequence is known as 'translation enhancing element' (TEE).³¹⁶

The pCOLD I vector (Figure 2.20) was selected as the expression plasmid for the proline hydroxylases, as it contains an *N*-terminal hexahistidine-tag (*N*-His₆-tag) sequence with a factor Xa cleavage site; the soluble *N*-His₆-proline hydroxylase proteins produced could then be initially purified using effective IMAC methods.

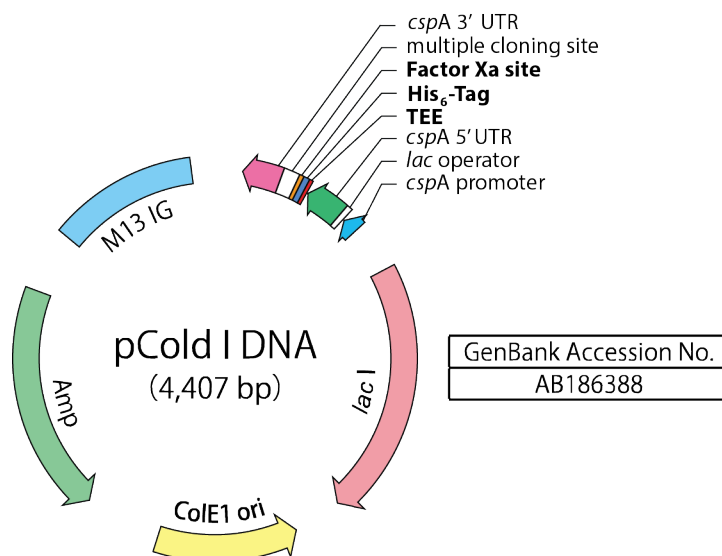
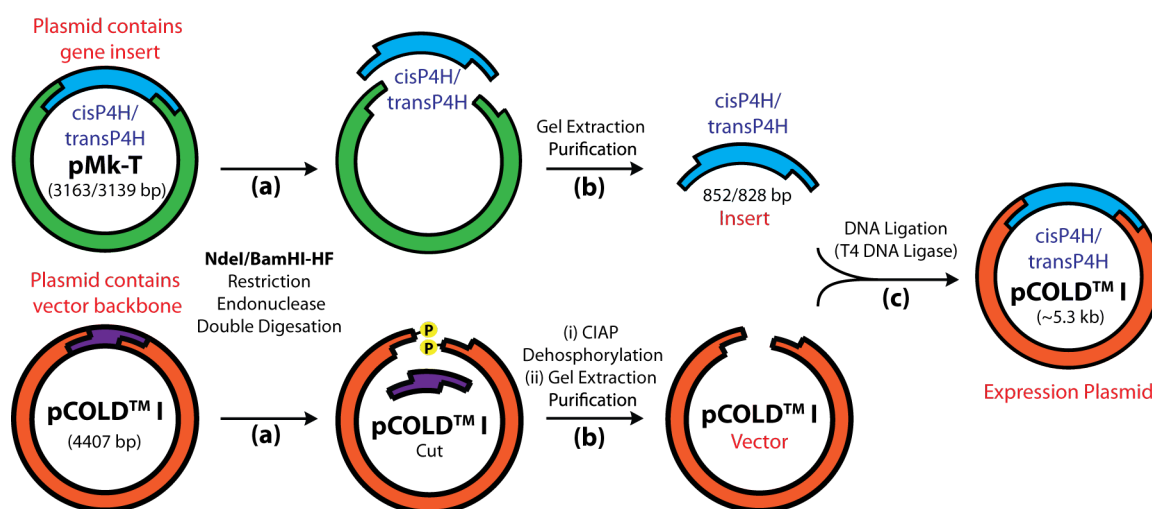


Figure 2.20: The Takara pCOLD I *E. coli* cold shock expression vector;³¹¹ the pCOLD I vector contains an *N*-terminal His₆-tag sequence, a *cspA* promoter, and an ampicillin selection marker. Target protein expression is induced at low temperature (~15 °C) with 0.5 mM IPTG.

A subcloning strategy was designed to insert the *cisP4H* and *transP4H* genes into a pCOLD I vector (Scheme 2.4). The method, first, involves isolating linear fragments of *cisP4H* and *transP4H* DNA from circular plasmids using restriction endonucleases to cut these genes out pMk-T vectors and then purifying the fragments; linear pCOLD I DNA obtained using a similar BamHI-HF/NdeI-double digestion method.



Scheme 2.4: The strategy used for subcloning *cisP4H* and *transP4H* gene into a Takara pCOLD I vector.

(a) The pMk-T plasmids containing *cisP4H* and *transP4H* genes are treated with both HindIII and NdeI restriction endonucleases to cut the proline hydroxylase inserts out of the plasmids; linear pCOLD I vector DNA are similarly prepared using a HindIII-NdeI double reaction.

(b) The quenched, pCOLD I restriction digestion products are then treated with Calf Intestinal Alkaline Phosphatase (CIAP) to dephosphorylate the 5'-ends of the pCOLD I DNA strands and then both the linear pCOLD I vector and proline hydroxylase gene inserts are gel purified; dephosphorylating the pCOLD I vector prevents self-ligation of the linear vector fragments (*i.e.* vector + vector) during the ligation step (reducing background colonies).

(c) The purified *cisP4H/transP4H* inserts and the pCOLD I vector DNA are then combined in ratio of 3:1 and ligated together in a reaction catalysed by T4 DNA ligase; the newly formed plasmids are then transformed into BL21-Gold (DE3) *E. coli* competent cells and plated on ampicillin plates for antibiotic selection of the successful clones.

Firstly, the circular pCOLD I vector was cut open using NdeI and HindIII restriction endonucleases, forming linear pCOLD I DNA (Figure 2.21). The linear pCOLD I was then treated with Calf Intestinal Alkaline Phosphatase (CIAP) to dephosphorylate the 5'-ends of the DNA fragments to prevent self-ligation of the pCOLD I vector strands (*i.e.* prevents vector-vector adducts) when the *cisP4H/transP4H* inserts were later ligated to the pCOLD I vector strands; this step therefore reduces the number of background colonies when the ligation mixture is transformed into *E. coli* competent cells. The *cisP4H* and *transP4H* genes were also 'cut out' of the pMk-T plasmid using NdeI/HindIII double restriction digestion reactions. The *cisP4H*

and *transP4H* gene inserts and the linear pCOLD I vector DNA fragments were run on a TAE-agarose gel; the bands corresponding to these 3 genes were then rapidly excised (to limit damage from UV-light exposure) with a clean razor blade and gel purified using a commercial kit (Figure 2.22).

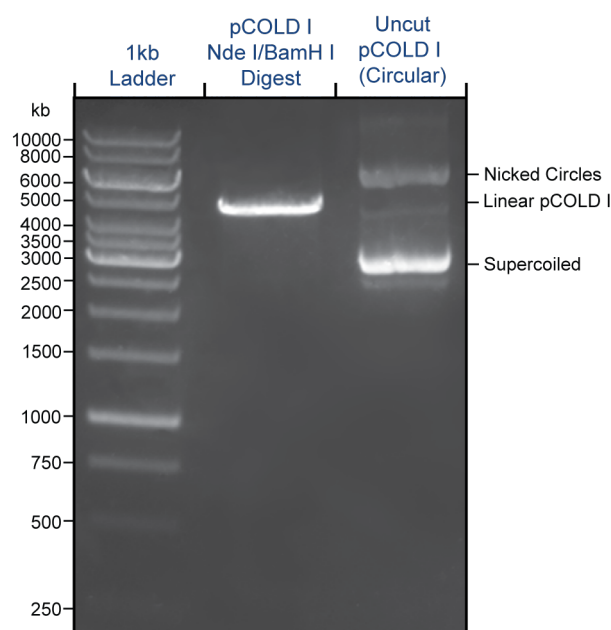


Figure 2.21: Agarose-gel-electrophoresis visualisation of the pCOLD I NdeI/BamHI double restriction digestion products.

5 μ L samples were loaded on a 0.8% (w/v) agarose-TAE gel (run with a p.d. of 100 V). Linear pCOLD I fragments (produced by complete restriction digestion) run at $\sim$ 4.4 kb, whereas circular plasmid runs at either a slightly smaller size ($\sim$ 3 kb - supercoiled) or a slightly larger size ($\sim$ 6 kb - nicked circles).

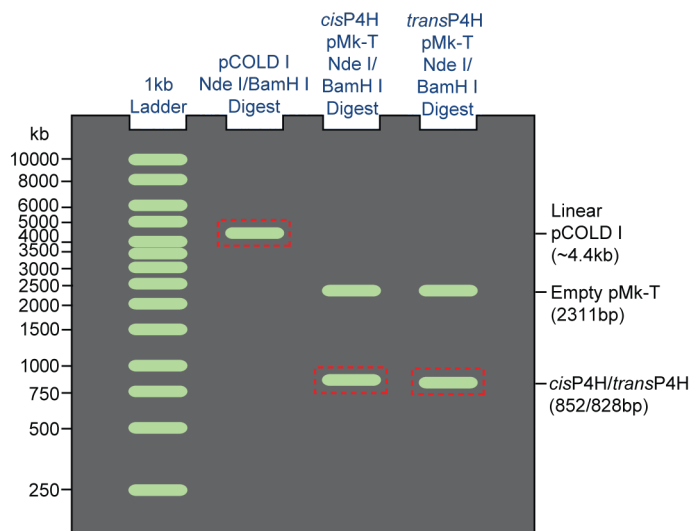


Figure 2.22: The gel purification of the linear pCOLD I vector and the *cisP4H* and *transP4H* gene fragments.

The NdeI/BamHI double restriction digestion products were run on an 0.8% (w/v) agarose-TAE gel containing SYBR Safe DNA stain; the DNA bands were visualised with UV-light and excised with a clean razor blade and purified using a commercial gel purification kit.

The purified linear pCOLD I vector was combined with the *cisP4H* and *transP4H* inserts in a 1:3 ratio and a ligation reaction carried out using T4 DNA ligase; successful reactions produced clones comprising a pCOLD I vector containing a correct sequences for the *cisP4H* and *transP4H* genes. Transformation of the DNA ligation mixture products into BL21-Gold (DE3) *E. coli* competent cells produced numerous colonies, which were then screened for pCOLD I vectors with correct *cisP4H/transP4H* inserts. Miniprep-purified plasmids from colony growths were restriction-digested with NdeI and BamHI enzymes and the products were analysed by Agarose-gel electrophoresis (Figure 2.23); colonies containing pCOLD I plasmids with successfully cloned *cisP4H* and *transP4H* gave bands at ~0.85 kb in the restriction digests (corresponding to the *cisP4H/transP4H* - colonies 2 and 5 (*cisP4H*) and colonies 1, 2, and 5 (*transP4H*))(Figure 2.23). Plasmids from colonies 2 and 5 (*cisP4H*) and colonies 1 and 2 (*transP4H*) were sent for sequencing and the sequencing confirmed that these colonies did, in fact, contain correct genes coding

for *cis*P4H and *trans*P4H.

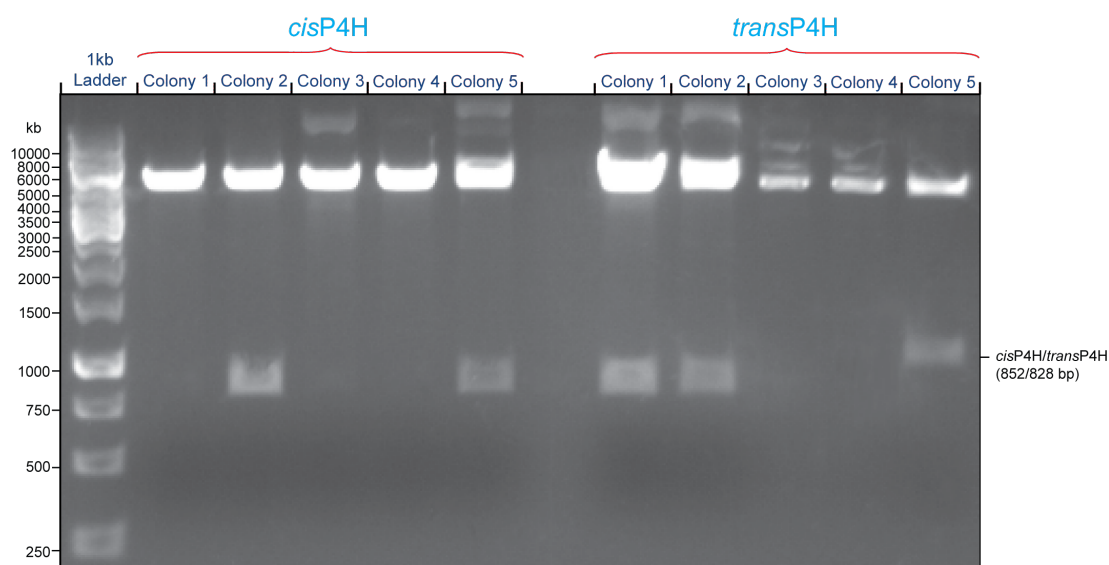


Figure 2.23: Agarose-gel-electrophoresis analysis of NdeI/HindIII-digested colony plasmids; the plasmids were originally produced in experiments where the *cis*P4H and *trans*P4H genes were ligated into the pCOLD I vector.

Plasmids (pCOLD I) from colonies containing successfully cloned *cis*P4H and *trans*P4H gene inserts gave a band at ~0.85 kb in analytical restriction digests; colonies 2 and 5 (*cis*P4H) and colonies 1, 2, and 5 (*trans*P4H) appeared to contain the correct genes. 5 μ L samples were loaded on a 0.8% (w/v) agarose-TAE gel (run with a p.d. of 100 V).

2.2.2 *cis*P4H and *trans*P4H Nickel-Affinity Purification

Terrific broth (TB) (batches of 4 $\times$ 1 L) cultures were inoculated with overnight 2-YT starter cultures (10 mL per 1 L TB); fresh transformants were used to seed the starter cultures [BL21-Gold (DE3) *E. coli* competent cells with *cis*P4H/*trans*P4H pCOLD I plasmids]. The shaking (180 rpm) cell cultures were grown for ~4 h at 37 $^{\circ}$ C, then cooled to 15 $^{\circ}$ C, induced for target protein expression with 0.5 mM IPTG, and incubated for a further ~24 h. Cells were harvest by centrifugation and frozen at -80 $^{\circ}$ C.

An immobilised-nickel-affinity chromatography method was developed for purifying hexahistidine-tagged *cis*P4H and *trans*P4H; a 5 mL GE Healthcare HisTrap FF column was selected for it's high protein-binding affinity and low nickel leak-

age. Buffers conditions were selected which would promote effective solubilisation of the proline hydroxylase enzymes during cell lysis and the initial purification steps; the buffers relatively high salt (500 mM NaCl), supplementary glycerol [20% (v/v) and 10% (v/v) for the cell lysis and wash/elution buffers, respectively], and small amount [0.1% (w/v)] Tween-20 detergent during cell lysis. The buffer pH and imidazole concentrations were based in part on criteria recommended by the His-column manufacturer; HEPES-NaOH-based buffers (50 mM) were selected for their effective and temperature-resistant buffering capacity, being prepared from a 1 M stock solution of HEPES-NaOH, pH 7.5 (at 25 °). Sonication was used to break resuspended cells [in cooled (4 °C) cell lysis buffer]; the column-binding buffer contained 5 mM imidazole to help prevent non-specific binding of contaminants to the nickel-sepharose 6 column resin when the *cis*P4H and *trans*P4H extract solutions were loaded onto the HisTrap FF column, and improve *cis*P4H/*trans*P4H purity. After all the cell lysate solution had been loaded onto the His-column, the column was washed with an excess of binding buffer until the UV_{280nm}-absorbance trace returned to baseline levels (removing the excess lysate solution). The HisTrap column was then treated with wash buffer containing a slightly higher imidazole concentration (30 mM) than the column binding buffer until the UV_{280nm}-absorbance chromatogram returned to baseline levels; this steps removes more strongly held impurities from the His-column resin. The 5 mL HisTrap FF column was then treated with a linear gradient of increasing imidazole concentration (30 mM-400 mM) eluting the strongly bound *cis*P4H and *trans*P4H target proteins from the column (Figures 2.27 and 2.25). Both purifications were deemed to produce *cis*P4H/*trans*P4H at a purity of >85%; the yields were estimated to be ~110 mg *cis*P4H and ~90 mg *trans*P4H from the 4 × 1 L TB cultures. The purest fractions (2-16 for *cis*P4H and 2-12 for *trans*P4H) were pooled, spin-concentrated, and buffer exchanged (into 10 mM Tris-HCl, 150 mM NaCl, pH 7.5); a Tris-HCl-based buffer was used because of

its very simple ^1H -NMR spectrum (a sharp singlet) which would not overlap and obscure important peaks in NMR assay reaction. In contrast, HEPES-NaOH has a complicated, multi-peak spectrum, which overlaps with interesting product peaks. The buffer-exchanged *cis*P4H and *trans*P4H were spin concentrated to (~ 71 mg/mL and ~ 25 mg/mL, respectively, and flash frozen at -80°C for assays.

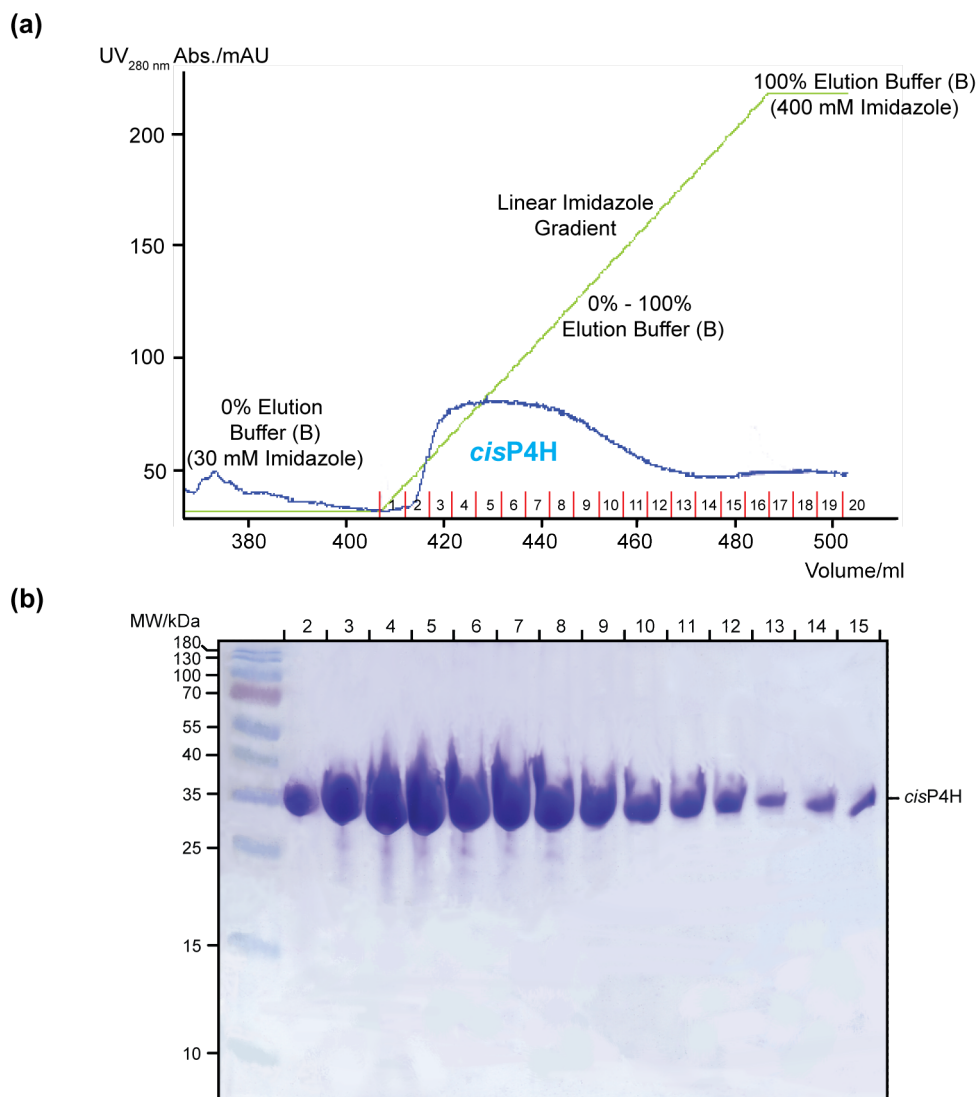


Figure 2.24: Immobilised-nickel-affinity chromatography of *cisP4H*.

After all of the cell lysate had been loaded onto the column, the column was washed with excess binding buffer and wash buffer until the UV_{280nm} absorbance trace reached baseline levels. The column was then treated with (a) a linear gradient of increasing imidazole (30 mM-400 mM) to elute the *N*-His₆-*cisP4H*, which was manifested as a large, broad peak in the UV_{280nm} -absorbance chromatogram. (b) The elution fractions (5 μ L samples) were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] and found contain an abundance of *cisP4H*.

Column: 5 mL HisTrap FF

Binding Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 5 mM Imidazole, 20% (v/v) glycerol, 0.1% Tween-20, 0.5 mM TCEP, pH 7.5.

Wash Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 30 mM Imidazole, 10% (v/v) glycerol, pH 7.5.

Elution Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 400 mM Imidazole, 10% (v/v) glycerol, pH 7.5.

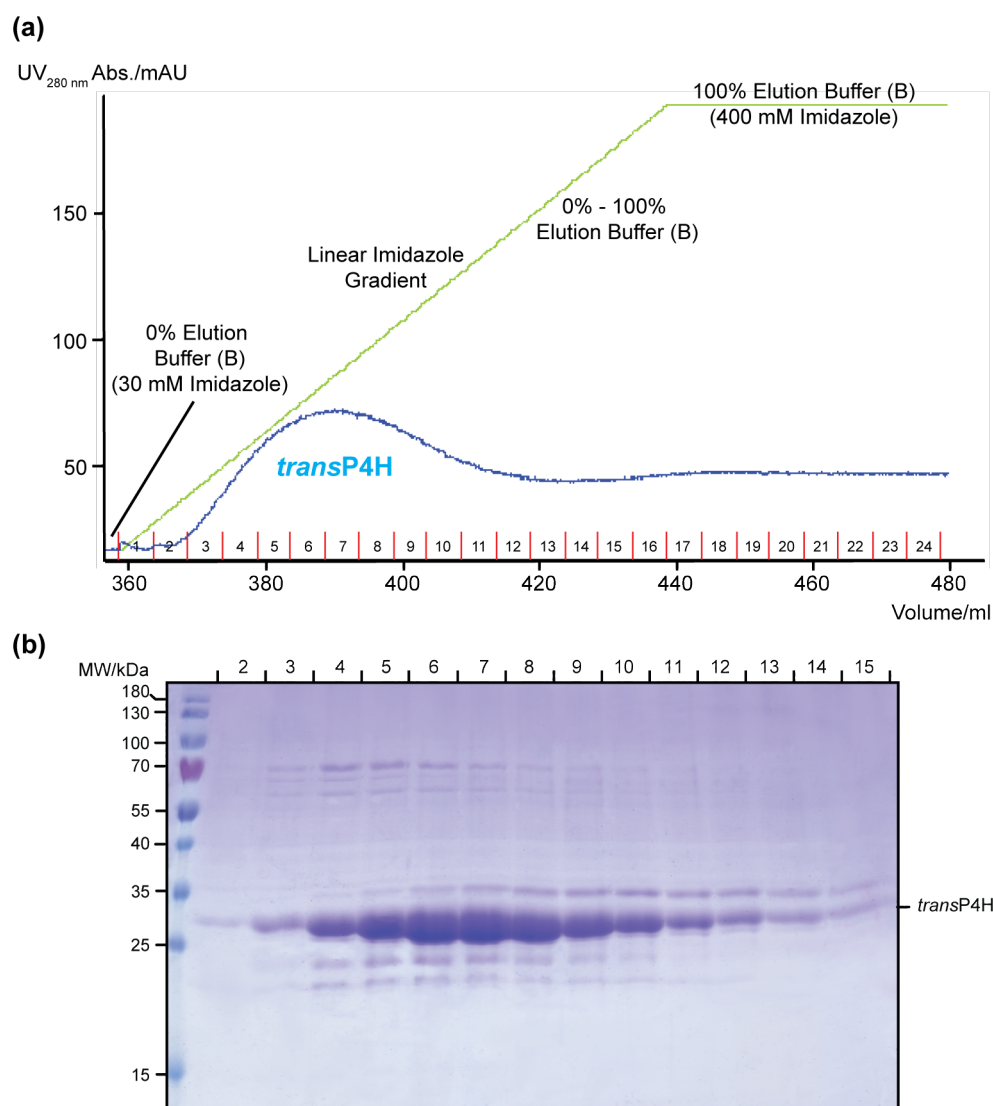


Figure 2.25: Immobilised-nickel-affinity chromatography of *transP4H*.

After all of the cell lysate had been loaded, the column was washed with excess binding buffer and wash buffer (until the UV_{280nm}-absorbance trace reached baseline levels). The column was then treated with (a) a linear gradient of increasing imidazole (30 mM-400 mM) to elute the *N*-His₆-*transP4H*, which was manifested as a large, broad peak in the UV_{280nm}-absorbance chromatogram. (b) The elution fractions (5 μ L samples) were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] and found contain an abundance of *transP4H*.

Column: 5 mL HisTrap FF

Binding Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 5 mM Imidazole, 20% (v/v) glycerol, 0.1% Tween-20, 0.5 mM TCEP, pH 7.5.

Wash Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 30 mM Imidazole, 10% (v/v) glycerol, pH 7.5.

Elution Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 400 mM Imidazole, 10% (v/v) glycerol, pH 7.5.

2.2.2.1 *cis*P4H Purification Development

In order to enhance the purity of the *cis*P4H produced, extra purification steps were explored on top of the initial IMAC step; one motivation for having very pure *cis*P4H was to try and increase the probability of *cis*P4H crystallisation.

A size-exclusion chromatography step using a Pharmacia (320 mL) HiLoad 26/60 Superdex 75 was added after the initial His-column step in the *cis*P4H purification (from Section 2.2.2); this gel filtration method improved the purity of the *cis*P4H, eluting from the column with a sharp peak (Figure 2.26 a). SDS-PAGE analysis revealed the protein was >95% pure (Figure 2.26 b). The purest fractions (A8-B12) were pooled and concentrated (~90 mg at ~37 mg/mL). Another batch of *cis*P4H was purified using the same, two-step protocol, except that, in this batch, *cis*P4H was treated with Factor Xa overnight (at 4 °C) to cleave the *cis*P4H's His₆-tag before the size-exclusion step. The IMAC purification fractions were carefully buffer-exchanged into Factor Xa cleavage buffer (20mM Tris-HCl, 100 mM NaCl, pH 7.4) using a 10 kDa spin-concentrator; the next morning, the His₆-tag-cleaved *cis*P4H fractions were then concentrated to ~1 mL and loaded on the 320 mL Superdex 75 gel filtration column. The yield of untagged *cis*P4H was roughly the same as His₆-tagged *cis*P4H (~90 mg at ~35 mg/mL). Both tagged and untagged *cis*P4H proteins were produced for crystallisation trials in an attempt to solve crystal structures for *cis*P4H; it was thought that the two *cis*P4H variants may crystallise differently, one being more readily crystallised than the other, and so both were tested to improve the odds of successful *cis*P4H crystallisation.

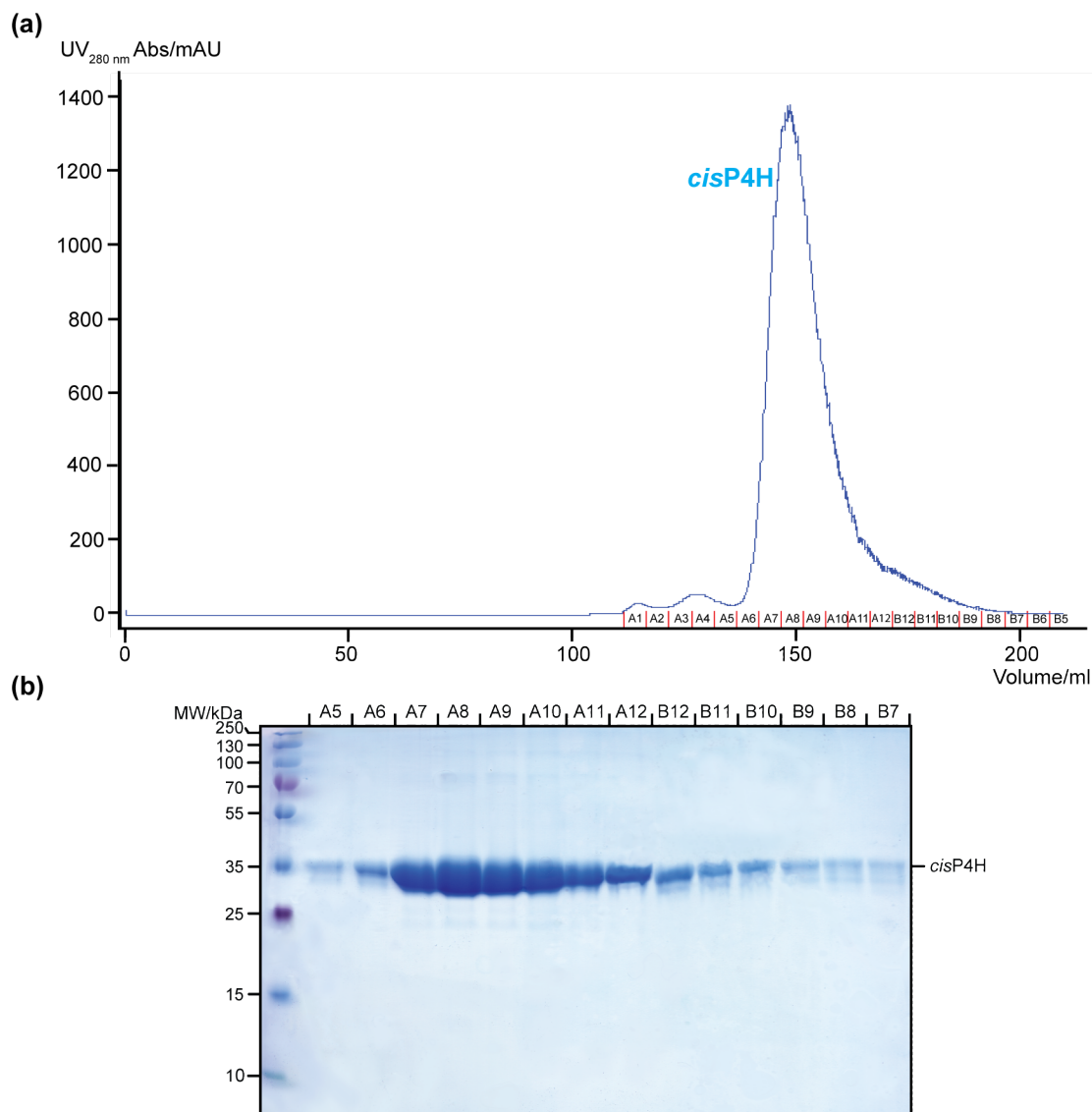


Figure 2.26: Size-exclusion purification of *cisP4H* as a second column-step after IMAC purification:

(a) the UV_{280nm}-absorbance chromatogram for the purification;

(b) and the tris-glycine SDS-PAGE [15% (w/v) acrylamide] analysis of the *cisP4H* purification fractions. The presence of *cisP4H* is indicated by a marker.

Column: (320 mL) Pharmacia HiLoad 26/60 Superdex 200 size-exclusion.

Buffer: 50 mM MES-NaOH, 200 mM NaCl, pH 6.5.

The method for *cisP4H* purification was further altered slightly to see if an extra intermediate Mono Q anion-exchange chromatography step would improve purity significantly; the total purification method consisted of an IMAC step (Figure 2.27), a Mono Q AEC step (Figure 2.28), and a final size-exclusion chromatography step (Figure 2.29); the purest fractions (A9-B12) were spin-concentrated (to ~35 mg/mL) and frozen at -80 °C. The final purity of *cisP4H* (fractions A9-B12) using this extended, 3-column purification method was, at most, only very slightly higher (maybe 1%-2%) than the *cisP4H* (fractions A8-B12) produced using the 2-column (Ni-Sepharose and Superdex 75) method; the purest *cisP4H* fractions of both methods were shown to be >95% by SDS-PAGE analysis (Figures 2.26 b and 2.29 b). The *cisP4H* yield of the method using an extra purification step was slightly reduced (by ~5%) in comparison to the previous, 2-step purification, as one would expect from extra material transfer steps (*i.e.* some *cisP4H* is lost on the surface of each vessel is in). Given that the purity of *cisP4H* produced (>95%) using a simpler, 2-column purification is very similar to that purified using the 3-column method, the 2-column (Ni-Sepharose and Superdex 75) method is probably more suitable for *cisP4H* production because it is faster and it yields slightly more *cisP4H* (extra ~5%).

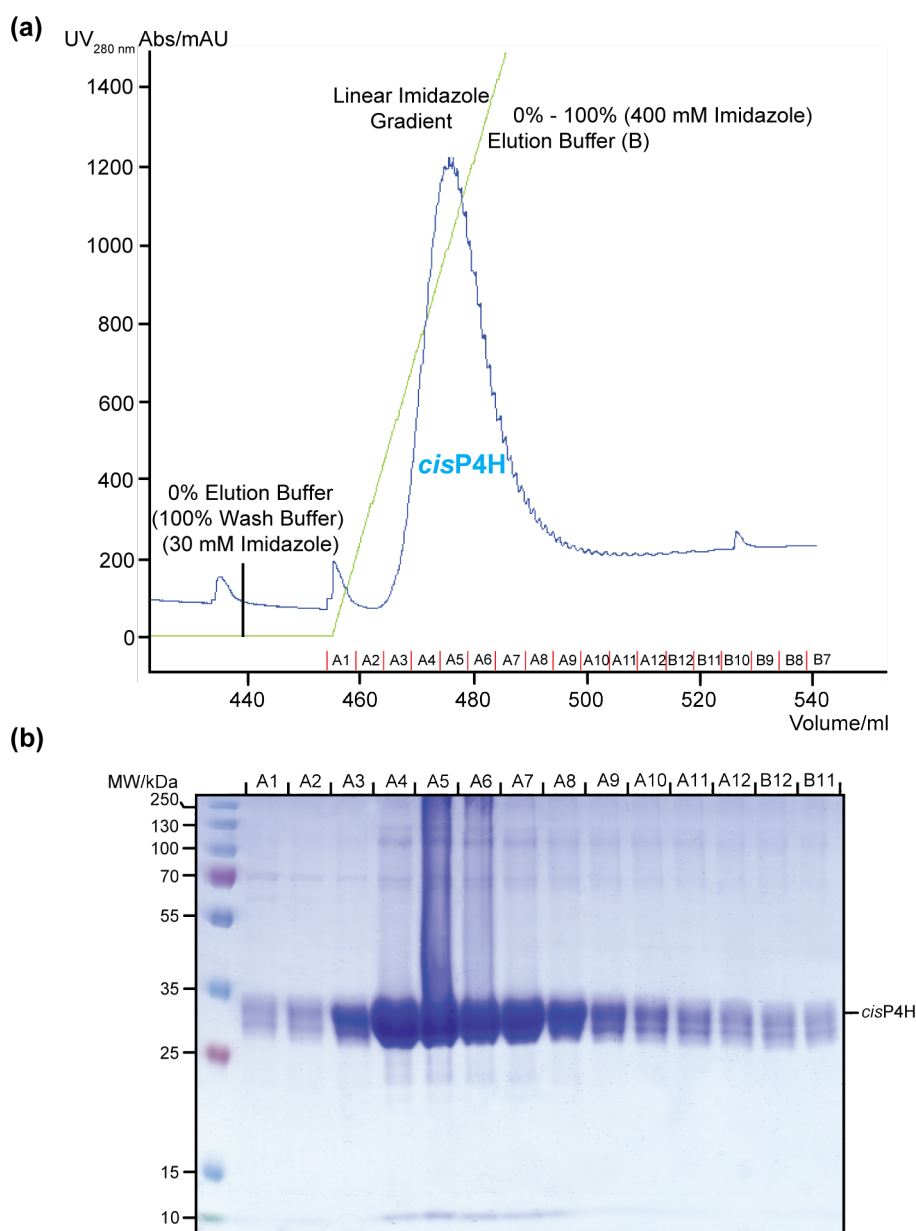


Figure 2.27: Immobilised-nickel-affinity chromatography of *cisP4H*.

After all of the cell lysate had been loaded, the column was washed with excess binding buffer and wash buffer until the UV_{280nm}-absorbance trace reached baseline levels. The column was then treated with (a) a linear gradient of increasing imidazole (30 mM-400 mM) to elute the *N*-His₆-*transP4H*, which was manifested as a large, broad peak in the UV_{280nm}-absorbance chromatogram. (b) The elution fractions (5 μ L samples) were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] and found contain an abundance of *transP4H*.

Column: 5 mL HisTrap FF

Binding Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 5 mM Imidazole, 20% (v/v) glycerol, 0.1% Tween-20, 0.5 mM TCEP, pH 7.5.

Wash Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 30 mM Imidazole, 10% (v/v) glycerol, pH 7.5.

Elution Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 400 mM Imidazole, 10% (v/v) glycerol, pH 7.5.

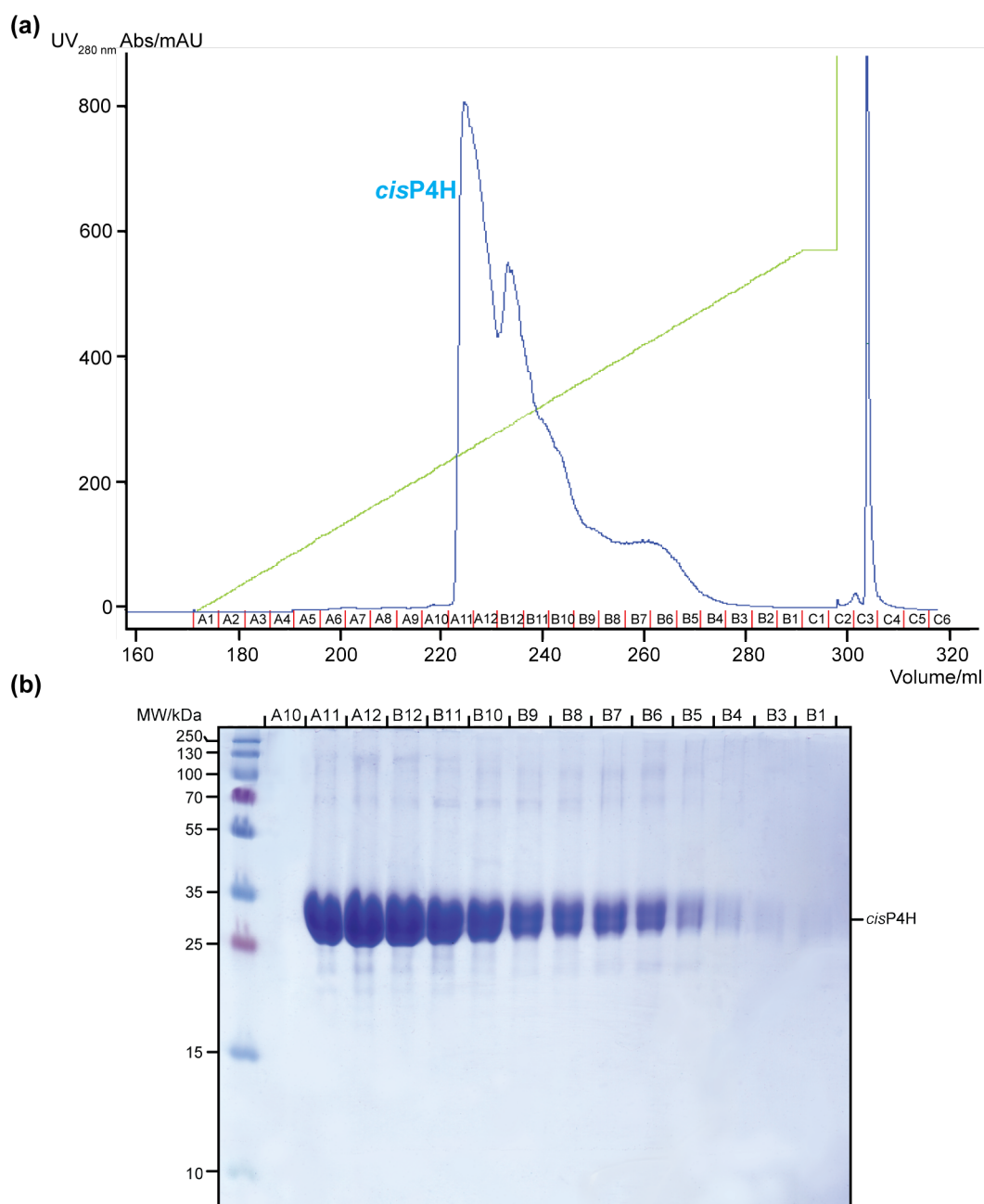


Figure 2.28: Mono Q anion-exchange chromatography of *cisP4H*:

(a) the UV_{280nm}-absorbance FPLC chromatogram for the *cisP4H* gel filtration fractions;

(b) The fractions were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] to assess protein purity. 5 μ L samples of each fraction were run in each lane. The bands ran at the expected position.

Column: Pharmacia (8 mL) Mono Q anion-exchange.

Buffer A: 20 mM Tris-HCl, pH 7.6.

Buffer B: 20 mM Tris-HCl, 1 M NaCl, pH 7.6.

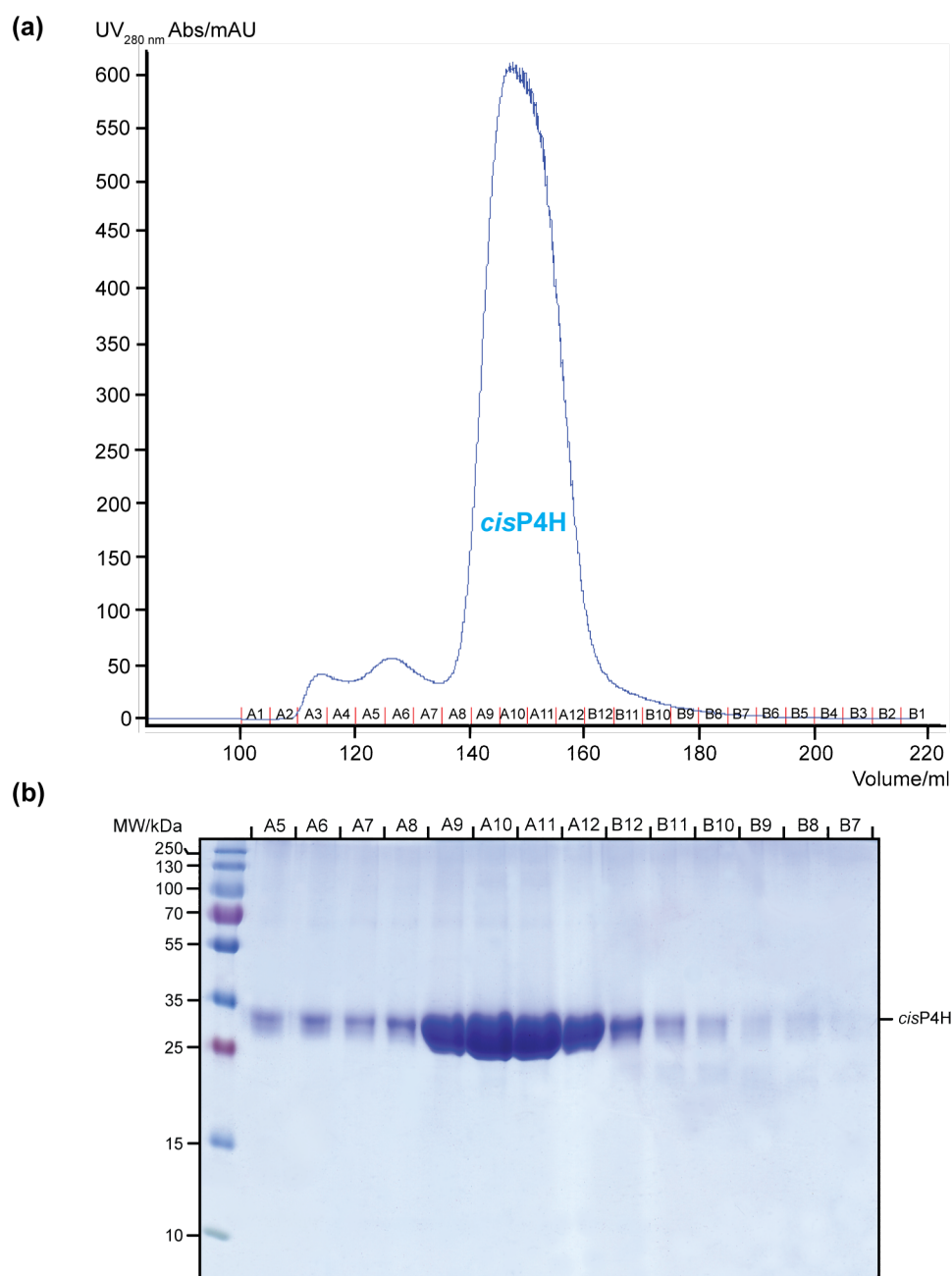


Figure 2.29: Size-exclusion purification of *cisP4H* as a second column-step after IMAC purification:

(a) the UV_{280nm} -absorbance chromatogram for the purification;

(b) and the tris-glycine SDS-PAGE [15% (w/v) acrylamide] analysis of the *cisP4H* purification fractions (5 μ L samples). The presence of *cisP4H* is indicated by a marker.

Column: Pharmacia (320 mL) HiLoad 26/60 Superdex 75

Buffer: 50 mM MES-NaOH, 200 mM NaCl, pH 6.5.

2.2.2.2 *trans*P4H Purification Development

To improve the purity of *trans*P4H produced, extra purification steps were considered to follow the initial nickel-affinity chromatography. Proline hydroxylases have been improved in purity using (30 mL) DEAE Sepharose FF and (320 mL) Superdex 75 columns in polishing steps (see Sections 2.1.2.2, 2.1.2.3, and 2.2.2.1); a purification scheme incorporating these columns was therefore considered to improve *trans*P4H purity. Following the His₆-column purification of *trans*P4H, the *trans*P4H fractions were buffer-exchanged (into 50 mM MES-NaOH, pH 6.5 - 'Buffer A') and then loaded onto a 30 mL Pharmacia DEAE Sepharose FF column; this DEAE Sepharose FF anion-exchange column was then washed with excess buffer A and then the column-bound *trans*P4H was eluted using a linear salt gradient [0%-100% elution buffer (B) - 0 mM-500 mM NaCl]. The *trans*P4H eluted as a sharp peak (Figure 2.30 a), consistent with the *trans*P4H being pure and homogeneous. SDS-PAGE analysis of the *trans*P4H fractions (Figure 2.30 b) confirmed that the *trans*P4H was very pure (>95%) with only a very faint, high-molecular-weight band being present.

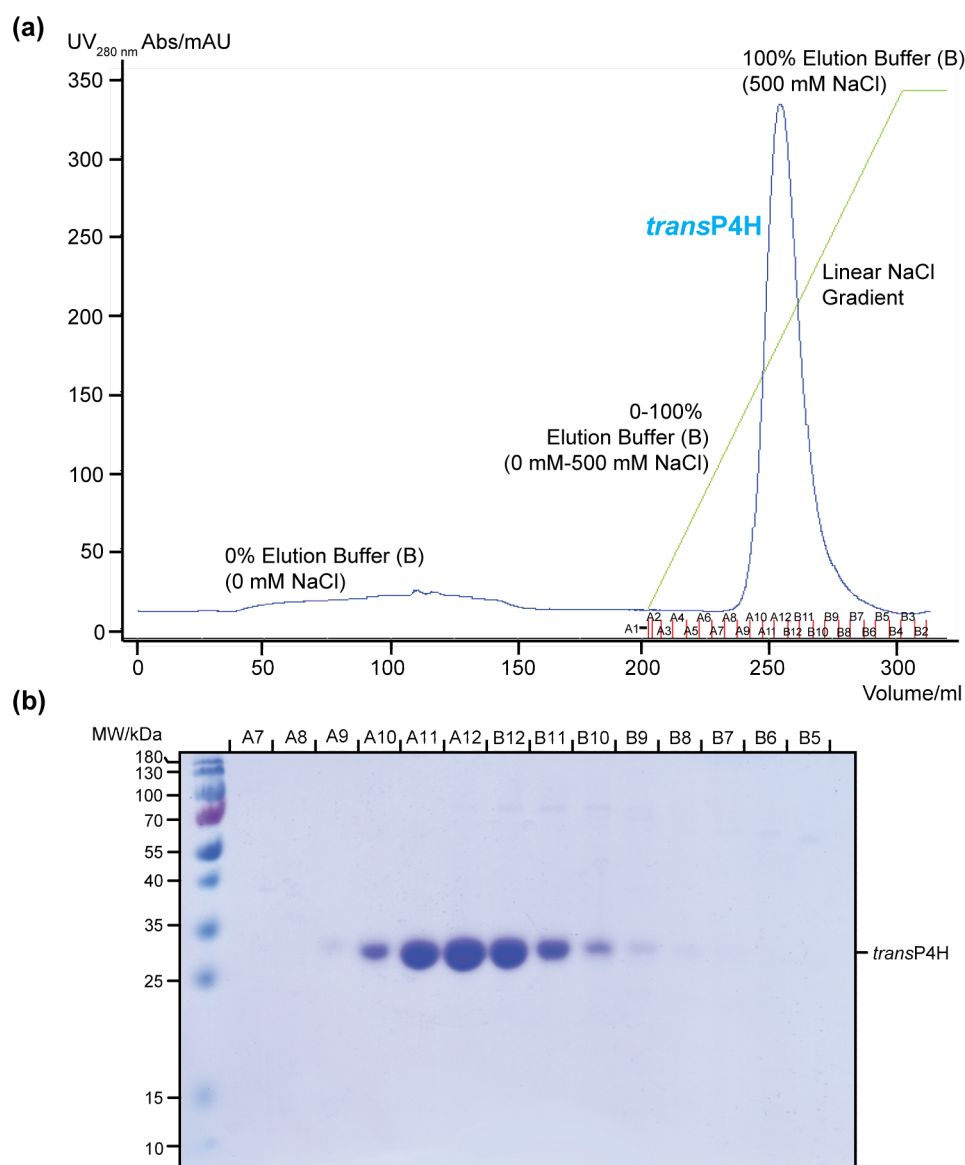


Figure 2.30: DEAE Sepharose FF anion-exchange chromatography of *transP4H*:

- (a) the UV_{280nm}-absorbance FPLC chromatogram for the *transP4H* gel filtration fractions;
- (b) the fractions were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] to assess protein purity. 5 μ L samples of each fraction were run in each lane. The marker denotes the target protein; the bands ran at a mass of the correct order of magnitude.

Column: Pharmacia (30 mL) DEAE Sepharose FF Anion-Exchange

Buffer A: 50 mM MES-NaOH, pH 6.5.

Buffer B: 50 mM MES-NaOH, 500 mM NaCl, pH 6.5.

All of the fractions containing *trans*P4H (A8-B9) from the DEAE Sepharose FF anion-exchange purification step were pooled and spin-concentrated (to ~1 mL); the concentrated *trans*P4H fractions were then loaded onto a Pharmacia (320 mL) HiLoad 26/60 Superdex 75 size-exclusion column and purified further. It was hoped that this size-exclusion chromatography step would remove the very tiny amount of high-molecular-weight contaminant from the *trans*P4H; the *trans*P4H eluted from the size-exclusion column as a sharp peak (Figure 2.31 a) and SDS-PAGE analysis revealed the *trans*P4H was even purer still (Figure 2.31), since the very slight high-MW contaminant eluted before the *trans*P4H (*i.e.* the high-MW contaminant was mostly in the fractions before the *trans*P4H fractions). The purest *trans*P4H fractions (A9-A12) were pooled, spin-concentrated (to ~33 mg/mL), and frozen at -80 °C. The *trans*P4H was very pure (>95%) and homogeneous, and so the 3-column purification method was deemed to be successful.

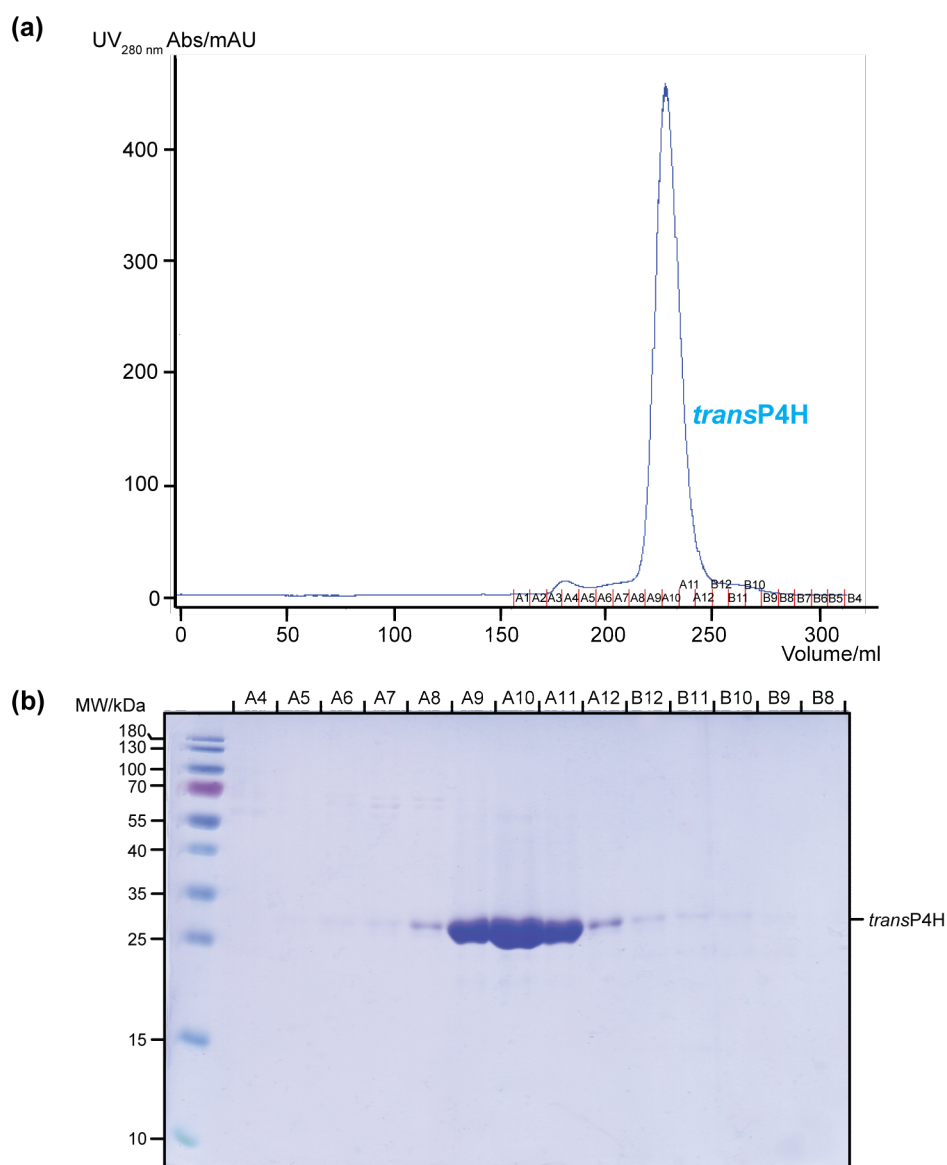


Figure 2.31: Size-exclusion purification of *transP4H* as a second column-step after IMAC purification:

(a) the UV_{280nm}-absorbance chromatogram for the purification;

(b) and the tris-glycine SDS-PAGE [15% (w/v) acrylamide] analysis of the *transP4H* purification fractions (5 μ L samples). The presence of *transP4H* is indicated by a marker.

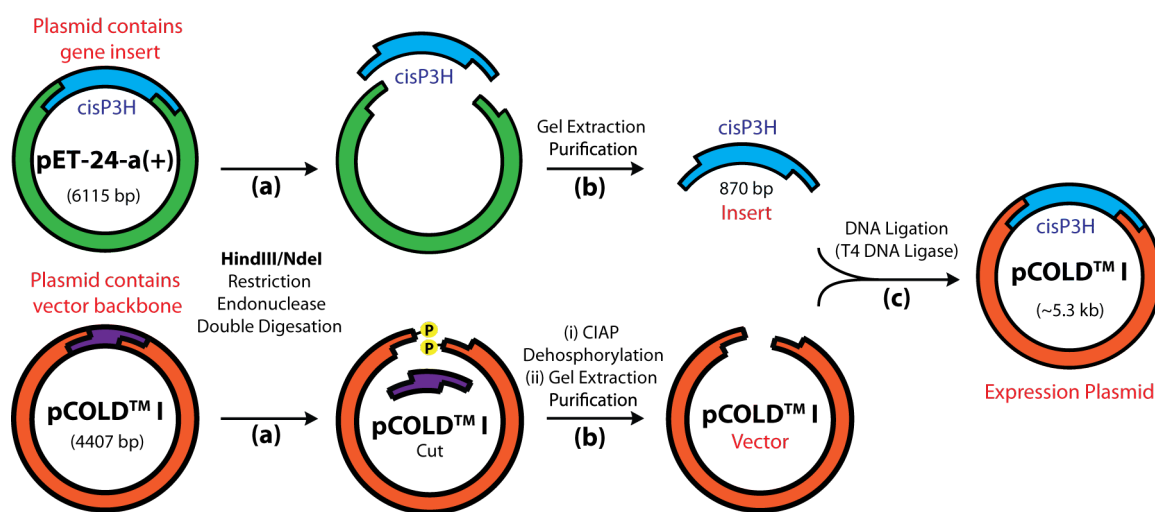
Column: Pharmacia (320 mL) HiLoad 26/60 Superdex 75

Buffer: 50 mM MES-NaOH, 200 mM NaCl, pH 6.5.

2.2.3 Takara pCOLD I *cis*P3H Type I Expression

The pCOLD I vector was effective in producing large quantities of *cis*P4H and *trans*P4H (Section 2.2.2), and it has been reported to be useful for preparing hexahistidine-tagged *cis*P3H Type I;¹⁹⁸ therefore, I explored its use for preparation of *cis*P3H Type I.

The subcloning strategy used to insert the *cis*P3H Type I gene into the pCOLD I vector (Scheme 2.5 was very similar to that devised to subclone the *cis*P4H and *trans*P4H genes into the pCOLD I vector (Scheme 2.4). Firstly, linear samples of pCOLD I vector DNA and *cis*P3H Type I insert were prepared by restriction digestion of plasmids containing those genes [*i.e.* the complete, circular pCOLD I vector and the *cis*P3H Type I pET-24-a(+) plasmid]. In this case, slightly different restriction enzymes were used (NdeI and HindIII), since the *cis*P3H Type I 5'- and 3'-cloning sites were different; the pET-24-a(+) *cis*P3H construct used 5'-*Nde*I and 3'-*Hind*III cloning sites. The purified fragments of linear pCOLD I vector and *cis*P3H Type I insert DNA were ligated together and transformed into BL21-Gold (DE3) competent cells; colonies were then screened for the correct *cis*P3H Type I gene insert. Plasmids isolated from colonies (using miniprep kits) were subject to NdeI/HindIII-restriction digestions and the products of these reactions were then visualised on a 0.8% (w/v) agarose-TAE gel; colony plasmids with the correct *cis*P3H Type I gene insert produced a band ~0.9 kb in size after restriction digestion (corresponding to the *cis*P3H Type I gene). Colonies containing plasmids with the correct insert were then sent for commercial sequencing and confirmed to contain the correct *cis*P3H Type I sequence.



Scheme 2.5: The strategy for subcloning the *cisP3H* Type I gene into a Takara pCOLD I vector.

(a) The pET-24-a(+) plasmids containing *cisP3H* Type I genes is treated with both HindIII and NdeI restriction endonucleases to cut the proline hydroxylase insert out of the PET-24-a(+) plasmid; linear pCOLD I vector DNA is similarly prepared using a HindIII-NdeI double reaction.

(b) The quenched, pCOLD I restriction digestion products is then treated with Calf Intestinal Alkaline Phosphatase (CIAP) to dephosphorylate the 5'-ends of the pCOLD I DNA strands and then both the linear pCOLD I vector and proline hydroxylase gene inserts are gel purified; dephosphorylating the pCOLD I vector prevents self-ligation of the linear vector fragments (*i.e.* vector + vector) during the ligation step (reducing background colonies).

(c) The purified *cisP3H* insert and the pCOLD I vector DNA are then combined in ratio of 3:1 and ligated together in a reaction catalysed by T4 DNA ligase; the newly formed plasmids are then transformed into BL21-Gold (DE3) *E. coli* competent cells and plated on ampicillin plates for antibiotic selection of the successful clones.

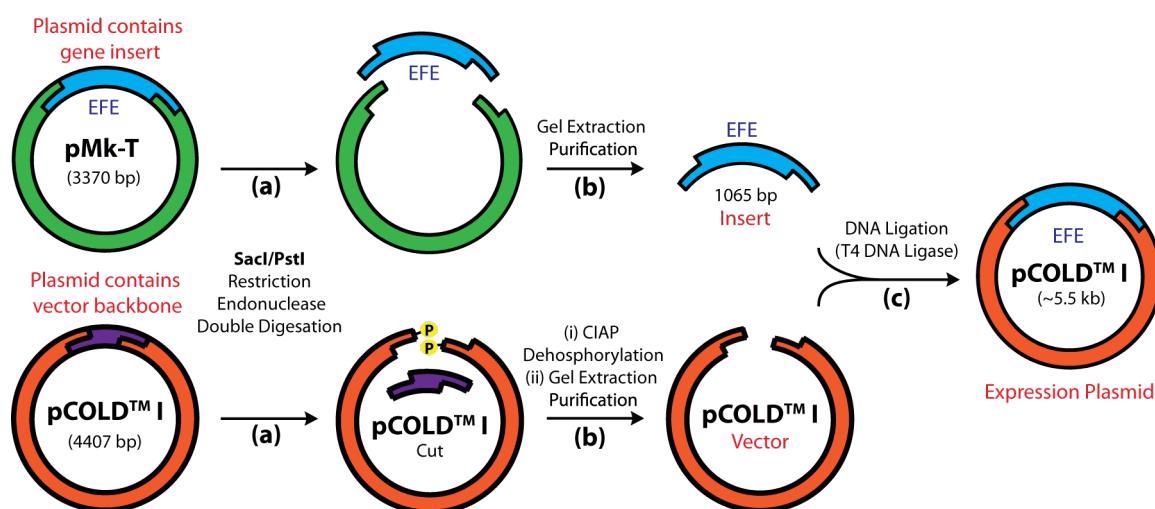
The expression of *cisP3H* Type I using the pCOLD I expression vector was tested for a number a manufacturer-recommended conditions and conditions based on those success before for *cisP4H* and *transP4H* proteins; nutritious TB medium and ZYM-5052 auto-induction medium were tested first and expression was negligible. Unfortunately, the pCOLD I *cisP3H* Type I vector was ineffective at producing hexahistidine-tagged *cisP3H* Type I and this method was not pursued further; perhaps the codon bias of the *E. coli* hosts hindered the effective transcription of the *cisP3H* Type I gene, and the method may have been more successful if the *cisP3H* Type I gene had been codon-optimised for *E. coli* hosts.

2.3 Ethylene-Forming Enzyme (EFE) Production

A full-length construct coding for *Pseudomonas syringae* pv. *phaseolicola* EFE was designed with appropriate 5'- and 3'-restriction cloning sites (*SacI* and *PstI*, respectively); the codon-optimised genes (for *E. coli* hosts) were then commercially synthesised using the GeneArt Gene Synthesis service (Life Technologies, Inc.).

2.3.1 Subcloning the *efe* Gene into Takara pCOLD I

Based on the large success of producing *cisP4H* and *transP4H* using a pCOLD I expression system, the *efe* gene was subcloned into the same pCOLD I vector to produce *N*-terminally hexahistidine-tagged ethylene-forming enzyme (EFE). An almost identical cloning strategy was used to insert the full-length *efe* gene into the pCOLD I vector (Scheme 2.6). The method, first, involves isolating linear fragments of *efe* DNA from circular pMk-T plasmids by using restriction endonucleases (*SacI* and *PstI*) to cut the *efe* gene out of the pMk-T vector; the linear *efe* fragments are then purified. Linear pCOLD I DNA is obtained using a similar *SacI*/*PstI*-double digestion method and is also dephosphorylated with CIAP before being purified. The linear pCOLD I vector and the *efe* insert are combined in a 1:3 ratio and ligated together in a reaction catalysed by T4 DNA ligase; the ligation reaction products are transformed into BL21-Gold (DE3) competent cells and the colonies screened for plasmids containing correctly cloned *efe* genes.



Scheme 2.6: The strategy for subcloning the *efe* gene into a Takara pCOLD I vector. (a) The pMk-T plasmid containing *efe* is treated with both SacI and PstI restriction endonucleases to cut the proline hydroxylase insert out of the pMk-T plasmid; linear pCOLD I vector DNA is similarly prepared using a SacI-PstI double reaction. (b) The quenched, pCOLD I restriction digestion products is then treated with Calf Intestinal Alkaline Phosphatase (CIAP) to dephosphorylate the 5'-ends of the pCOLD I DNA strands and then both the linear pCOLD I vector and proline hydroxylase gene inserts are gel purified; dephosphorylating the pCOLD I vector prevents self-ligation of the linear vector fragments (*i.e.* vector + vector) during the ligation step (reducing background colonies). (c) The purified *efe* insert and the pCOLD I vector DNA are then combined in ratio of 3:1 and ligated together in a reaction catalysed by T4 DNA ligase; the newly formed plasmids are then transformed into BL21-Gold (DE3) *E. coli* competent cells and plated on ampicillin plates for antibiotic selection of the successful clones.

Colonies containing transformed *efe*-pCOLD I ligation products were screened for plasmids with the correct *efe* gene inserts; plasmids were isolated from each colony (by miniprep methods of small growths) and SacI/PstI restriction digestion reactions were performed using each colony's plasmid. Visualisation of the products of the SacI/PstI restriction digests on a 0.8% (w/v) agarose-TAE NA gel revealed three plasmids contained the correct *efe* insert; the plasmids isolated from colonies 1,2, and 8 gave bands at ~1.1 kb on the agarose-TAE gel after restriction digestion, indicating the presence of the *efe* gene in these plasmids (Figure 2.32). Commercial gene sequencing reactions confirmed that these colonies did, in fact, contain pCOLD I plasmids with correct *efe* gene sequences.



Figure 2.32: Agarose-gel-electrophoresis analysis of *SacI*/*PstI*-digested colony plasmids; the plasmids were originally produced in experiments where the *efe* gene was ligated into the pCOLD I vector. Plasmids (pCOLD I) from colonies containing successfully cloned *efe* gene inserts gave a band at ~1.1 kb in analytical restriction digests; colonies 1, 2, and 8 appeared to contain the correct *efe* gene. 5 μ L samples were loaded on a 0.8% (w/v) agarose-TAE gel (run with a p.d. of 100 V).

2.3.2 EFE Production

The production of recombinant ethylene-forming enzyme was tested using the EFE pCOLD I construct in BL21-Gold (DE3) *E. coli* hosts; cells were cultured using terrific broth. The cultures were inoculated with samples taken from 2-YT overnight, starter cultures, which, themselves, were initiated with fresh transformants for healthy cell growth. The growths were incubated with 180 rpm shaking at 37 °C until the cell density reached OD₆₀₀ ~0.75, and then cooled to 15 °C and induced for EFE expression with 0.5 mM IPTG for 24 h; these growth conditions were recommended by the pCOLD I expression vector, Takara, and were used successfully to produce the proline hydroxylases (*cis*P4H/*trans*P4H).

Cells were lysed using a sodium dodecyl sulphate (SDS) solution and then samples of the water-soluble component of the cell lysate were taken for tris-glycine SDS-PAGE analysis (Figure 2.33); increasing volumes of lysate were loaded across the gel, as expression levels were still unknown, and therefore it was not clear yet how much sample must be loaded to be seen. All wells containing cell lysate had

a noticeable band that ran at a mass of ~ 41.5 kDa corresponding to soluble, *N*-terminally His₆-tagged EFE protein. Given the apparent success in expressing EFE protein, a purification method was considered to produce EFE protein.

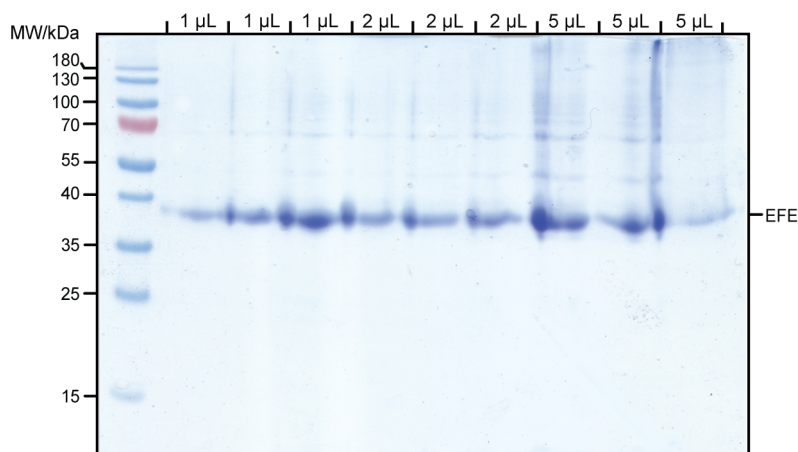


Figure 2.33: SDS-PAGE analysis of cell lysate from EFE expression trials. Varying amounts of lysate were loaded (1 μ L, 2 μ L, and 3 μ L - in triplicate) as the EFE band visibility threshold was still unknown. A noticeable band ~ 41.5 kDa in size confirmed the expression of soluble *N*-terminally His₆-tagged EFE. The gel was a 15% (w/v) acrylamide tris-glycine gel run using a p.d. of 220 V.

2.3.2.1 Recombinant EFE Purification

A purification method largely based on that used for the proline hydroxylases (see Section 2.2.2) was first attempted; the buffer pH and imidazole concentrations were within the criteria recommended by the His-column manufacturer. Low glycerol [5% (v/v)] and no detergent (Tween-20) was added to the EFE nickel-affinity purification buffers, because it thought the EFE should be fairly water-soluble (based upon the amino acid sequence and previous work^{242,255}); other components of the EFE purification buffers were the same as the proline hydroxylase buffers, however, such as the buffer type (50 mM HEPES-NaOH, pH 7.5), the salt concentrations (500 mM NaCl), and the imidazole concentrations (5 mM, 30 mM, and 400 mM - binding, wash, and elution buffers, respectively).

A large-scale, 6×1 L terrific broth batch was prepared; each 1 L flash of terrific

broth was inoculated with 10 mL samples of overnight, 2-YT culture (which, itself, was initiated with fresh transformants for healthy cells). The shaking (180 rpm), 1 L TB flasks were incubated at 37 °C (for ~4 h) until OD₆₀₀ ~0.75, and then cooled to 15 °C ready for cold-shock induction with 0.5 mM IPTG; the 15 °C induction lasted 24 h as was used for the proline hydroxylase enzymes (see Section 2.2.2). Harvested cells were frozen at -80 °C prior to EFE purification.

The harvested *E. coli* cells containing recombinant EFE were resuspended in cold (4 °C) binding buffer on ice and broken using sonication. Centrifuge-clarified, 0.45 µL-filtered cell lysate was then loaded onto an equilibrated GE Healthcare 5 mL His-Trap FF column (Figure 2.34 a); the His-column was then thoroughly washed with excess binding buffer until the UV_{280nm}-absorbance trace reached baseline levels once again (*i.e.* contaminants had been rinsed from the column). The wash buffer was then passed over the His-column; a significant elution peak was observed (Figure 2.34 a and b); given the large size of the wash-step UV_{280nm}-absorbance peak, it was likely that most of the protein being eluted was EFE, and fraction collection was quickly initiated (Figure 2.34 a and b). After the wash step, an increasing linear gradient of imidazole [30 mM-400 mM Imidazole - 0%-100% elution buffer (B)] was applied to the column the elute any remaining column-bound EFE protein; another large peak was seen in the UV_{280nm}-absorbance chromatogram and was thought to correspond to hexahistidine-tagged EFE (Figure 2.34 a and b). Tris-glycine SDS-PAGE analysis of wash fractions and elution step fraction confirmed that *N*-His₆-EFE was, in fact, present in both the wash step and elution step fractions (~120 mg EFE) and at least >80% pure. Given that some EFE eluted from the His-column during the wash step, it was thought a lower concentration of imidazole would be more appropriate for column wash steps in future purifications; this could be achieved by washing the column with a large excess of binding buffer to remove contaminants from the column whilst still retaining column-bound EFE

protein. The wash and elution step fractions were pooled and spin-concentrated (to ~1 mL) for the next purification step.

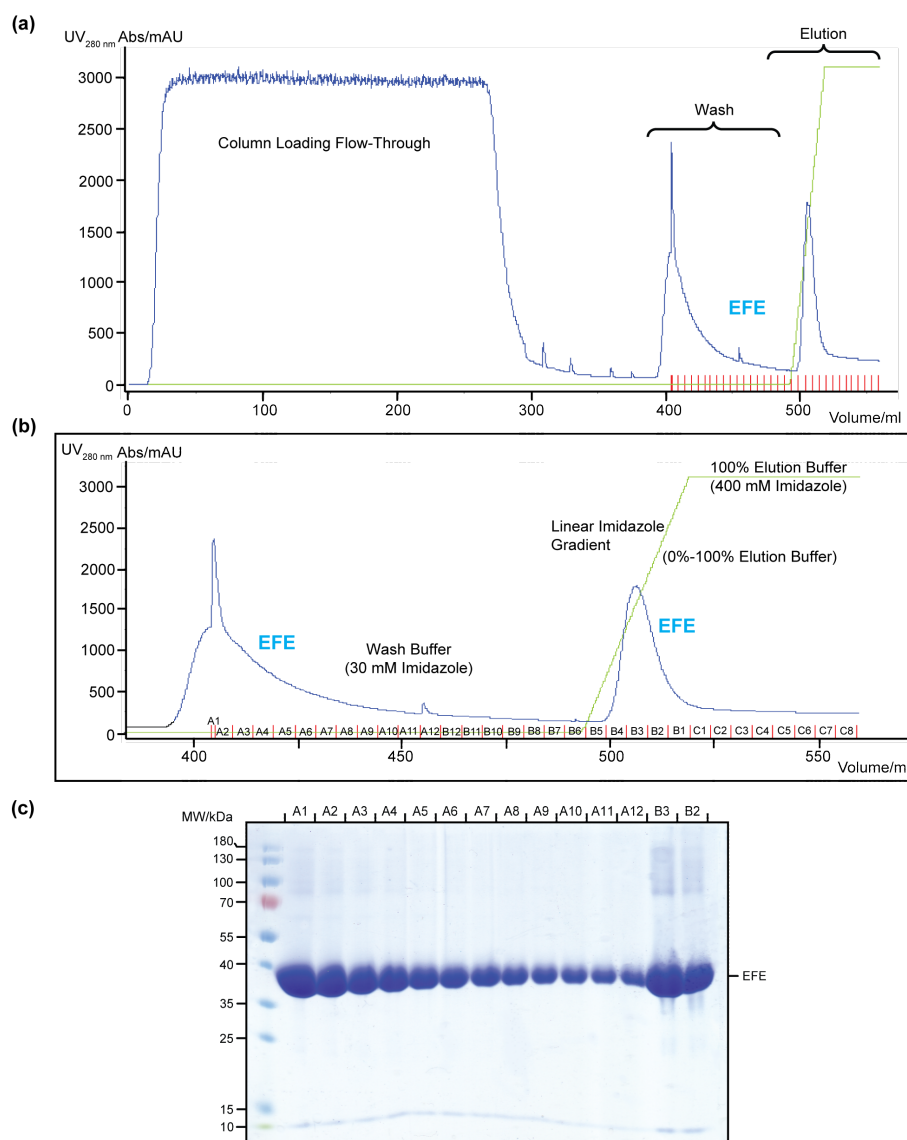


Figure 2.34: Immobilised-nickel-affinity chromatography of EFE:

(a) After all of the cell lysate had been loaded, the column was washed with excess binding buffer (until the UV_{280nm}-absorbance trace reached baseline levels). The column was then washed with wash buffer (30 mM Imidazole) and some EFE appeared to elute; fractions of this wash peak were collected.

(b) The column was then treated with a linear gradient of increasing imidazole (30 mM-400 mM) to elute the remaining *N*-His₆-EFE from the His-column, which was manifested as a large, broad peak in the UV_{280nm}-absorbance chromatogram.

(c) The wash and elution fractions (5 μ L samples) were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] and found to contain an abundance of *N*-His₆-EFE (~41.5 kDa).

Column: 5 mL HisTrap FF. Binding Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 5 mM Imidazole, 5% (v/v) glycerol, 0.5 mM TCEP, pH 7.5. Wash Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 30 mM Imidazole, 5% (v/v) glycerol, pH 7.5. Elution Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 400 mM Imidazole, 5% (v/v) glycerol, pH 7.5.

The IMAC EFE purification fractions were then loaded onto a (~320 mL) Pharmacia HiLoad 26/60 Superdex 200 size-exclusion column; the EFE protein eluted to produce a large and relatively sharp peak (~320 mL) in the UV_{280nm}-absorbance trace (Figure 2.35 a). Tris-glycine SDS-PAGE analysis of the elution fractions (Figure 2.35 b) showed that the EFE protein was largely separated from the high-molecular weight impurity (roughly twice the mass of EFE at ~85 kDa - perhaps a dimer aggregate); the protein was estimated to be >90% pure in the purest fractions (C1-C10), which were pooled (~105 mg EFE), spin-concentrated (to ~42 mg/mL), and frozen at -80 °C. Interestingly, the concentrated EFE protein is a dark blue colour when it is concentrated.

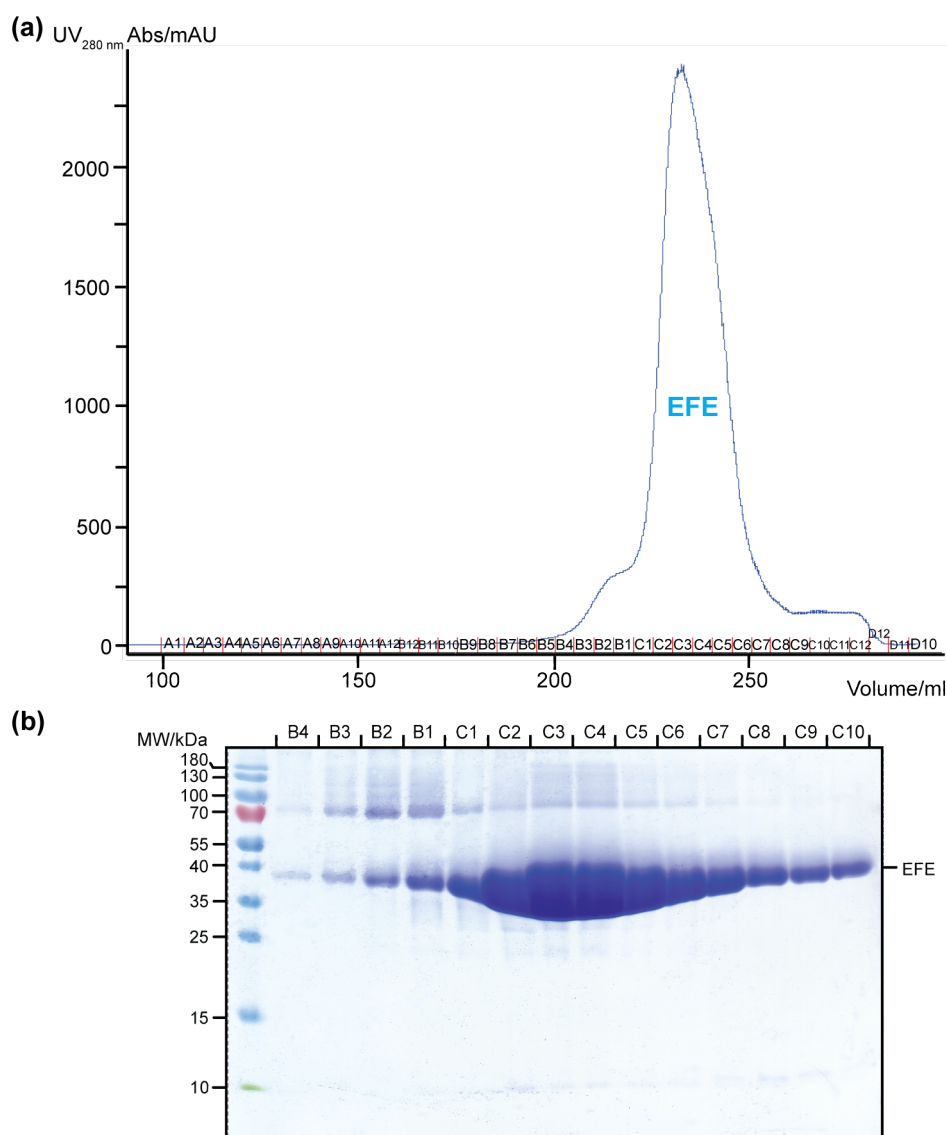


Figure 2.35: Size-exclusion purification of EFE as a second column-step after IMAC purification:

(a) the UV_{280nm}-absorbance chromatogram for the purification;
 (b) and the tris-glycine SDS-PAGE [15% (w/v) acrylamide] analysis of the EFE purification fractions (5 μ L samples). The presence of EFE is indicated by a marker.

Column: Pharmacia (320 mL) HiLoad 26/60 Superdex 200

Buffer: 50 mM HEPES-NaOH, 250 mM NaCl, pH 7.0.

Another EFE purification was attempted based upon what was learned from the initial, large-scale EFE purification; the same growth conditions were used as those previously, but fewer flasks were grown (4×1 L TB), since EFE expression was so high-yielding. In the nickel-affinity purification step, the column was thoroughly washed with a large excess of binding buffer (5 mM imidazole) in an attempt to rinse the column of impurities whilst still retaining column-bound *N*-His₆-EFE; the binding of the EFE had been shown to be relatively weak during the initial nickel-affinity EFE purification and so the column was only washed with buffer with a low concentration of imidazole (protein eluent). The target EFE protein was then eluted with a linear gradient of increasing imidazole (5 mM-400 mM - 0%-100% elution buffer), as before, and the EFE produced a large and broad peak in the UV_{280nm}-absorbance chromatogram (Figure 2.36 a). This slightly altered approach worked well, with all EFE being retained on the column, and EFE being purified with an acceptable level of contamination (<20% by SDS-PAGE analysis) (Figure 2.36 b).

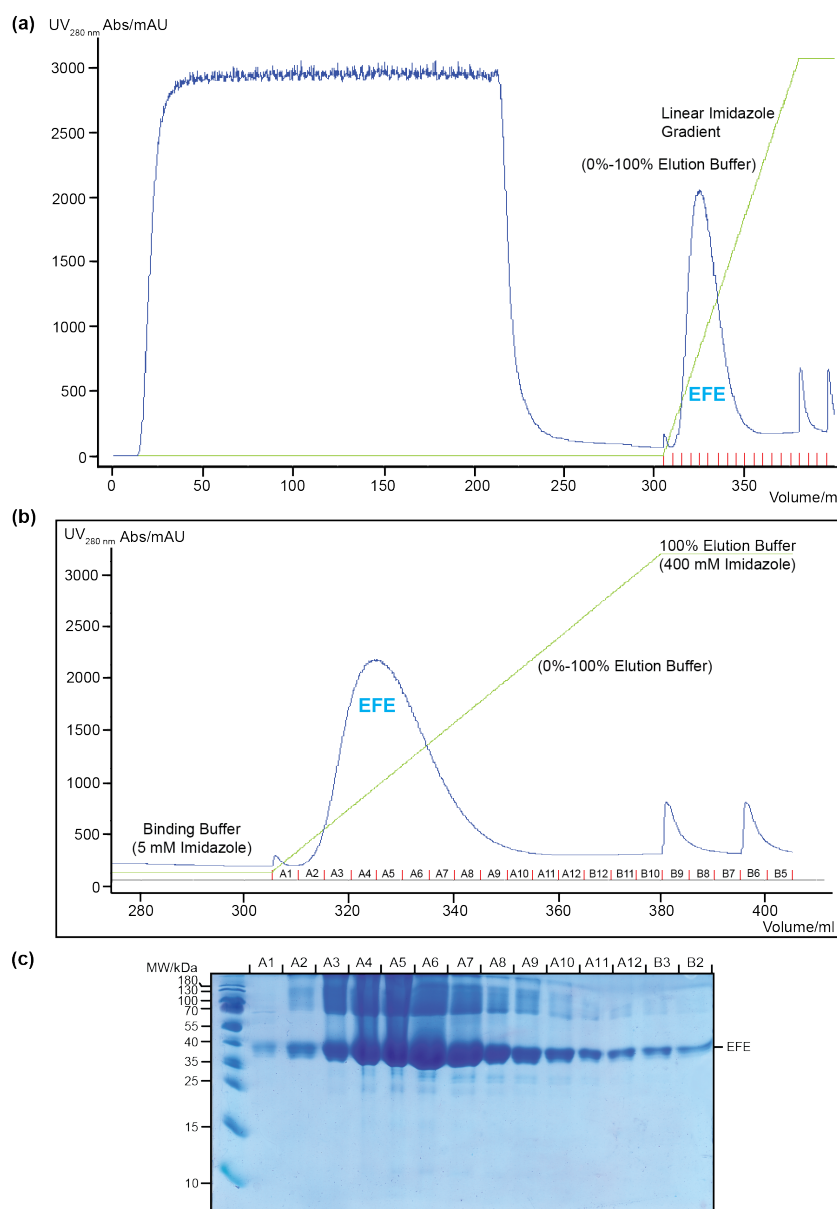


Figure 2.36: Immobilised-nickel-affinity chromatography of EFE.

(a) After all of the cell lysate had been loaded, the column was washed with excess binding buffer (until the UV_{280nm}-absorbance trace reached baseline levels).

(b) A linear gradient of increasing imidazole (5 mM-400 mM) to elute the *N*-His₆-EFE from the His-column, which was manifested as a large, broad peak in the UV_{280nm}-absorbance chromatogram.

(c) The elution fractions (5 μ L samples) were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] and found to contain an abundance of *N*-His₆-EFE (~41.5 kDa).

Column: 5 mL HisTrap FF. Binding Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 5 mM Imidazole, 5% (v/v) glycerol, 0.5 mM TCEP, pH 7.5. Elution Buffer: 50 mM HEPES-NaOH, 500 mM NaCl, 400 mM Imidazole, 5% (v/v) glycerol, pH 7.5.

The EFE protein from the nickel-affinity purification step was further purified using a DEAE Sepharose anion-exchange column; buffers were selected based upon the guidelines set by the DEAE Sepharose resin manufacturer (Pharmacia), having a pH (8.0) that was at least one pH unit higher than the theoretical pI of the EFE protein (6.41) being purified (so that the EFE would effectively bind the anion-exchange column). Despite this, much of the EFE did not bind the 30 mL DEAE Sepharose FF column during protein loading and ended up in the column flow-through (Figure 2.37 a); EFE appeared to bind the DEAE Sepharose resin fairly weakly. The remaining, column bound was then eluted using a linear, increasing salt gradient (0 mM-500 mM NaCl - 0%-100% elution buffer); the bound EFE eluted to produce a sharp peak in the UV_{280nm}-absorbance chromatogram (Figure 2.37 a). SDS-PAGE analysis confirmed that there was indeed EFE in the column flow-through as well as in the elution fractions; the protein purity did slightly improve, however, despite these purification difficulties. Perhaps the EFE may bind the DEAE Sepharose resin in buffer of a higher pH, but could also just bind DEAE Sepharose in an inherently weak manner. The column flow-through and the major elution fractions (A6-B6) were pooled and concentrated (to ~1 mL) for the next purification step.

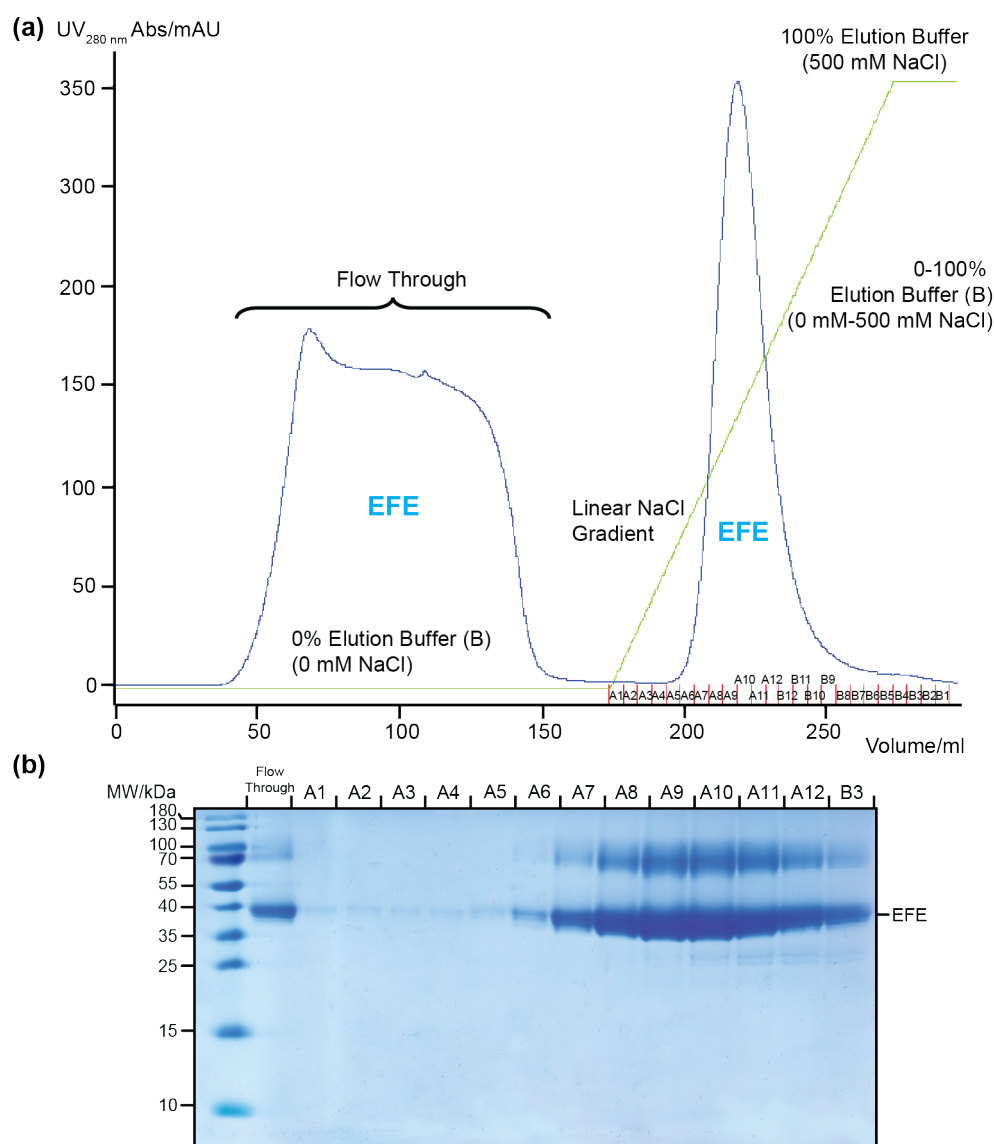


Figure 2.37: DEAE Sepharose FF anion-exchange chromatography of EFE:
 (a) the UV_{280nm} -absorbance FPLC chromatogram for the EFE gel filtration fractions;
 (b) the fractions were analysed by tris-glycine SDS-PAGE [15% (w/v) acrylamide] to assess protein purity. 5 μ L samples of each fraction were run in each lane. The marker denotes the EFE target protein; the bands ran at a the correct MW ($\sim$ 41.5 kDa).

Column: Pharmacia (30 mL) DEAE Sepharose FF Anion-Exchange

Buffer A: 50 mM HEPES-NaOH, pH 8.0.

Buffer B: 50 mM HEPES-NaOH, 500 mM NaCl, pH 8.0.

The pooled and concentrated DEAE Sepharose anion-exchange fractions were then further purified by size-exclusion chromatography; in this EFE purification, a gel filtration column using a resin (Superdex 75) that more effectively separates proteins of a similar MW to EFE was used in attempt to further resolve the EFE elution peak from the high-MW contaminant bands. The size-exclusion purification using Superdex 75 resin did work more effectively than when using a Superdex-200-based column, with most of the high-MW impurity being eluted before the major EFE peak at an elution volume of roughly 140 mL (peak centre)(Figure 2.38 a); the EFE peak then began to elute after ~150 mL of buffer had flowed through the size-exclusion column, producing a large peak in the UV_{280nm}-absorbance chromatogram, which centred at a volume of ~170 mL (Figure 2.38 a). SDS-PAGE analysis of the elution fractions confirmed that the EFE protein was largely separated from sample impurities (Figure 2.38 b) and was estimated to be >90% pure (yield 90 mg). The purest fractions (C8-D6) were pooled, spin-concentrated (to ~41 mg/mL) and frozen at -80 °C.

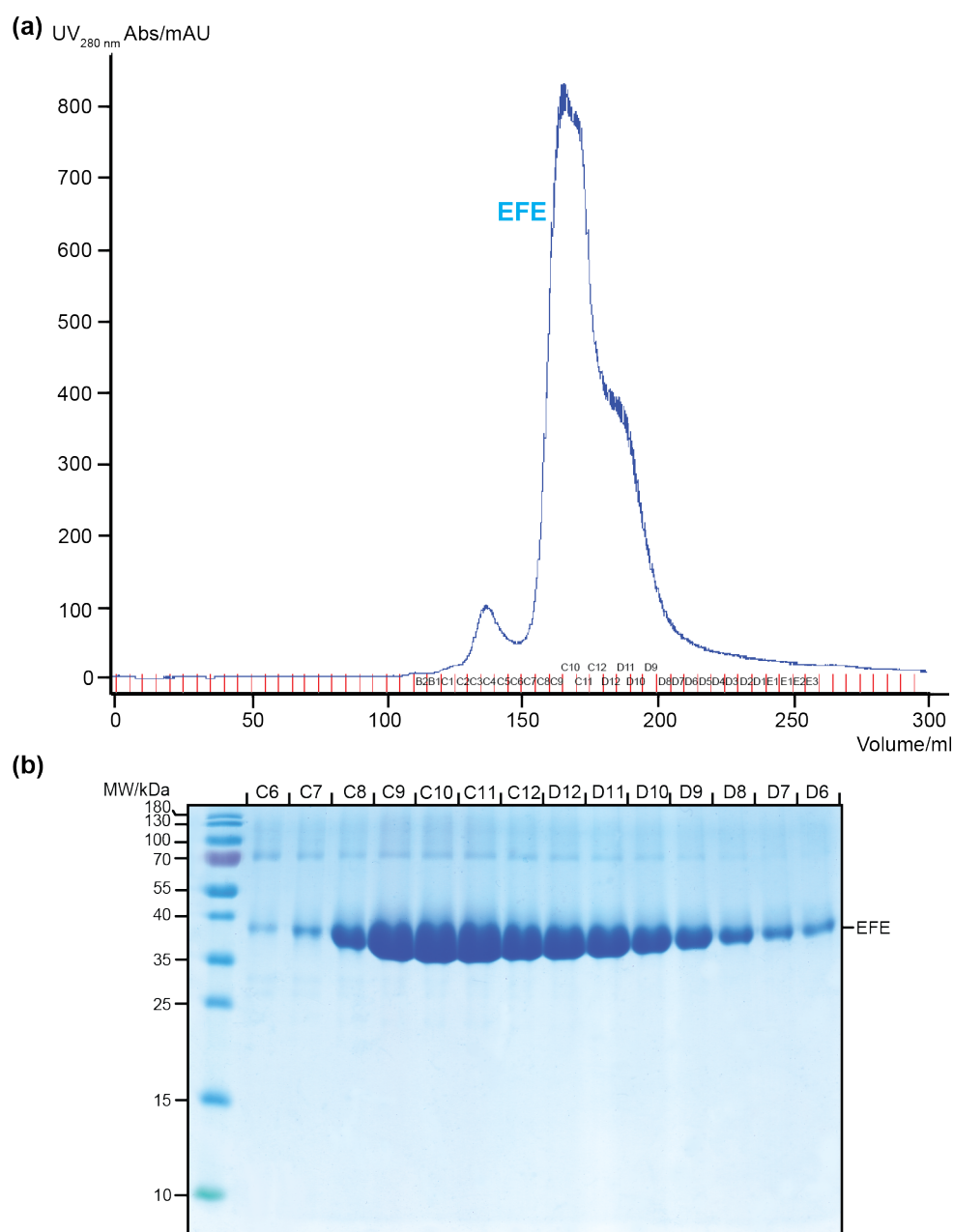


Figure 2.38: Size-exclusion purification of EFE as a second column-step after IMAC purification:

- (a) the UV_{280nm}-absorbance chromatogram for the purification;
- (b) and the tris-glycine SDS-PAGE [15% (w/v) acrylamide] analysis of the EFE purification fractions (5 μ L samples). The presence of EFE is indicated by a marker.

Column: Pharmacia (320 mL) HiLoad 26/60 Superdex 75

Buffer: 50 mM HEPES-NaOH, 250 mM NaCl, pH 7.0.

2.4 Summary

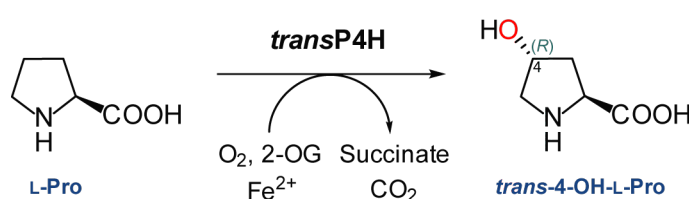
Overall, the methods in this chapter are sufficient for the large-scale (>10 mg) production of 3 of the target proline hydroxylases (*cis*P3H Type I, *cis*P4H, and *trans*P4H), as well as the EFE from *Pseudomonas syringae* *pv.* *phaseolicola*; the protein was used for both biochemical studies and crystallographic experiments.

Chapter 3

The L-Proline Hydroxylases (PHs)

3.1 PH Discoveries

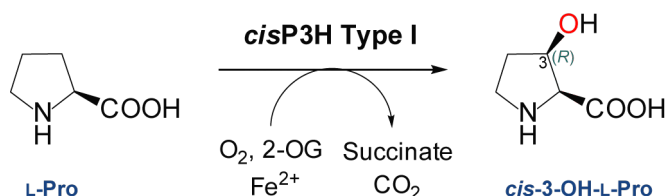
As reviewed in Chapter 1, proline hydroxylases have been established in the biosynthesis of peptide antibiotics (see Section 1.2.2.2).^{52,180–182} *Streptomyces griseoviridis* P8648 cell-free lysates were subsequently shown to contain a Fe(II)/2-OG-dependent L-proline hydroxylase. (now known as *trans*P4H) (Scheme 3.1);^{191,193} this contrasts with mammalian prolyl hydroxylation.^{139,140,157}



Scheme 3.1: The reaction catalysed by *trans*P4H: free L-proline is hydroxylated to produce *trans*-4-hydroxy-L-proline with a retention of stereochemistry.

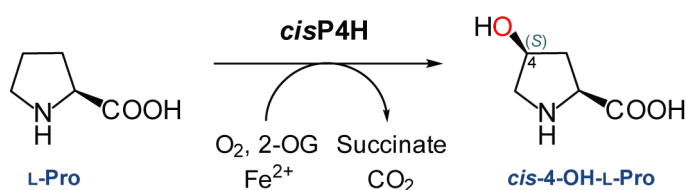
Other microbial L-proline hydroxylases have subsequently been reported; Mori *et al.* found a novel activity, which hydroxylates L-proline to produce *cis*-3-hydroxy-L-proline).²² These L-proline *cis*-3-hydroxylases (*cis*P3Hs) were thought to be ODDs,²² and a target gene was cloned from *Streptomyces sp.* strain TH1 and the corresponding *cis*P3H was produced;²³ *in vitro* reactions showed that the purified enzyme (now

known as 'cisP3H Type I') catalysed hydroxylation of L-proline in the presence of ferrous iron and 2-OG to give *cis*-3-hydroxy-L-proline (Scheme 3.2).²³



Scheme 3.2: The reaction catalysed by L-proline *cis*-3-hydroxylase Type I (*cis*P3H Type I): *cis*P3H Type I converts L-Pro to *cis*-3-hydroxy-L-proline.

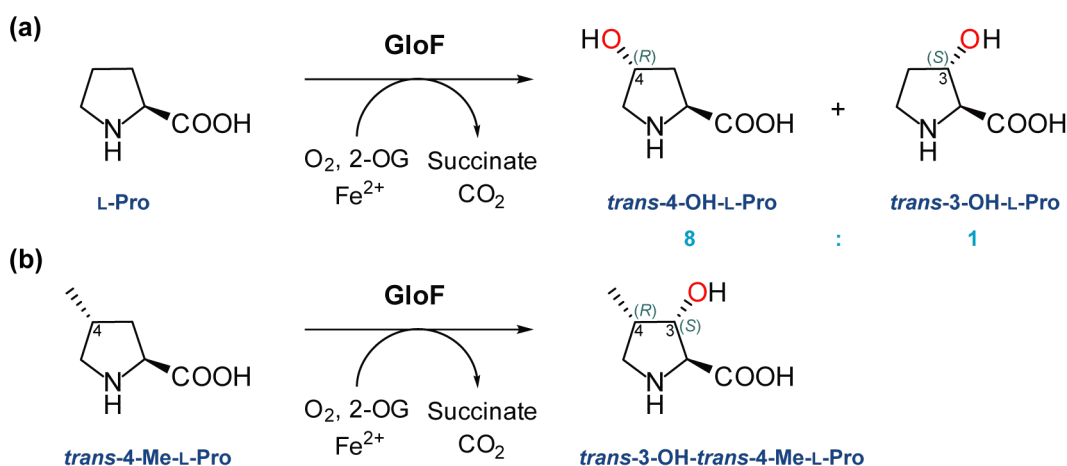
An isozyme gene (870 bp) of *cis*P3H Type I was cloned from *Streptomyces* *sp.* strain TH1, and the protein that it codes for ('*cis*P3H Type II') was produced.²³ Despite extensive screening of microorganisms exhibiting L-proline *cis*-4-hydroxylase activity,^{22,206} microorganisms which produce *cis*-4-hydroxy-L-proline from free L-proline were not reported until 2009.²⁰⁷ Two genes from the nitrogen-fixing, legume-root-nodule bacteria, *Mesorhizobium loti*²⁰⁸ and *Sinorhizobium meliloti*,²⁰⁹ were cloned and overexpressed in *E. coli*. Analyses of reactions catalysed by these two enzymes showed that they catalysed the Fe(II)-/2-OG-dependent hydroxylation of L-proline to produce *cis*-4-hydroxy-L-proline (Scheme 3.3).



Scheme 3.3: The reaction catalysed L-proline *cis*-4-hydroxylase (*cis*P4H): *cis*P4H converts L-proline to *cis*-4-hydroxy-L-proline.

Fungal proline hydroxylases have also been reported.²¹¹ Echinocandins are cyclic, nonribosomal hexapeptides, based mostly on non-proteinogenic amino acids, including 2,3-*trans*-3,4-*cis*-3-hydroxy-4-methyl-L-proline, *trans*-3-hydroxy-L-proline, and *trans*-4-hydroxy-L-proline.¹¹⁰ Crude extracts from a pneumocandin-producing fungus, *Glarea lozoyensis*, was found to contain a 2-OG-dependent L-proline *trans*-

3-hydroxylase activity (beside a *trans*P4H);²¹¹ the extract required ascorbate and dithiothreitol for enzymatic turnover, and inclusion of ferrous iron was required for optimal catalytic activity,²¹¹ suggesting that the enzyme(s) were ODD(s). Recently, a proline hydroxylase, GloF, was identified from the pneumocandin biosynthesis cluster of *Glarea lozoyensis* ATCC 74030;²¹² overexpression and characterisation of GloF revealed it was the first eukaryotic Fe(II)- and 2-OG-dependent proline hydroxylase to be discovered.²¹² GloF was shown to catalyse the simultaneous *trans*-3- and *trans*-4-hydroxylation of L-proline in a ratio of 1:8 (Scheme 3.4 a), respectively.²¹² GloF also converts *trans*-4-methyl-L-proline into the corresponding product, *trans*-3-hydroxy-*trans*-4-methyl-L-proline (Scheme 3.4 b).²¹²



Scheme 3.4: The reactions catalysed by GloF, a L-proline *trans*-3/4-hydroxylase: GloF also converts *trans*-4-methyl-L-proline into *trans*-3-hydroxy-*trans*-4-methyl-L-proline.

3.2 Proline Hydroxylase Assays

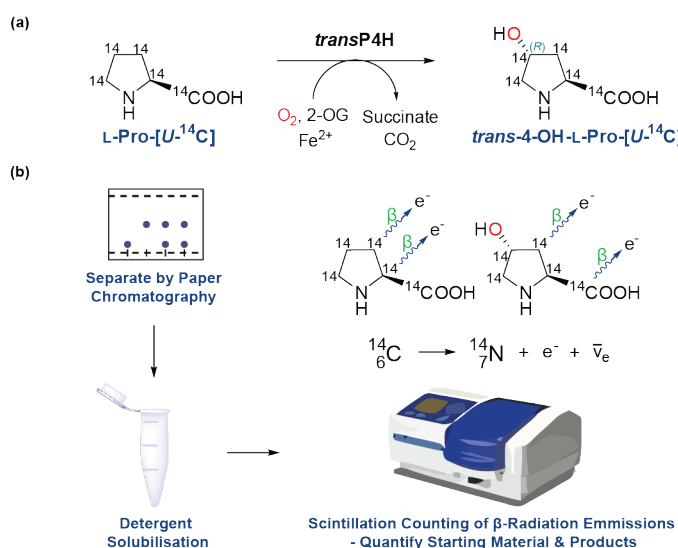
Various methods have been developed for proline hydroxylases: some follow turnover of L-proline and others follow the conversion of 2-OG into succinate. As summarised below, following the successful production of proline hydroxylases (see Chapter 2), I then focused on developing assays to follow the reactions of these enzymes. Initially, I carried out a literature survey of relevant assays.

3.2.1 [^{14}C]-Radioactivity Detection

Early methods for proline and hydroxyproline separation used paper chromatography, and the positions of the separated components were detected using specific colouring agents;^{317,318} these methods were later expanded to detect hydroxyamino acids from protein hydrolysates.³¹⁹ Early L-proline hydroxylase assays used [^{14}C]-radiolabelled substrates, and established paper chromatography methods were used to isolate reaction products (e.g. L-hydroxyproline, succinate, *etc.*) and starting materials (e.g. L-proline, 2-OG, *etc.*) for quantification; scintillation counting was then used for compound quantification to calculate reaction turnovers.^{53,256}

L-Proline hydroxylation by *trans*P4Hs in cell-free extracts⁵³ and purified enzymes²⁵⁶ have been monitored using a radioactive L-proline-[$U\text{-}^{14}\text{C}$] turnover assay.^{53,256} Reactions (100 μL) with cell-free extracts containing *trans*P4H were carried out with all necessary reaction components (*i.e.* L-Pro, 2-OG, $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, sodium ascorbate, catalase) spiked with 0.1 pCi L-proline-[$U\text{-}^{14}\text{C}$];⁵³ reactions were carried out at 25 °C in 20 mM TES-NaOH, pH 7.5 buffer (Scheme 3.5 a).⁵³ The L-proline starting material and *trans*-4-hydroxy-L-proline product of the *trans*P4H reactions were separated by paper chromatography, solubilised from the DE-81 paper with SDS detergent, and then quantified using a liquid scintillation spectrometer (Scheme 3.5 b).⁵³ The assays which were implemented by Lawrence *et*

al. using purified *trans*P4H used a similar procedure to that used for cell-free extract assays, but with a higher specific activity of L-proline- $[U-^{14}\text{C}]$ (typically 206 mCi mmol⁻¹);²⁵⁶ after incubation, L-proline and *trans*-4-hydroxy-L-proline (100 mM stocks) were then added to a final concentration of 17 mM to prepare them for separation by ion-exchange paper chromatography.²⁵⁶ Levels of isolated L-proline and *trans*-4-hydroxy-L-proline were quantified using a scintillation counter; turnovers and *trans*P4H activities were then calculated by subtracting the *trans*-4-hydroxy-L-proline- $[U-^{14}\text{C}]$ background levels of heat-inactivated-*trans*P4H control samples from the corresponding *trans*-4-hydroxy-L-proline:L-proline ratios measured for complete reactions (*i.e.* samples with active *trans*P4H).



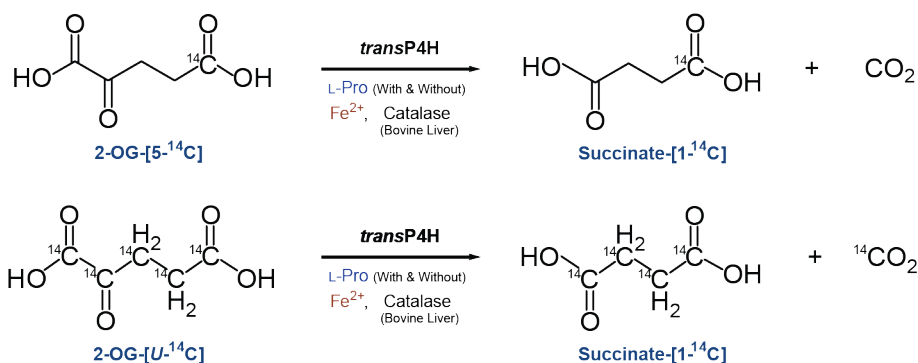
Scheme 3.5: Radioactive assay using $[^{14}\text{C}]$ -labelled L-Pro: (a) *trans*P4H catalyzes the hydroxylation of L-proline- $[U-^{14}\text{C}]$ to produce *trans*-4-hydroxy-L-proline- $[U-^{14}\text{C}]$, and (b) a scintillation counter is used to quantify the products and starting material.

Reactions (100 μL) have been carried out at 25/26 $^\circ\text{C}$ with the following conditions: 1 mM L-Pro, 0.5 mM 2-OG, 0.5 mM $(\text{NH}_4)_2\text{FeSO}_4$, 1 mM sodium ascorbate, 10 pg or 0.2 mg/mL catalase.^{53,256} Reactions with cell-free extract used 0.33 uCi L-Pro- $[U-^{14}\text{C}]$.⁵³ Reactions were initiated with 20 μL of cell-free extract/purified *trans*P4H.^{53,256}

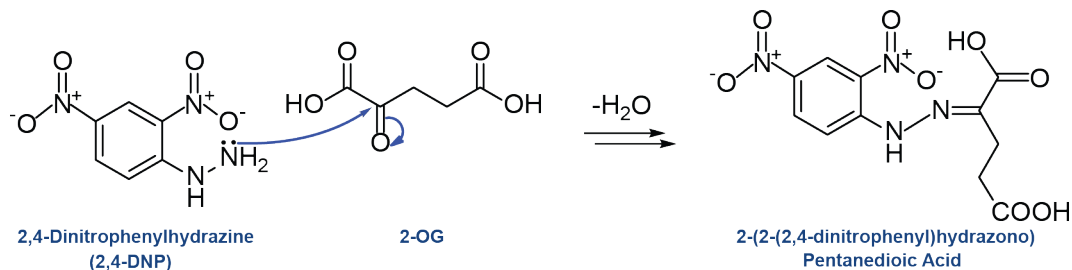
The catalytic decarboxylation of 2-OG as catalysed by *trans*P4H has also been monitored using $[^{14}\text{C}]$ -labelled 2-OG substrates (2-OG- $[5-^{14}\text{C}]$ and 2-OG- $[U-^{14}\text{C}]$);²⁵⁶ *trans*P4H catalyses the conversion of 2-OG- $[5-^{14}\text{C}]/[U-^{14}\text{C}]$ to succinate- $[1-^{14}\text{C}]/[U-^{14}\text{C}]$.

^{14}C] (Scheme 3.6 a), respectively.²⁵⁶ After reaction quenching, the remaining 2-OG-[5- ^{14}C] was precipitated using 2,4-dinitrophenyl hydrazine (2,4-DNP), which reacts to form a hydrazone with the C-2-carbonyl group of 2-OG (Scheme 3.6 b);^{256,320} the remaining succinate-[1- ^{14}C] is then quantified by liquid scintillation counting.^{256,320} This assay was initially developed using the collagen prolyl 4-hydroxylase, PP4H, as a test system;³²⁰ the procedure is simple, quick, cheap, and radiochemically safe, and produces highly reproducible results (repeated measurements have a standard deviation of ~2% of the mean).³²⁰

(a)



(b)



Scheme 3.6: Radioactive assay using [^{14}C]-labelled 2-OG: (a) *trans*P4H catalyses the decarboxylation of 2-OG-[5- ^{14}C]/[U- ^{14}C] to produce succinate-[1- ^{14}C]/[U- ^{14}C],^{256,320} respectively, and (b) the remaining [^{14}C]-labelled 2-OG starting material was precipitated using 2,4-DNP to form a hydrazone.

Reactions (50 μL) were carried out at 26 $^\circ\text{C}$ at for 30 min under the following conditions: 25 mM TES-NaOH, pH 7.5, 0.1 mM 2-OG (sodium salt - containing 0.08 μCi of 2-OG-[5- ^{14}C]/[U- ^{14}C] - specific radioactivities = ~ 260 mCi mmol $^{-1}$), 0.2 mg/mL bovine liver catalase, 0.5 mM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM L-Pro, and 10 μL protein solution.²⁵⁶

These *trans*P4H [^{14}C]-radioactivity assays were used to confirm the strict requirement for ferrous iron and 2-OG for catalysis, as well as show that *trans*P4H is

inhibited by aromatic analogues of 2-OG;¹⁹³ the IC_{50} s of pyridine-2,4-dicarboxylate (2,4-PDCA), pyridine-2,5-dicarboxylate, and 3,4-dihydroxybenzoate (3,4-DHBA) (Figure 3.1) were measured to be 5 μ M, 49 μ M, and 32 μ M, respectively, and pyridine-2,6-dicarboxylate was found to be a catalytic activator.¹⁹³ 2,4-PDCA was also shown to be a competitive inhibitor of *trans*P4H with a $K_i = 5 \mu$ M.¹⁹³ In addition, Lawrence *et al.* reported that the addition of L-ascorbate does not stimulate L-proline-coupled 2-OG turnover by *trans*P4H from *Streptomyces griseoviridis* P8648,¹⁹³ but L-ascorbate was shown to stimulate L-Pro-uncoupled 2-OG turnover.¹⁹³

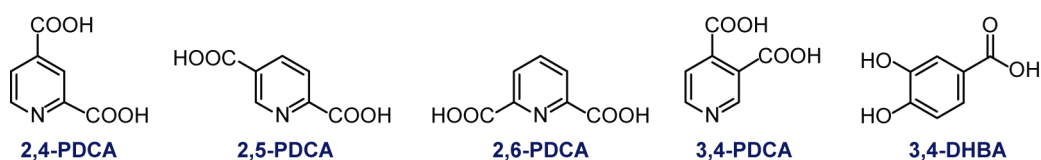


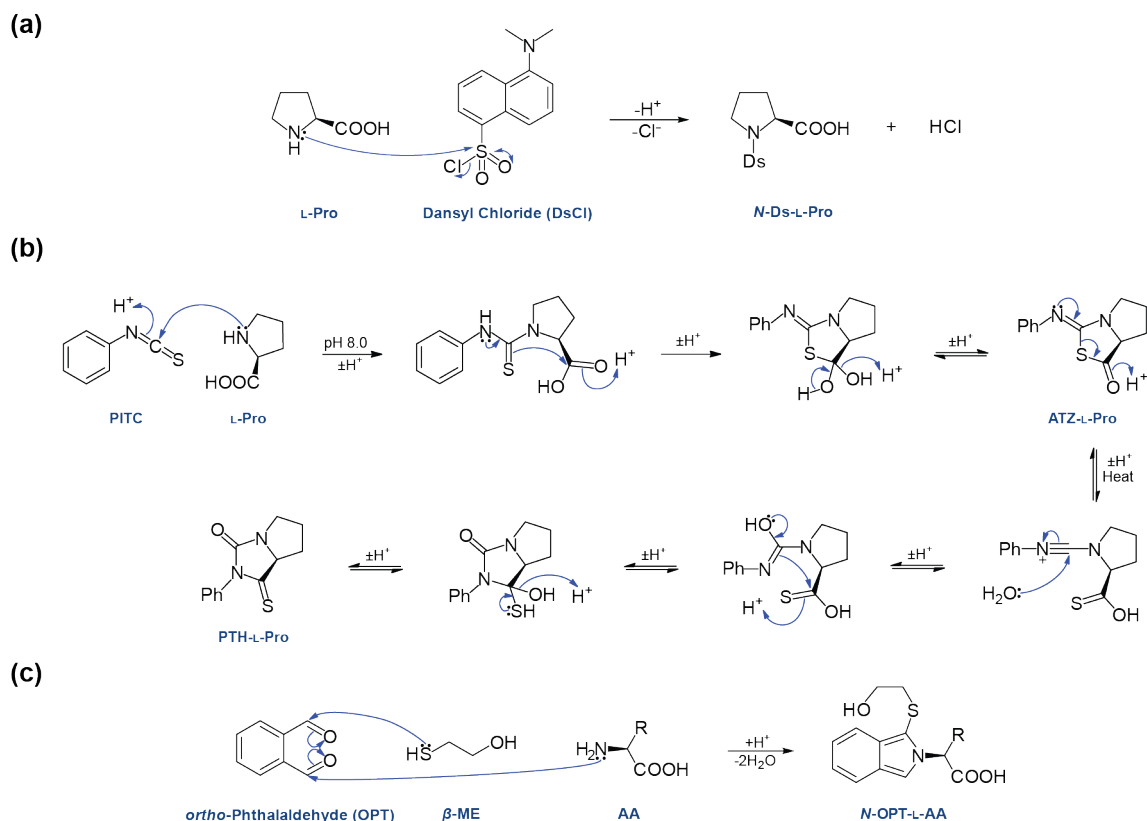
Figure 3.1: *trans*P4H inhibitors: aromatic 2-OG analogues including pyridine dicarboxylic acids (PDCAs) and 3,4-dihydroxybenzoic acid (3,4-DHBA).¹⁹³

3.2.2 Reverse-Phase HPLC Assays

Many HPLC detection methods have been developed for amino acids, including proline. In a number of methods, amino acids are typically subjected to *N*-derivatisation, a step which adds a hydrophobic, aromatic ring system to the amino nitrogen atom of the proline compound;^{321–332} derivatisation improves reverse-phase HPLC compound separation (by improving column binding strength), as well as improving compound detection by UV-absorption spectroscopy or mass spectrometry.

An early amino acid HPLC-detection method involved *N*-dansylation of the amino nitrogen atom with dansyl chloride (DsCl) (Scheme 3.7 a).³²¹ Another early method used phenylisothiocyanate (PITC) to derivatise the amino and carboxylic acid groups of amino acids (Scheme 3.7 b), initially forming a 5-membered anilinothiazolidine (ATZ) species (kinetic product),³²² which, on heating, rearranges to form a more thermodynamically stable phenylthiohydantoin (PTH) compound;³²² this re-

action is essentially the same as that used in Edman Degradation peptide sequencing.³³³ Some amino acids have also been derivatised for HPLC detection using *ortho*-phthalaldehyde (OPT) and β -mercaptoethanol, which combine with the amino group to form a 2*H*-isoindole derivative.^{323–327}



Scheme 3.7: Amino acid derivatisation methods for HPLC detection and quantification:

- (a) addition of an amino dansyl (Ds) group by treatment with dansyl chloride (DsCl);³²¹
- (b) by treatment with phenylisothiocyanate (PITC) to form an anilinothiazolinone (ATZ) species, which rearranges to a phenylthiohydantoin (PTH) compound on heating;³²²
- (c) OPT derivatisation by treatment with phthalaldehyde (OPT) and β -mercaptoethanol, which combine with the amino group to form a 2*H*-isoindole derivative.^{323–327}

As well as radiolabelling experiments, proline hydroxylase assays have been developed, which involve the HPLC-based detection of hydroxyproline products and L-proline starting materials; these methods do not require ‘special’ labelled substrates, improving detection sensitivity, reducing costs, as well as removing the need for sophisticated detection equipment for sample analysis. Many of these

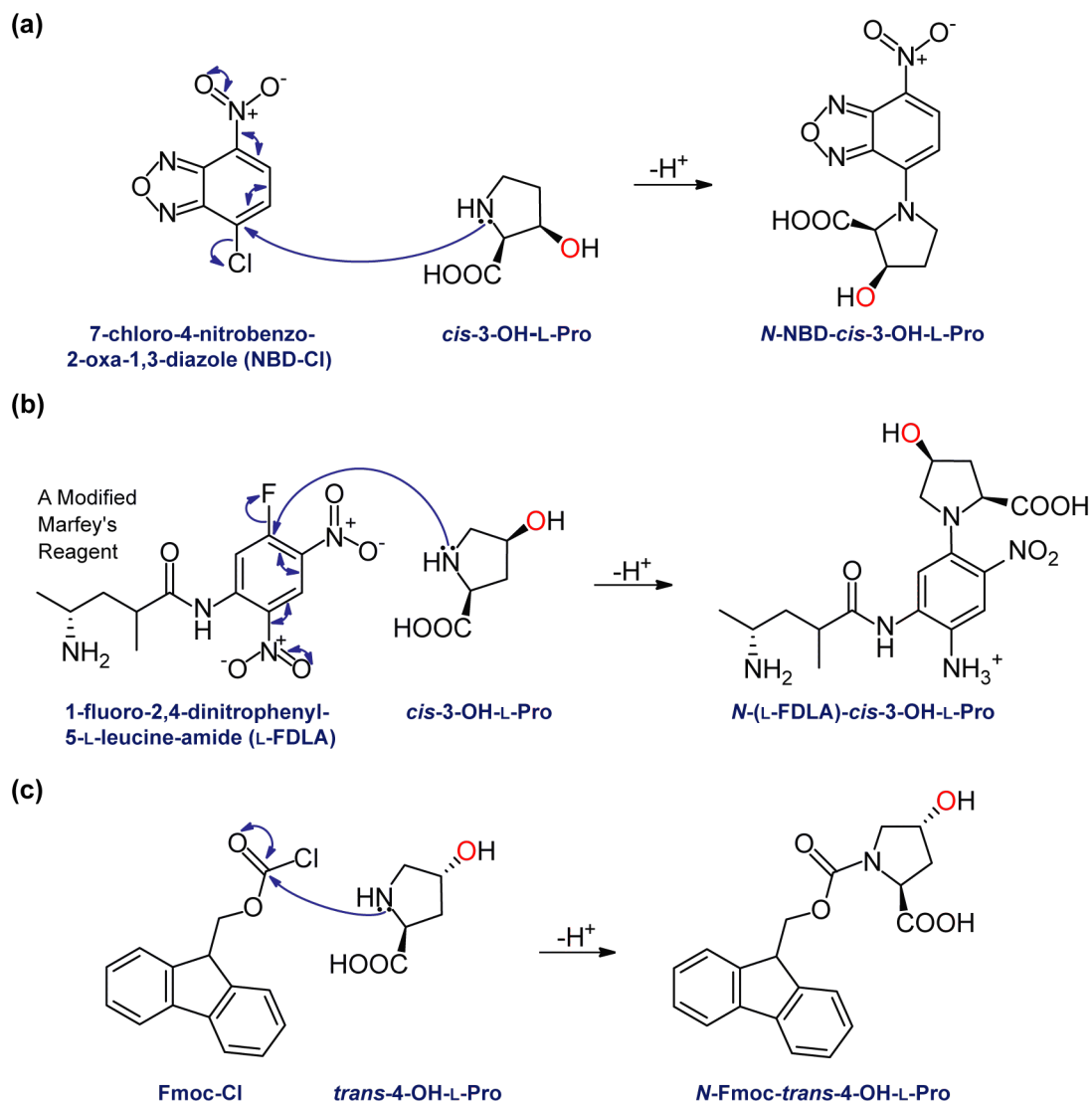
methods ‘borrow’ from methods used to analyse samples containing endogenous proline-based compounds.

An HPLC-based method for detection (and quantification) of hydroxyproline (Hyp) stereoisomers was developed to reliably monitor endogenous Hyp levels;^{329,334} this can be rather difficult, because Hyps are usually a minor component of cellular amino acids, and the chiral resolution of Hyp stereoisomers is difficult. In nature, *trans*-4-L-Hyp, *cis*-4-L-Hyp, *cis*-4-D-Hyp, *trans*-3-L-Hyp, and *cis*-3-L-Hyp have been reported in mammalian connective tissues, plant tissues, and microbial sources.¹⁰⁹ A pre-column derivatisation method using 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl) was reported for the specific detection of Pro and Hyp;³³⁴ this method selectively adds a NBD motif onto the amino N-atom of the proline compound (Scheme 3.8 a), which allows for HPLC detection (using two C18 reverse-phase columns connected in series).³³⁴ The derivatised imino acid fractions are then quantitated spectrophotometrically using light with a wavelength of 495 nm, enabling quantities above 40 pmol to be detected (for *trans*-4-Hyp, *trans*-3-Hyp, Pro, and other analogues). However, this particular method only resolves *cis* and *trans* isomers, and the chiral resolution of all stereoisomers of Hyp and Pro by this method is not possible.³³⁴ Later, another group reported a Hyp/Pro-detection method, which involved NBD derivatisation (Scheme 3.8 a) after a chiral-resolution, ligand-exchange chromatography step;³²⁹ a method based upon this system has since been used to follow the conversions of L-proline to *cis*-3-hydroxy-L-Pro^{22,23,195,201} and *trans*-4-hydroxy-L-Pro^{194,195,197} by *cis*P3Hs^{22,23,195,201} and *trans*P4H,^{194,195,197} respectively. Another derivatisation procedure uses a modified Marfey’s Reagent, 1-fluoro-2,4-dinitrophenyl-5-L-leucine amide (FDLA), to introduce a L-DLA group on the amino nitrogen of amino acids (Scheme 3.8 b);^{330–332} the derivatised amino acids can then be detected using an LC/MS method,^{330–332} and their absolute configuration determined.^{330–332} The FDLA method was originally applied to a test

case, which involved the successful determination of the absolute configuration of 3-amino-6-hydroxy-2-piperidone (Ahp) in aeruginopeptin 228-A.³³⁰ Later, in 2009, the modified Marfey's derivatisation method was used to detect L-proline and *cis*-4-hydroxy-L-proline from *cis*P4H reactions (Scheme 3.8 b);²⁰⁷ reaction components were separated by reverse-phase HPLC and detected by UV-absorption spectroscopy (wavelength 210 nm).²⁰⁷ A method for detecting *N*-containing solutes, such as amino acids, was developed using 9-fluorenylmethyl chloroformate (Fmoc) derivatisation and HPLC.³²⁸ This method was later adapted for the detection of the starting material and hydroxylation products from various proline hydroxylase reactions (Scheme 3.8 c) (including *cis*P3H Type II, *cis*P4H, and (*trans*P4H));^{198,295} Klein *et al.*'s method was also later used to characterise a novel L-proline *trans*-4- and *trans*-3-hydroxylase, GloF.²¹²

3.3 PH Assay Development

Although previous proline hydroxylase assays have been relatively effective detection methods, they were relatively limited in detection sensitivity, since they either required labelled reagents^{53,193,256,320} or amino acid chemical derivatisation steps;^{22,23,194,195,197,198,201,207,295,328–332,334} these methods do not directly detect native, unmodified substrates or products, and may not work for the detection of all products formed (*e.g.* the labelled atom or the group being derivatised may not end up in a final product, or the position may become no longer reactive to derivatisation). Other assays were, therefore, pursued to accomplish the direct detection and quantification of substrates and products from proline hydroxylase reactions.



Scheme 3.8: Proline derivatisation methods for HPLC detection and quantification.

(a) addition of an amino NBD group by treatment with 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole (NBD-Cl) (*i.e.* a nucleophilic substitution reaction);^{22, 23, 194, 195, 197, 201, 329, 334}

(b) addition of an amino L-FDLA motif by treatment with a modified Marfey's Reagent, 1-fluoro-2,4-dinitrophenyl-5-L-leucine amide (FDLA) (in which the L-A of 1-fluoro-2,4-dinitrophenyl-5-L-alanine amide (FDAA) is replaced by L-L);^{207, 330–332}

(c) amino Fmoc-derivatisation by treatment with Fluorenylmethyloxycarbonyl chloride (Fmoc-Cl).^{198, 295, 328}

3.3.1 Proline Hydroxylase NMR Assays

Nuclear magnetic resonance (NMR) experiments are being used increasingly in enzyme assays, because they can provide information on the connectivity, stereochem-

istry, and functional groups.^{335,336} Solution NMR experiments, observing isotopically abundant nuclei, such as those from ^1H and ^{19}F atoms, can be used to reliably quantify substances,^{335,336} since the integrals of a compound's resonance peaks are directly proportional to that substance's concentration.^{335,336} Moreover, no chemical modification of analytes is required for high-resolution NMR spectra to be generated;^{335,336} however, substances are generally analysed in an 'NMR-invisible' solvent (*i.e.* containing no spin- $\frac{1}{2}$ nuclei, such as D_2O , CDCl_3 , *etc.*), so that the signals generated by the solvent do not obscure the resonances from the analytes;^{335,336} NMR solvents are readily available and relatively inexpensive, however. Another benefit of NMR analysis is that sometimes precious samples are not consumed by the process.

3.3.1.1 Proline Hydroxylase End-Point NMR Assays

NMR spectroscopy has been used previously used to follow the reaction of ODD enzymes (for both small-molecule and peptide substrate systems).^{17,337–339} The initial aim of my assay was to establish if NMR could be carefully used to monitor proline hydroxylase activity. Initially, reactions were carried out in Eppendorf tubes, and incubated overnight for ~ 14 h at room temperature; reactions were carried out for prolonged periods to ensure significant enzymatic substrate turnover would take place (*i.e.* producing 'end-point' reactions), in order to increase product-detection sensitivity. Reactions were carried out using conditions based on those used successfully in previous studies;^{22,23,128,191,193,198,207,256,258,295} deuterated Tris- d_{11} -DCl buffer in D_2O was used to facilitate NMR analyses. Eppendorfs tubes containing reactions were incubated in a closed drawer, out of the way of direct light sources; published studies suggest that ascorbate is photosensitive to degradation,^{340–346} and anecdotal observations from the Schofield lab suggest that some 2-OG may also be photosensitive to degradation in some buffered assay solutions. After incubation,

reactions were transferred to NMR tubes and analysed. Initial assays were used to test the activity of a small-scale, trial batch of *cis*P3H (from *Streptomyces* sp. strain TH1); ^1H -NMR spectra measured for samples containing *cis*P3H Type I contained peaks corresponding to the reaction products, *cis*-3-hydroxy-L-proline and succinate (Figure 3.2 a), confirming that the trial enzyme batch was active (hence, the purification method used to isolate the *cis*P3H Type I was effective). NMR analyses of *cis*P3H Type I reactions also showed a reduction in the resonances corresponding to the L-proline substrate (Figure 3.2 a) compared to the ‘no-enzyme’/‘no 2-OG’ negative control reaction (Figure 3.2 b), consistent with L-Pro consumption during enzymatic turnover; 2-OG was omitted from the initial negative control sample in order to simplify its ^1H -spectrum (Figure 3.2 b), and ease the assignment of L-Pro and *cis*-3-OH-L-Pro signals in the reaction mixture. The peaks corresponding to 2-OG were reduced to near base-line levels, consistent with 2-OG being ‘fully’ converted to succinate, which manifests as a clear singlet at $\delta_{\text{H}} = \sim 2.35$ ppm (Figure 3.2 a). Given that the end-point NMR assay worked well for *cis*P3H Type I reactions, it was used as a method for substrate testing in later experiments.

3.3.1.2 PH NMR Timecourse Assay

A continuous, timecourse NMR assay was then pursued; such an assay involves starting a reaction (often by adding enzyme solution), and then measuring spectra at multiple time points to follow the progress of the reaction. Other ODD reactions have been followed successfully using 1D-NMR timecourse methods,^{17,337–339} using peak integration to quantify substrate and product concentrations; under the correct conditions, an enzyme reaction may occur in a manner where the conversion of substrate(s) to product(s) is approximately ‘linear’ with time for a certain time period (so-called ‘linear range’), allowing ‘initial rate’ data to be calculated (from the slope of a ‘product-creation’, concentration-versus-time graph). In some cases,

Samples analysed contained:

(a) a 'complete' *cis*P3H Type I reaction mixture;

- ¹H-NMR spectra were measured at 298 K for 500 μL reactions in 5 mm NMR tubes using a 500 MHz

Type I reactions are used to demonstrate assay behaviour. A high-field 700 MHz Bruker AV III 700 spectrometer (equipped with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe) was used to analyse samples for maximum detection sensitivity and high signal-to-noise ratios; samples were analysed in 3 mm NMR tubes. Reactions typically used an excess of Fe^{2+} with respect enzyme concentration (at least ~ 5 - 10 -fold), in order to saturate all enzyme iron(II)-binding sites, and maximise catalytic activity. Fe^{2+} concentrations above $100\text{ }\mu\text{M}$ were not used, however, since excessive salt concentrations in NMR samples can lead to induced currents within the sample tube,^{335,336} leading to a loss of spectral resolution and measurement sensitivity;^{335,336} high-spin $[\text{Fe}(\text{H}_2\text{O})_6]^{2+}$ ions and $[\text{Fe}(\text{H}_2\text{O})_6]^{3+}$ ions (the oxidation product of $\text{Fe}^{2+}_{(\text{aq})}$) are also paramagnetic³⁴⁷ and cause spectral line-broadening,^{335,336} as well as a loss in spectral resolution and measurement sensitivity.^{335,336} NMR timecourse assays also used an excess of the reducing agent sodium ascorbate with respect to $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ concentrations, in order to maintain ferrous iron(II) concentrations and optimise proline hydroxylase catalysis, as added ascorbate does for many, but not all, ODDs.^{18,40} Integrals from the 1.50-3.68 ppm region of the proton spectra (Figure 3.3) measured were used for calculating concentrations, since the baseline is flat in this region and line shapes are good; at higher ppm values, there are other peaks from ascorbate and solvent (HDO peak is suppressed but still broad), which, in some cases, overlap with L-Pro/*cis*-3-OH-L-Pro peaks. For example, the β -proton peak (~ 4.6 ppm) of *cis*-3-OH-L-Pro (H_3 in Figure 3.3, respectively) overlaps with the HDO solvent peak (centred about ~ 4.7 ppm), and so its integral is difficult to determine accurately; similarly, the α -protons of L-Pro (4.12 ppm) and *cis*-3-OH-L-Pro (4.14 ppm) (H_2 and H_2 in Figure 3.3) are also overlapping slightly, and, therefore, peak integration is difficult for both of these peaks.

At relatively high concentrations of *cis*P3H (5 - $15\text{ }\mu\text{M}$) the reaction profile generally plateaus quite quickly (within the first ~ 500 - $1000\text{ s} \equiv 8.3 - 16.6\text{ min}$) (Figures

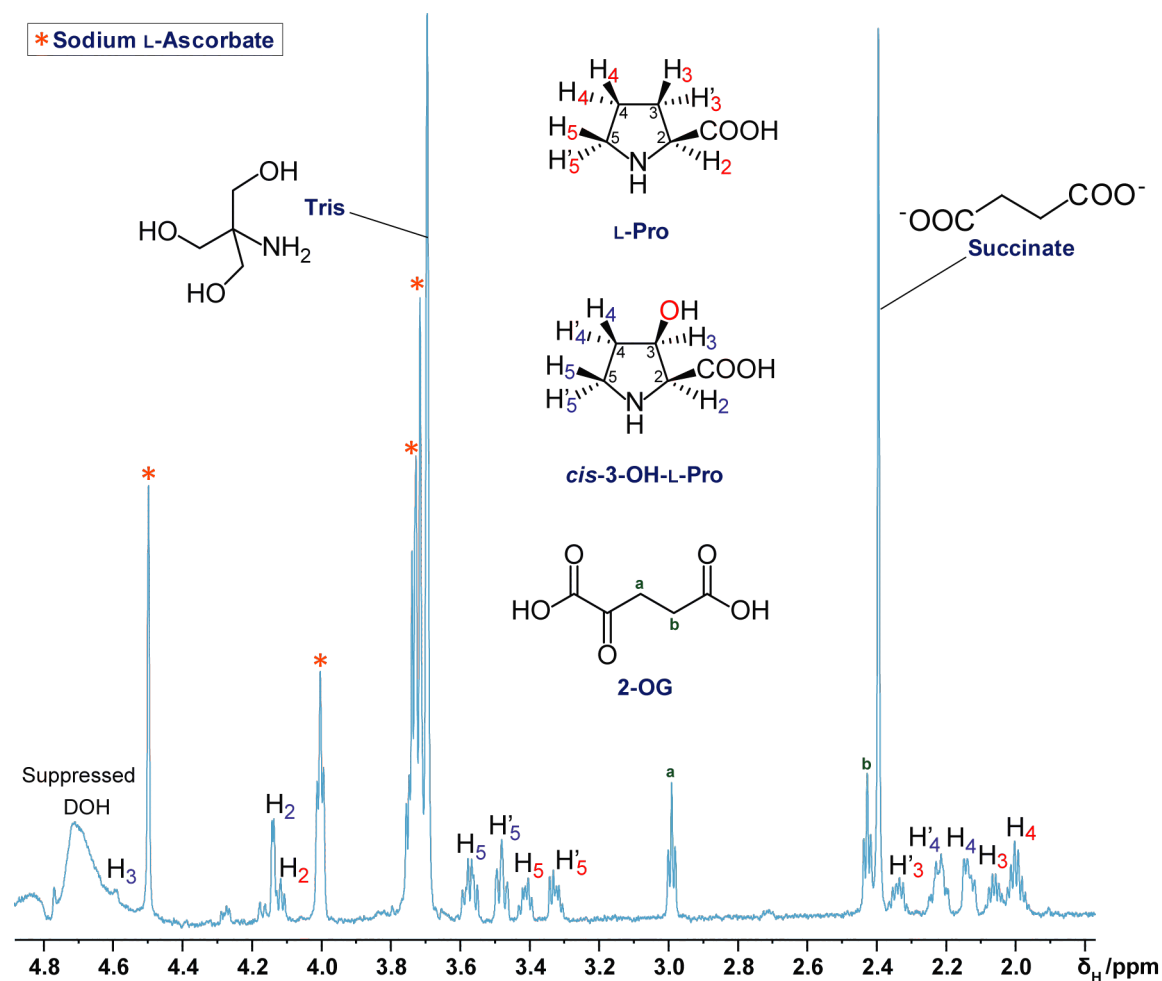


Figure 3.3: Example *cis*P3H Type I ^1H -NMR timecourse spectrum.

This spectrum (16 scans) was measured at 298 K on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe) with a sample in 3 mm NMR tubes; presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Chemical shifts were then referenced to DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). Reactions were carried in 50 mM Tris- d_{11} -DCl at room temperature (25 °C).

The reaction conditions for this particular assay were: 250 μM L-Pro, 250 μM 2-OG, 50 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 0.5 mM sodium ascorbate, and 1.0 μM *cis*P3H Type I. This spectrum time was taken 5375 s into the reaction.

3.4, 3.5, and 3.6). For example, if moderate concentrations (1.0 mM) of L-Pro and 2-OG are used in a timecourse experiment containing 15 μM *cis*P3H Type I, *cis*-3-OH-L-Pro and succinate levels reach a relatively modest maximum of ~ 170 μM and 300 μM (Figure 3.4), respectively, after ~ 1000 s; similarly, under conditions using 1.0 mM L-Pro, 1.0 mM 2-OG, and 10 μM *cis*P3H Type I, concentrations of both

cis-3-OH-L-Pro and succinate peak at roughly $\sim 850 \mu\text{M}$ (Figure 3.5) after ~ 1000 s. However, increasing the concentration of *cis*P3H Type I from $10 \mu\text{M}$ to $15 \mu\text{M}$ increases substrate-uncoupled 2-OG turnover (*i.e.* 2-OG turnover which does not lead to L-proline substrate hydroxylation) from $\sim 2\%$ to $\sim 28\%$. Thus, the preliminary analyses suggested that the catalytic efficiency of proline hydroxylation may be reduced when the substrate-to-enzyme ratio is reduced beyond a threshold level. The reason for this unclear, but may reflect protein oligomerisation. Timecourses which used high concentrations (5.0 mM) of both substrates and relatively high ($5.0 \mu\text{M}$) *cis*P3H Type I concentrations reached peak turnover even faster ($\sim 500 \text{ s}$ - roughly twice as fast) than those with lower substrate concentrations (1.0 mM) (but similar enzyme concentrations); concentrations of *cis*-3-OH-L-Pro and succinate level out at $\sim 300 \mu\text{M}$ and $\sim 330 \mu\text{M}$ (Figure 3.6), respectively, and substrate-uncoupled turnover was $\sim 4\%$. It is possible that product formation ‘stops’ quickly in the NMR tubes owing to limiting availability of molecular oxygen; NMR tubes are relatively long and thin, causing diffusion of molecular oxygen into the bulk reaction mixture to be insufficient for optimal catalysis in some ODD reactions,³⁴⁸ particularly if they are very active (as is the case for a number of small-molecule hydroxylases).^{17,207,338} Given that catalysis would be particularly fast for reactions with high enzyme concentrations and high substrate concentrations, NMR tubes containing reactions with these types of conditions would quickly consume and deplete the local oxygen in the reaction mixture, slowing catalysis.

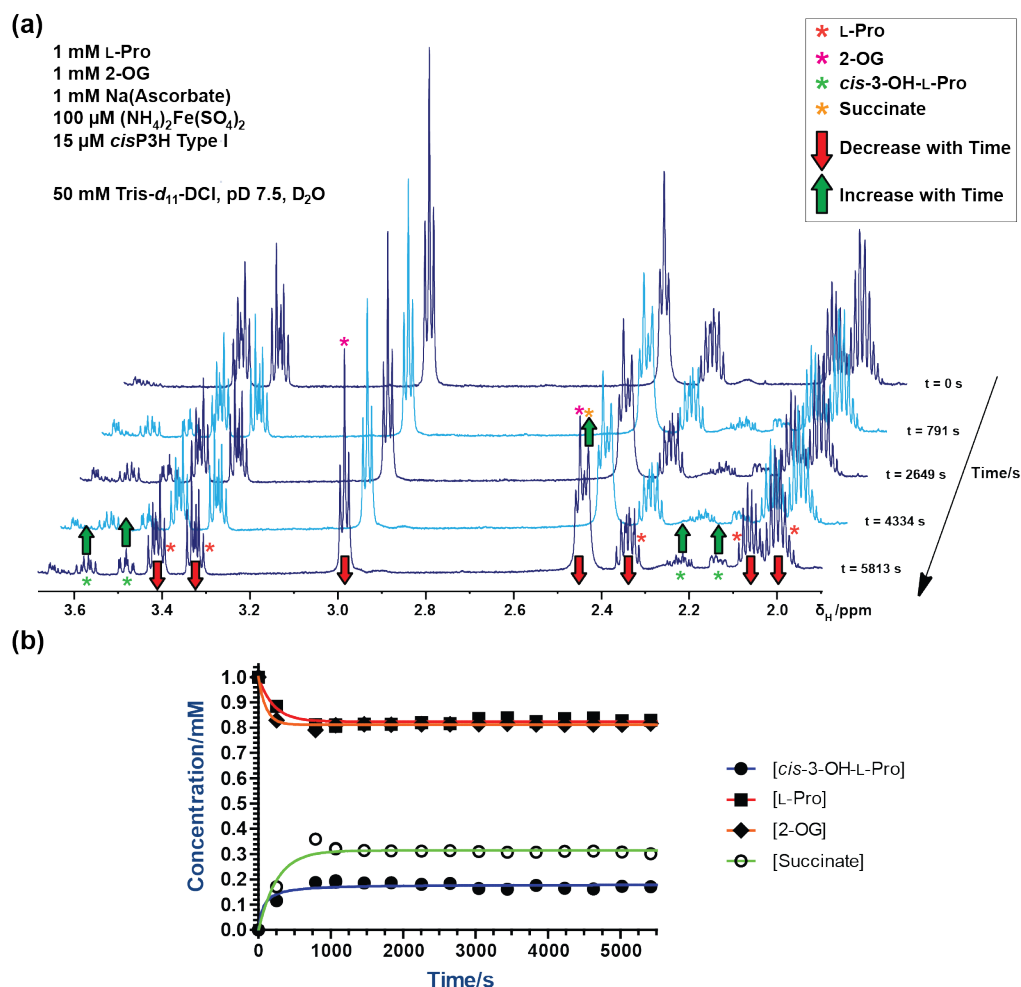


Figure 3.4: *cis*P3H Type I NMR Timecourse experiment using a high concentration (15 μ M) of *cis*P3H Type I and moderate substrate concentrations (1.0 mM L-Pro and 1.0 mM succinate).

(a) Proton NMR spectra (in the 1.50-3.68 ppm range) which show how the reaction mixture alters with time.

(b) Variations in substrate and product concentrations with time.

Spectra were measured at 298 K on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe) with samples in 3 mm NMR tubes; presaturation (with O1P = 4.701 ppm) was used to suppress broad, HDO-solvent peaks. Chemical shifts were then referenced to DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). Reactions were carried in 50 mM Tris- d_{11} -DCl at room temperature (25 °C).

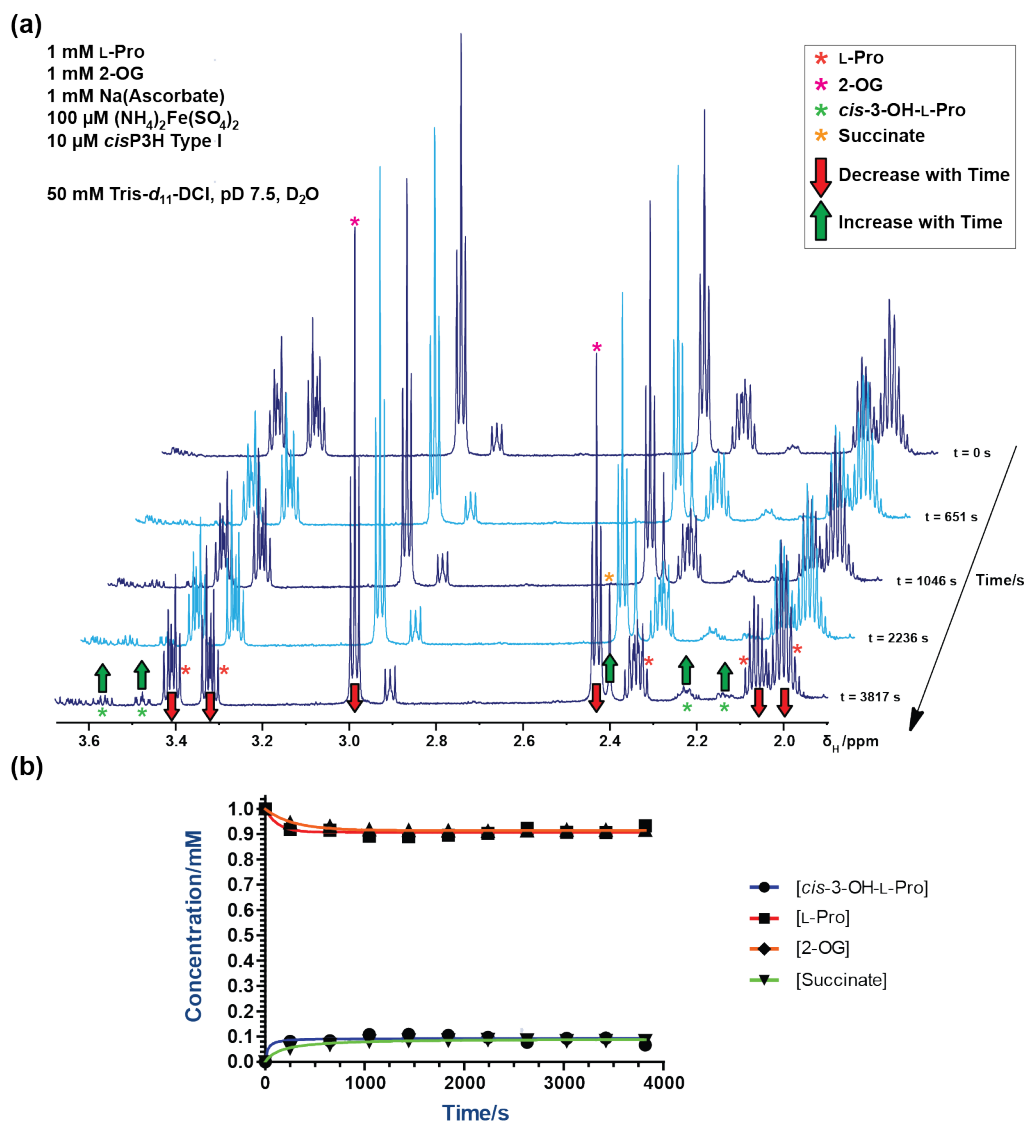


Figure 3.5: *cis*P3H Type I NMR Timecourse experiment using a high concentration (10 μ M) of *cis*P3H Type I and moderate substrate concentrations (1.0 mM L-Pro and 1.0 mM succinate).

(a) Proton NMR spectra (in the 1.50-3.68 ppm range) which show how the reaction mixture alters with time.

(b) Variations in substrate and product concentrations with time.

Spectra were measured at 298 K on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe) with samples in 3 mm NMR tubes; presaturation (with O1P = 4.701 ppm) was used to suppress broad, HDO-solvent peaks. Chemical shifts were then referenced to DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). Reactions were carried in 50 mM Tris- d_{11} -DCl at room temperature (25 $^{\circ}\text{C}$).

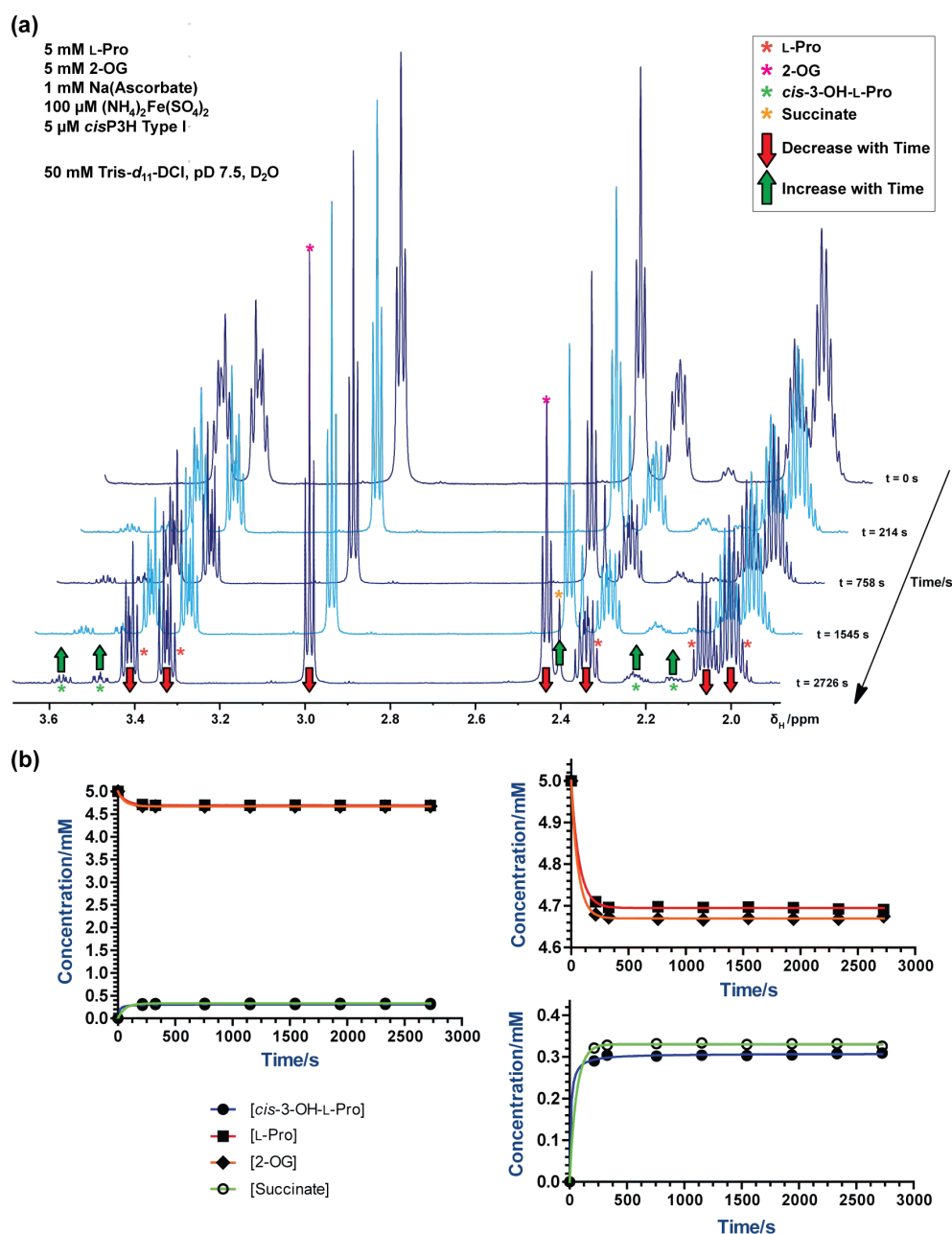


Figure 3.6: *cis*P3H Type I NMR Timecourse experiment using a high concentration (15 μM) of *cis*P3H Type I and moderate substrate concentrations (5.0 mM L-Pro and 5.0 mM succinate).

(a) Proton NMR spectra (in the 1.50-3.68 ppm range) which show how the reaction mixture alters with time.

(b) Variations in substrate and product concentrations with time.

Spectra were measured at 298 K on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe) with samples in 3 mm NMR tubes; presaturation (with O1P = 4.701 ppm) was used to suppress broad, HDO-solvent peaks. Chemical shifts were then referenced to DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid). Reactions were carried in 50 mM Tris- d_{11} -DCl at room temperature (25 $^\circ\text{C}$).

If the concentrations of reaction substrates and *cis*P3H Type I are lowered to modest levels (such as 250 μ M L-Pro, 250 μ M 2-OG, and 1.0/0.5 μ M *cis*P3H Type I), the reaction rates are lowered enough such that molecular oxygen consumption by the *cis*P3H Type I reaction is lowered enough that these reactions can continue for longer than those using higher substrate concentrations (1.0-5.0 mM); local oxygen concentrations in the NMR tube take much longer to drop to a level which is limiting to catalysis, and turnover ‘plateaus’ much later than in reactions using the higher substrate concentrations (resulting in higher percentage substrate turnovers). For example, *cis*P3H Type I reactions using 250 μ M L-Pro, 250 μ M 2-OG, and 1.0 μ M *cis*P3H Type I reach maximum turnover at ~ 5000 s ($\equiv 1$ h 23 min 20 s) (Figure 3.7); succinate and *cis*-3-OH-L-Pro levels reach a maximum of 145 μ M with an initial uncoupled turnover of ~ 3 s. Similarly, under conditions of 250 μ M L-Pro, 250 μ M 2-OG, and 0.5 μ M *cis*P3H Type I, maximum turnover peaked at ~ 5000 s, with ~ 150 μ M succinate and ~ 170 μ M *cis*-3-L-Hyp being produced (Figure 3.8). Given that, in this case, the maximal succinate concentration was slightly lower (by ~ 20 μ M) than the *cis*-3-OH-L-Pro concentration, it is probable that the errors of peak integration are higher than those for reactions using higher substrate concentrations (owing to signal:noise considerations). If the *cis*P3H Type I obeys ‘current mechanistic models’ for 2-OG oxygenases, succinate concentrations should be greater or equal to that of the hydroxylated product, since 2-OG turnover is supposed to produce the ferryl species [Fe(IV)=O], which hydroxylates the L-proline substrate.^{13–15,38} The limited sensitivity of NMR spectroscopy compared with some other analytical techniques^{335,336} means it can be difficult to resolve small concentration differences; measurements of substances at concentrations of ~ 150 μ M (as in Figure 3.8) with this degree error (up to 20 μ M) could have a $\sim 13\%$ error, which is relatively significant.

The continuous NMR timecourse assay can be limited by oxygen availability;

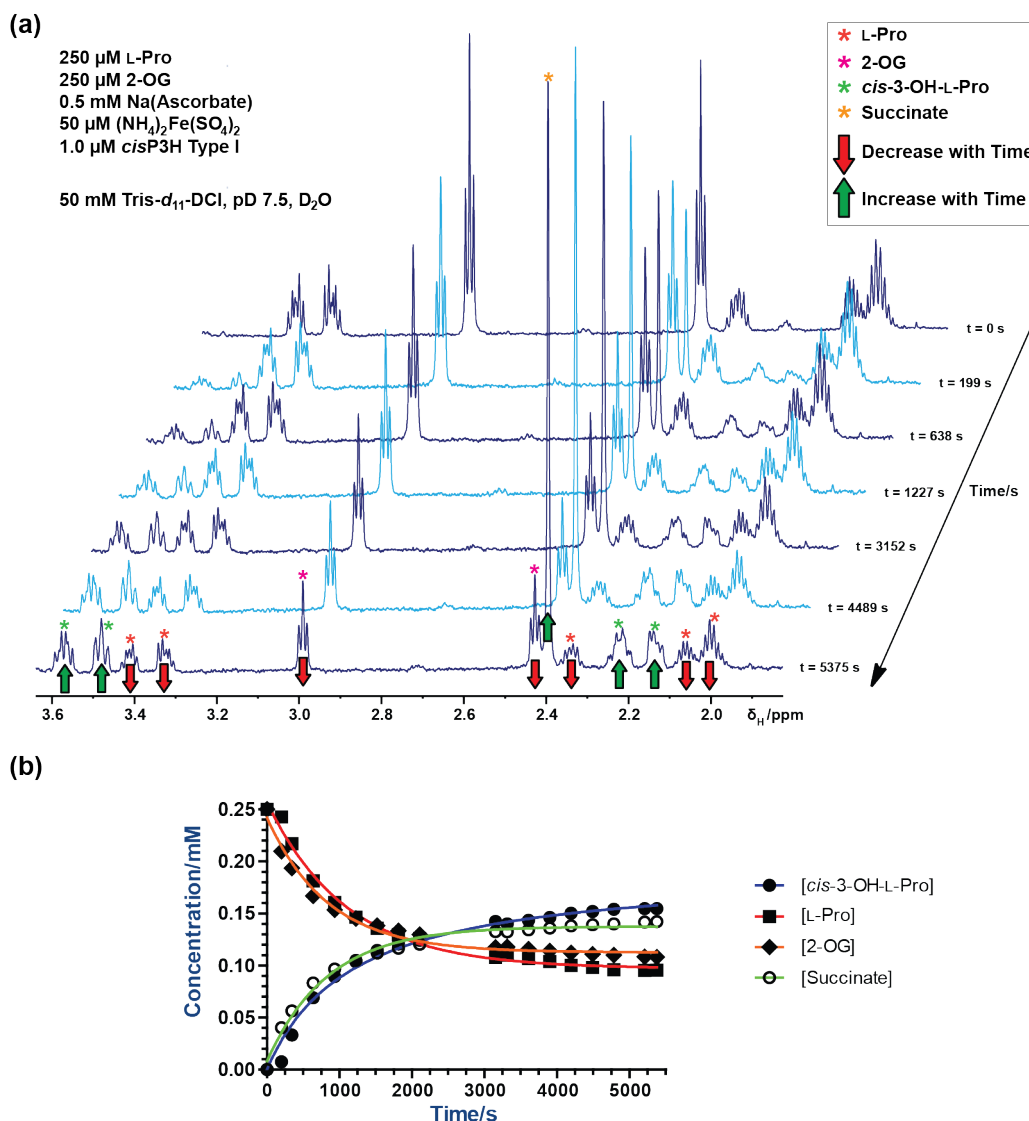


Figure 3.7: *cis*P3H Type I NMR Timecourse experiment using a high concentration (1.0 μM) of *cis*P3H Type I and moderate substrate concentrations (250 μM L-Pro and 250 μM succinate).

(a) Proton NMR spectra (in the 1.50-3.68 ppm range) which show how the reaction mixture alters with time.

(b) Variations in substrate and product concentrations with time.

Spectra were measured at 298 K on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe) with samples in 3 mm NMR tubes. Reactions were carried in 50 mM Tris- d_{11} -DCl at room temperature (25 $^\circ\text{C}$).

moreover, conditions which do not have significant issues with this problem (*i.e.* conditions with lower concentrations of substrate and enzyme) could produce lower accuracy concentration measurements, owing to a decrease in peak signal-to-noise

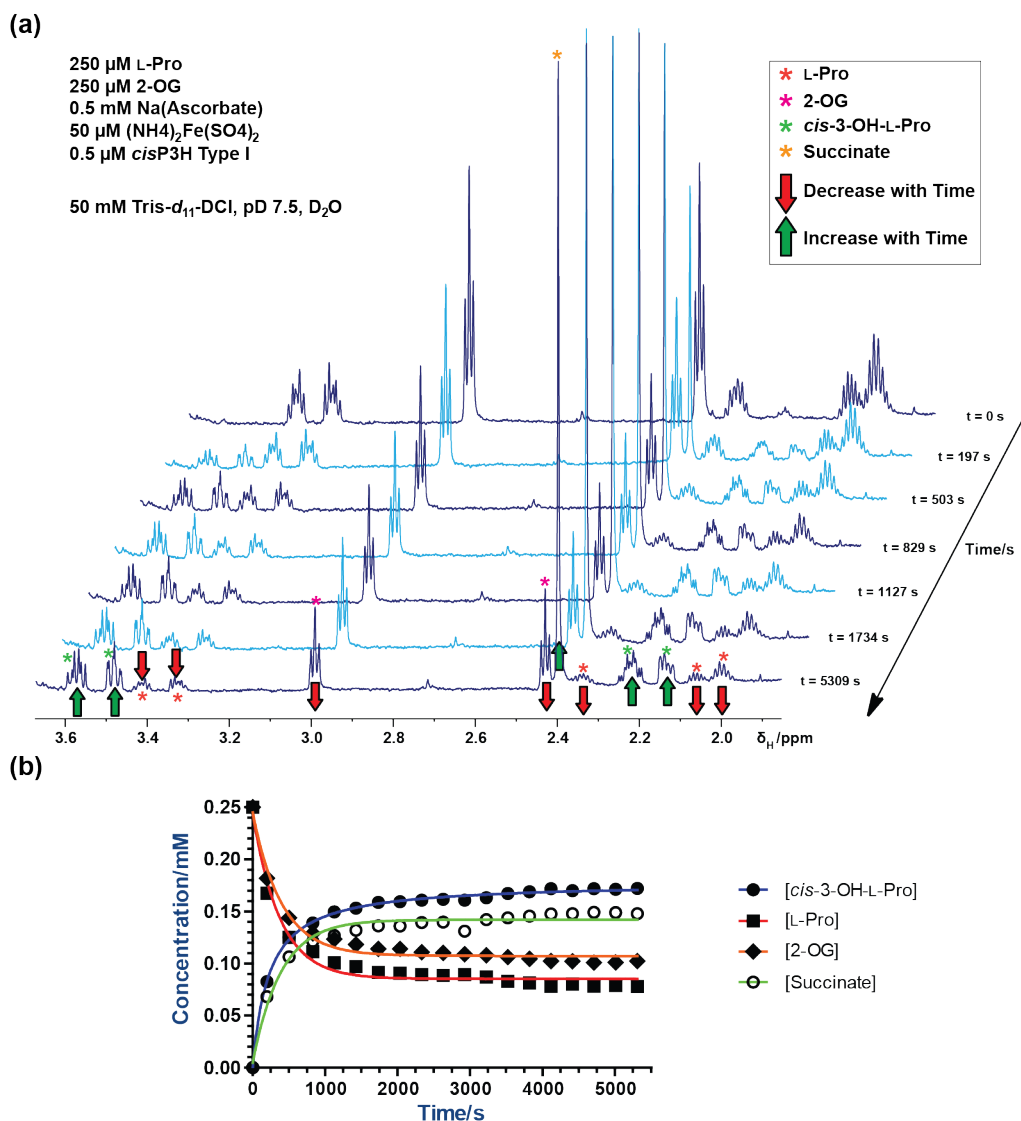


Figure 3.8: *cis*P3H Type I NMR Timecourse experiment using a high concentration (0.5 μM) of *cis*P3H Type I and moderate substrate concentrations (250 μM L-Pro and 250 μM succinate).

(a) Proton NMR spectra (in the 1.50-3.68 ppm range) which show how the reaction mixture alters with time.

(b) Variations in substrate and product concentrations with time.

Spectra were measured at 298 K on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe) with samples in 3 mm NMR tubes. Reactions were carried in 50 mM Tris- d_{11} -DCl at room temperature (25 $^\circ\text{C}$).

ratios and peak integration errors. These issues would make using this proline hydroxylase NMR timecourse assay potentially unsuitable for kinetic parameter analysis experiments, or, at the very least, the experiments would be challenging.

3.3.2 LC/MS-Based PH Assays

Liquid Chromatography-Mass Spectrometry (LC/MS) has become a routine technique for the detection of a wide range of biological molecules with the development of electrospray ionisation (ESI);^{349–353} the use of high-performance liquid chromatography (HPLC) combined with ESI mass spectrometry has proven to be a highly sensitive and accurate method of detection, and is widely used for biological assays and drug discovery projects.^{349–353} Although MS methods ‘consume’ samples, very little material (typically ng to μg) is required for LC/ESI-MS detection, allowing most of a sample to be retained.^{349–353} It was therefore thought that a sensitive, LC/MS-based proline hydroxylase assay may be an accurate way to detect and quantify proline hydroxylase reaction products and starting materials; for effective column retention of these polar amino acid analytes (e.g. L-proline and L-hydroxyprolines), an appropriate HPLC column and separation conditions would be necessary. It was thought that a method employing a mixed-mode HPLC column may be useful and various options were researched.

SiELC Technologies PrimesepTM HPLC columns are designed for mixed-mode separations of a wide range of polar and non-polar compounds; different separation modes can be specified based upon the mobile phase used (*i.e.* solvent formulations).³⁵⁴ These versatile columns efficiently separate organic and inorganic ions in ion-exchange and ion-exclusion modes, and can be used to provide efficient separation in normal phase, reverse phase, and polar organic modes; different column separation modes provide different analyte-binding selectivities, allowing methods to be tailored for specific requirements.³⁵⁴ In most cases, separations require a simple mobile phase containing acetonitrile, water, and trifluoroacetic acid (TFA) or formic acid (FA) [the latter of which is compatible with common detection methods, such as MS, a Evaporating Light Scattering Detector (ELSD), UV-absorption,

and refractive index (RI) detection]; binding modes can be simply ‘tuned’ by varying organic modifier concentrations (like in conventional reverse-phase chromatography), and, additionally, by altering acid modifier types (*e.g.* FA) and concentrations.³⁵⁴ Primesep columns are suitable for both analytical and preparative scale separations in isocratic and gradient modes, and separations are easily scaled up;³⁵⁴ these columns are resistant to dewetting in 100% aqueous eluent, and are stable in pure organic and highly acidic conditions down to pH 1.5, which aids the separation of highly polar compounds.³⁵⁴ Any silanol or metal chelating interactions are also completely eliminated and do not affect the efficiency of the separation.³⁵⁴ In conventional ion-pairing, reverse-phase chromatography, column retention of polar, ionisable species is controlled by the concentration and the type of ion-pairing reagents used, *e.g.* 1-pentanesulfonic acid [$\text{CH}_3(\text{CH}_2)_3\text{CH}_2\text{SO}_2\text{OH}$], heptanesulfonic acid [$\text{CH}_3(\text{CH}_2)_5\text{CH}_2\text{SO}_2\text{OH}$], sodium dodecylsulfate (SDS), tetrabutylammonium hydroxide [$\text{N}(\text{n-Bu})_4\text{OH}$];³⁵⁵ by analogy, PrimesepTM HPLC mixed-mode columns have embedded ion-pairing groups in their stationary phases and require no ion-pairing reagents in the mobile phase to retain and separate ionisable polar compounds.³⁵⁴ For example, there are columns with different strengths of ion-bearing groups for cation exchange modes (decreasing retention strength order: PrimesepTM A, 100, 200, 300, and C) and for anion-exchange modes (decreasing retention strength order: Primesep B and B2);³⁵⁴ PrimesepTM C columns are specifically designed to form a strong complex with amines.³⁵⁴ PrimesepTM columns also use ‘SWITCH PhaseTM Technology’, a system where the column properties change depending on the mobile phase (solvent) pH.³⁵⁴ Resin-embedded carboxylic acids are fully ionised at a pH above ($\text{pH} > \text{pK}_a$) a transition point (*i.e.* less acidic), and the columns operate as a mixed-mode (ion-exchange and reverse-phase) column (Figure 3.9 b);³⁵⁴ the same resin-embedded carboxylic acid group lose charge when the mobile phase pH goes below ($\text{pH} > \text{pK}_a$) the transition point (*i.e.* more

acidic) (Figure 3.9 b).³⁵⁴ At the transition pH value ($\text{pH} = \text{pK}_a$), and the columns will function as a reverse-phase column with a non-ionized functional group (Figure 3.9 b).³⁵⁴ Very close to the transition pH ($\text{pH} \approx \text{pK}_a$), ~50% of the PrimesepTM embedded carboxylic acid groups are ionised.³⁵⁴

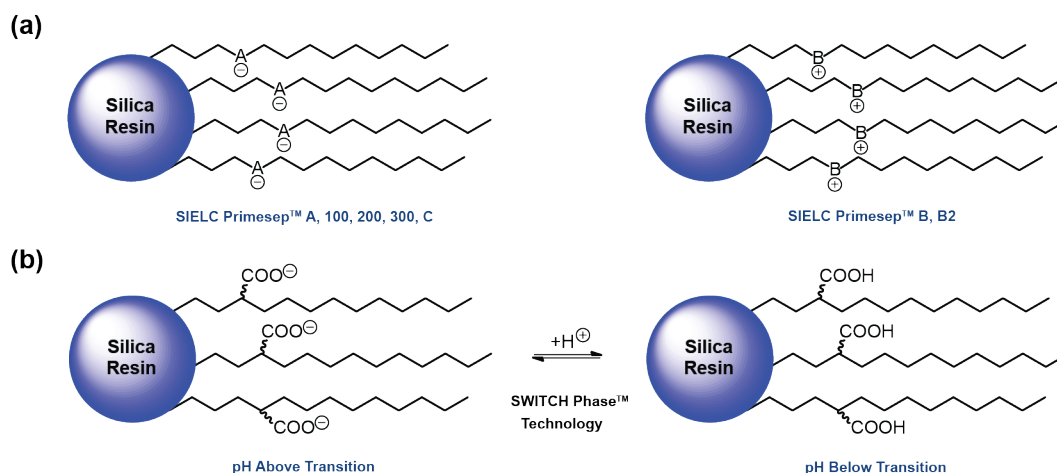


Figure 3.9: The architecture and properties of SiELC Technologies PrimesepTM HPLC columns:

(a) PrimesepTM columns contain a resin with ion-bearing groups to facilitate ion-exchange modes, including both cation-exchange (using embedded anions - PrimesepTM A, 100, 200, 300, C) and anion-exchange (using embedded cations - PrimesepTM B, B2);

(b) the ionisation states of embedded carboxylic acid groups can be altered with a change in pH over a pH threshold value (pH 0.0, 1.0, 2.0, 3.0, 3.5 for PrimesepTM A, 100, 200, 300, and C), and be ‘switched’ between operating as a mixed-mode column (*i.e.* reverse-phase and ion-exchange) or a reverse-phase column.

The most universal column of the SiEL PrimesepTM range is the PrimesepTM 100 column, which can be used for the separation of basic, acidic, neutral, or zwitterionic compounds; acids are separated owing to ion exclusion, and bases are controlled by cationic exchange. PrimesepTM 100 columns are therefore very useful for the isolation and separation of amino acids (and related compounds) from reactions mixtures and enzyme assays;³⁵⁴ these methods would not require any sample derivatisation for HPLC separation, allowing for direct detection of reaction products and starting materials. Methods using SiELC PrimesepTM 100 columns have been used for the separation of amino acid compounds similar to proline

from enzyme reaction mixtures (Figure 3.10 a).³⁵⁶⁻³⁶¹ For example, variants of *cis/trans*-5-(2-carboxymethyl)-L-proline (*c/t*-CMP) have been detected from Carboxymethylproline Synthase (CarB)³⁵⁶⁻³⁶⁰ and ThnE³⁶¹ enzymatic reactions using PrimesepTM 100 LC/MS methods; these enzymes are involved in the biosynthesis of carbapenem (β -lactam) antibiotics, namely 1-carbapen-2-em-3-carboxylic acid (Car)^{29,362,363} and thienamycin,^{364,365} respectively. Under the HPLC conditions used in these studies [*i.e.* water (H₂O)/acetonitrile (CH₃CN) with 0.05-0.1% (v/v) FA - pH ~2.0], the SiELC PrimesepTM 100 column behaves as a mixed-mode (reverse-phase and ion-exchange) column;^{354,356-361} the non-polar motifs (*e.g.* alkyl chains, aromatic rings *etc.*) of analytes are retained using hydrophobic interactions (*i.e.* reverse-phase interactions) with long, aliphatic chains in the column resin (Figure 3.10 b),³⁵⁴ whilst positively charged motifs (*e.g.* protonated amines) are retained using electrostatic interactions (*i.e.* cation-exchange interactions) with negatively charged groups embedded in the stationary phase (Figure 3.10 b).³⁵⁴ Many of these methods incorporated a stable internal standard (Figure 3.10 c) for the accurate quantification of starting materials and products;³⁵⁶⁻³⁶¹ the compound used (Figure 3.10 c), *para*-amino-salicylic acid (*p*ASA), contains similar functional groups (*e.g.* -NH₂, -COOH) and, thus, has similar retention properties to the analytes being studied.³⁵⁶⁻³⁶¹

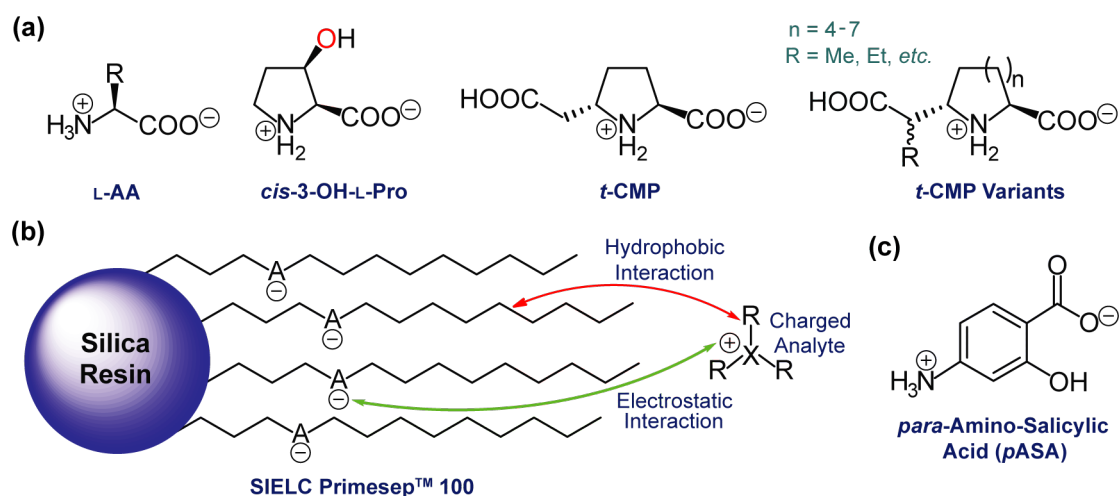


Figure 3.10: SiELC Primesep™ 100 column applications: (a) typical analytes retained by the column; (b) a schematic of the mixed-mode (reverse-phase + cation-exchange) mechanism of analyte retention; (c) a useful *para*-amino-salicylic acid (pASA) internal standard for LC/MS-quantification of analytes.

An LC/MS method for detection of L-proline, hydroxyprolines, and related products was developed based heavily upon work carried out by Batchelar *et al.*³⁵⁶ and Hamed *et al.*^{357–361} A 40 min HPLC method was developed, which uses a 4.6 mm $\times$ 250 mm SiELC Primesep™ 100 column (100 Å, 5 μ m) and a mobile phase containing water, acetonitrile, and 0.1% FA; analyte samples are typically bound to the column in a low-concentration of acetonitrile [$\sim 5\%$ (v/v) CH_3CN] and then eluted with a linear gradient of increasing acetonitrile (30%–70%). After the gradient elution step, the column is subsequently washed with 100% CH_3CN [with 0.1% (v/v) FA] to remove any remaining column-bound compounds, and then re-equilibrated with sample-binding solvent. The downstream flow of eluents from the column are then passed through a splitter and then analysed by ESI-MS; ESI a relatively mild ionisation method, allowing most analyte compounds to remain largely intact (giving prominent molecular ion peaks).^{349–351}

In vitro reactions (100 μ L) were carried out with purified *cis*P3H Type I, *cis*P4H, and *trans*P4H; initial reaction conditions [50 mM MES-NaOH, pH 6.5, 15 μ M enzyme (~ 0.5 mg/mL), 1 mM L-Pro, 10 mM 2-OG, 0.5 mM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM

sodium ascorbate] were based, in part, on those used effectively before,^{198,258,295} and concentrations were slightly adjusted to compliment the LC/MS detection methods. A large excess of 2-OG (compared with L-Pro and enzyme concentrations) was used to optimise L-Pro substrate turnover (consistent with previous work),^{198,295} and an excess of Fe²⁺ was added with respect to enzyme concentrations (~15 μ M) to saturate all iron(II)-binding sites to maximise catalytically active enzyme levels. An excess of sodium ascorbate (typically 1 mM-1.5 mM) was used with respect to the Fe(II) levels in order to maintain ferrous iron levels and optimise enzyme catalysis. Reactions were incubated at room temperature rather than 37 °C to stabilise the enzymes to degradation/inactivation, as well as stabilising ferrous iron and ascorbate with respect to oxidation; this should extend the length of the reaction with only slight decrease in initial rate for *cis*P3H Type I and *trans*P4H (*cis*P3H Type I and *trans*P4H have been reported to react slightly faster at 37 °C than at room temperature^{198,258,295} - in contrast, room temperature is the optimal temperature for *cis*P4H turnover^{198,207,295}). Reactions were quenched with cold methanol containing 0.25 mM *para*-amino-salicylic acid (*p*ASA - internal standard) in a 1:1 mixture (by volume - 100 μ L reaction + 100 μ L MeOH); samples were then incubated on ice for 10-15 min whilst the proline hydroxylases precipitated, and then samples were clarified using high-speed microcentrifugation.

Analysis of the LC/MS chromatograms measured for proline hydroxylase reactions revealed that the proline hydroxylases (*cis*P3H Type I, *cis*P4H, and *trans*P4H) consumed all (*cis*P3H Type I, *cis*P4H) or the vast majority (*trans*P4H) of the L-proline in a ~2 h period, and converted the L-Pro into the corresponding hydroxyprolines (Figures 3.11 a-c); extracted-ion count (XIC) chromatograms (showing 132.1 *m/z* ions) contain large peaks corresponding to the protonated molecular ions ($[M+H]^+$) of hydroxyprolines (132.1 *m/z*) (Figures 3.11 a-c). The retention times of *cis*-3-OH-L-Pro, *cis*-4-OH-L-Pro, and *trans*-4-OH-L-Pro (Figures 3.11 a-c)

were ~14 min, ~13 min, and ~10.5 min, respectively. Consistent with these findings, extracted-ion count (XIC) chromatograms (showing 154.1 m/z and 116.1 m/z ions) from LC/MS-analysed PH reactions contain either no peak corresponding to protonated proline ($[M+H]^+$ - 116.1 m/z) (in the case of *cis*P3H Type I and *trans*P4H reactions) or a very small one (in the case of *trans*P4H reactions), which is consistent with L-Pro being heavily consumed by the PH reactions (Figures 3.12 a-c); the *p*ASA internal standard, however, produces a prominent peak with a retention time of ~14.5 min and of the expected size (Figures 3.12 a-c). Negative control reactions (with omitted enzymes) showed no turnover, and contained only L-proline starting and *p*ASA starting material; this is consistent with the hydroxylation reactions being enzyme catalysed. As well as observing the previously reported wild-type behaviour of *trans*P4H,^{191, 193, 195, 256} there appears to be an additional (2nd) small peak ('bump') in the 132.1 m/z XIC for the *trans*P4H reaction (marked with a '*' - Figure 3.11 c); this small peak has a retention time of ~13 min, the same as that seen for *cis*-4-OH-L-Pro in the *cis*P4H reaction, suggesting that *trans*P4H can, to some extent, produce a very small amount of *cis*-4-OH-L-Pro, in addition to the *trans*-4-OH-L-Pro. However, further investigation is required in order to determine whether this is a 'true' indication of mixed stereoselectivity for *trans*P4H and not just an artifact of the LC/MS method (e.g. the tapering of the *trans*-4-OH-L-Pro peak owing to directional solvent flow in the HPLC methods). This result, if validated, would be the first detection of *cis*-4-hydroxylation by a *trans*-4-hydroxylase.

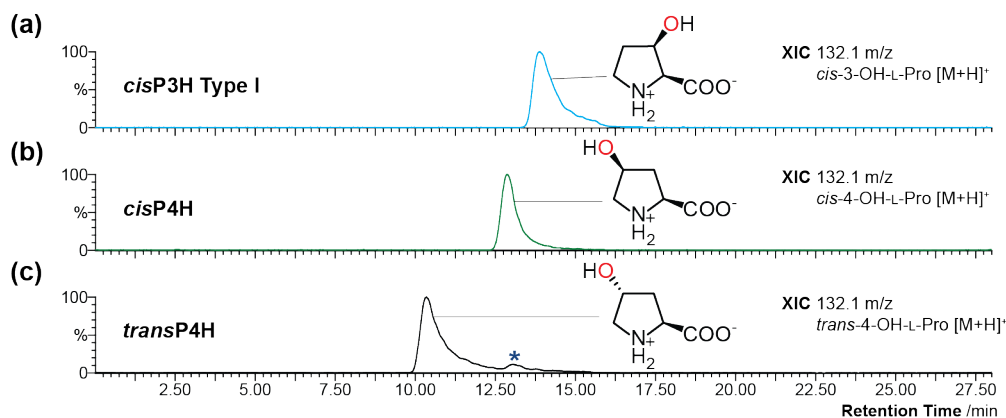


Figure 3.11: LC/MS-based analysis of proline hydroxylase reactions with L-Pro [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H].

Extracted-ion chromatograms (XICs) (showing 132.1 m/z ions) contain peaks corresponding to (a) *cis*-3-OH-L-Pro, (b) *cis*-4-OH-L-Pro, and (c) *trans*-4-OH-L-Pro, respectively.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 10 mM 2-OG, 1 mM L-Pro. 100 μ L reactions were carried out in at room temperature (25 $^\circ$ C) for 2 h in 50 mM MES-NaOH, pH 6.5 buffer.

Complete reactions were quenched with 100 μ L methanol (1:1) containing a *para*-aminosalicylic acid (0.25 mM *p*ASA) standard, and then microcentrifuged to remove precipitate. Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 $\text{\AA}$, 5 μ m) [4.6 x 250mm].

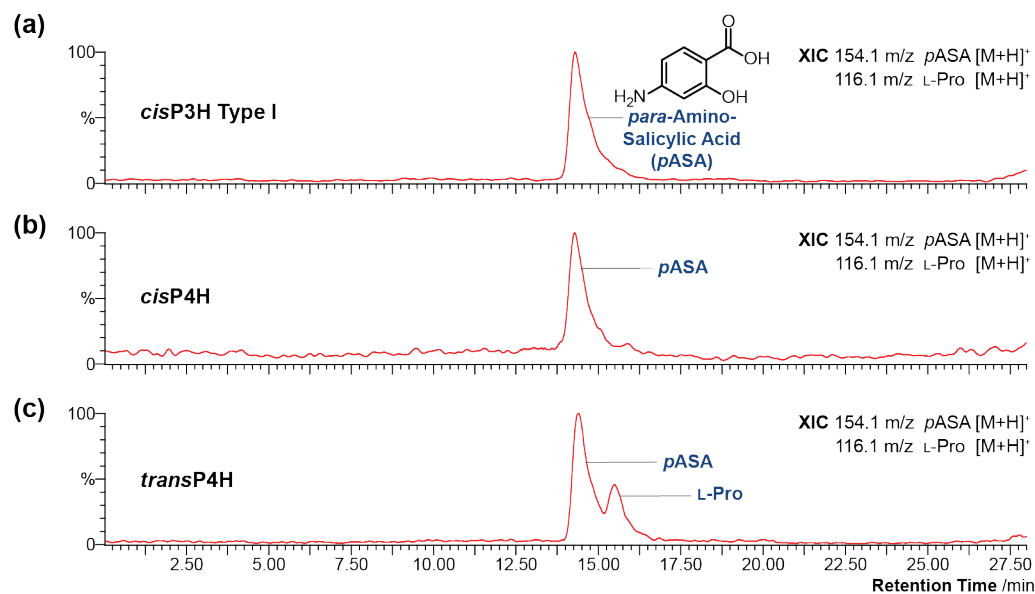


Figure 3.12: LC/MS-based analysis of proline hydroxylase reactions with L-Pro [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H].

Extracted-ion chromatograms (XICs) (showing 154.1 m/z and 116.1 m/z ions) contain peaks corresponding to *para*-amino-salicylic acid (*p*ASA) (internal standard) and L-Pro.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM L-Pro. 100 μ L reactions were carried out in at room temperature (25 °C) for 2 h in 50 mM MES-NaOH, pH 6.5 buffer.

3.4 PH Substrate Analogue Screening

A common property of dioxygenases (including many ODDs) is that many of these enzymes have relatively low substrate selectivities;^{13,15} many of these enzymes have been shown to accept substrates which are structurally related to their natural substrates, and to produce compounds of biological^{101,102} or pharmaceutical importance.^{13,15} In addition to the many alternative hydroxylation substrates which have been found, these ODD enzymes have been shown to catalyse the epoxidation of alkenes,^{139,366} the sulfoxidation of sulfides,³⁶⁷ the demethylation/dealkylation *N*^ε-alkyl-lysyl histone residues,³⁶⁸ and the ketenisation of alkyne functionalities³⁶⁹ present in unnatural substrates.^{13,15,370}

3.4.1 Substrate Screening by [*U*-¹⁴C]-2-OG Turnover

Pioneering studies carried out by Lawrence *et al.* initiated a search for alternative proline hydroxylase substrates;^{191,193,256} these preliminary studies employed an *in vitro* radioactive-[*U*-¹⁴C]-2-OG-turnover assay, which followed the oxidative decarboxylation of [*U*-¹⁴C]-2-OG to [¹⁴C]-CO₂ and [*U*-¹⁴C]-succinate by *S. griseoviridis* P8648 *trans*P4H using a scintillation counter.^{191,193,256} L-Proline analogues which are possible *trans*P4H substrates should stimulate 2-OG turnover; various L-proline analogues were screened, including unsaturated compounds, compounds containing heteroatoms, compounds of varying ring size, and small peptides containing L-Pro residues (Figure 3.13 and Table 3.1).^{191,193,256}

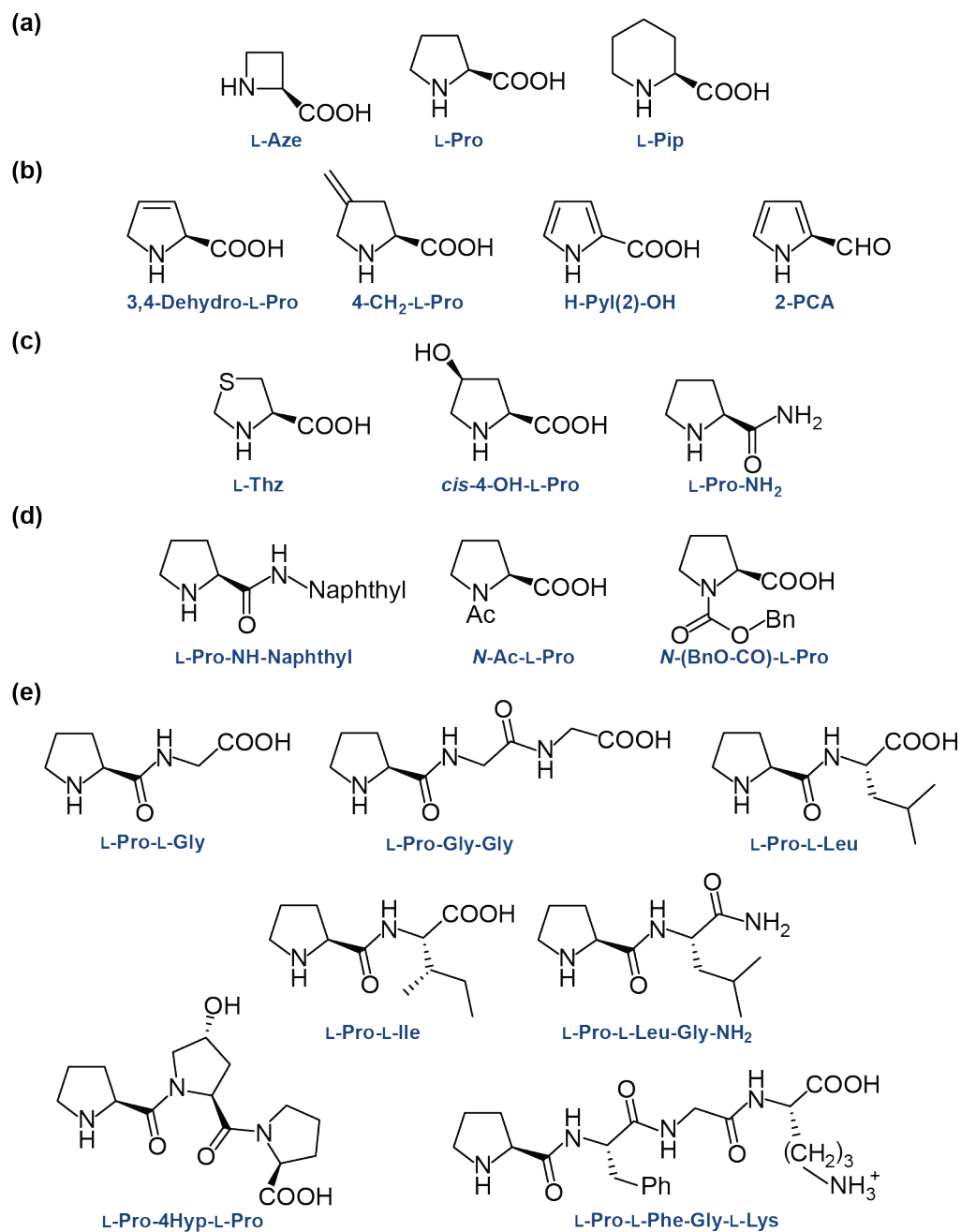


Figure 3.13: L-Proline analogues tested for stimulation of *S. griseoviridis* P8648 *trans*P4H 2-OG turnover: a radioactive [U - ^{14}C]-2-OG-turnover assay was used as a measure substrate quality.

Substrate	2-OG Turnover/%	*Relative Activity/%
None	3	5
L-Aze	9	17
L-Pro	51	100
L-Pip	51	100
3,4-Dehydro-L-Pro	67	134
4-CH ₂ -L-Pro	15	29
H-Pyl(2)-OH	5	10
2-PCA	4	7
L-Thz	10	20
cis-4-OH-L-Pro	1.5	3

Table 3.1: *S. griseoviridis* P8648 *trans*P4H-induced carboxylative turnover of 2-OG in the presence of L-proline.^{191,256}

*Relative activity values were referenced to L-proline turnover ('100% relative activity').

The reactions were followed by monitoring the production of [¹⁴C]-CO₂ from [*U*-¹⁴C]-2-OG by scintillation counting. Data was taken from Lawrence *et al.*^{191,256}

The 2-OG-turnover results measured by Lawrence *et al.* for ring-size-modified L-proline analogues (Figure 3.13 a) showed some increases in 2-OG turnover compared with substrate-uncoupled experiments (Table 3.1), suggesting these compounds may be substrates of the *trans*P4H.^{191,256} L-azetidine-2-carboxylic acid (L-Aze), a 4-membered-ring analogue of L-Pro, was shown to stimulate *trans*P4H 2-OG turnover with a relative rate of 17% (compared with 2-OG consumption in the presence of L-Pro, for which relative activity is defined as 100%), whereas uncoupled 2-OG turnover accounted for only ~5% of activity (Table 3.1);^{191,256} the percentage of *trans*P4H 2-OG turnover was also higher in the presence of L-Aze (9%) than substrate-uncoupled levels (3%) (Table 3.1).^{191,256} L-Pipecolic acid (L-Pip), a 6-membered-ring analogue of L-Pip, showed significant enhancement of *trans*P4H 2-OG turnover (relative activity = 100%) compared with substrate-uncoupled levels (5%), and matched levels seen with L-Pro reactions (Table 3.1);^{191,256} percentage 2-OG turnover (51%) was also significant in the presence of L-Pip, also matching

values measured with L-Pro present (Table 3.1).^{191,256} These findings strongly suggested that L-Pip may be an alternative substrate of *trans*P4H.

Interestingly, Lawrence *et al.* observed stimulation of *trans*P4H 2-OG turnover by 3,4-dehydro-L-proline (Figure 3.13 b) which exceeded values measured for L-Pro-containing reactions (Table 3.1);^{191,256} 3,4-Dehydro-L-Pro stimulated a relative activity of 134% (and 2-OG turnover = 67%) compared with the relative activity of 100% (and 2-OG turnover = 51%) with L-Pro^{191,256} (A similar ‘enhanced’ 2-OG turnover reaction has been reported for hyoscyamine 6 β -hydroxylase).³⁷¹ HPLC analysis of PITC-derivatised products from 3,4-Dehydro-L-Pro *trans*P4H reactions showed a peak with a retention time identical to that of synthetic, PITC-derivatised *trans*-3,4-epoxy-L-Pro, suggesting *trans*P4H catalyses the *trans*-epoxidation of 3,4-Dehydro-L-Pro to give *trans*-3,4-epoxy-L-Pro;^{191,256} no evidence was accrued that suggested that *cis*-3,4-epoxy-L-Pro is a product.^{191,256} In *trans*P4H reactions containing 4-methylene-L-proline (4-CH₂-L-Pro), moderate stimulation of 2-OG turnover above substrate-uncoupled levels was observed (relative activity = 29%, 2-OG turnover = 15%) (Table 3.1),^{191,256} suggesting that 4-CH₂-L-Pro is a *trans*P4H substrate.

With respect to L-Pro analogues with heteroatom substitutions (Figure 3.13 c), L-thiazolidine-4-carboxylic acid (L-Thz) stimulated slight *trans*P4H 2-OG turnover above substrate-uncoupled levels (relative activity = 20%, 2-OG turnover = 10%) (Table 3.1),^{191,256} and so L-Thz could potentially be a *trans*P4H substrate; on the other hand, L-prolinamide (L-Pro-NH₂) and *cis*-4-hydroxy-L-proline (*cis*-4-OH-L-Pro) did not show activity.^{191,256}

‘Functional-group-protected’ forms of L-Pro (Figure 3.13 d), such as *N*-acetyl-L-proline (*N*-Ac-L-Pro), L-proline naphthyl-2-amide (L-Pro-NH-Naphthyl), and *N*-benzyloxycarbonyl-L-proline [*N*-(BnO-CO)-L-Pro], were shown to be inactive;^{191,256} similarly, seven peptides containing L-Pro (Figure 3.13 e) were shown to be inactive at stimulating 2-OG turnover *trans*P4H.

3.4.2 Substrate Screening by HPLC

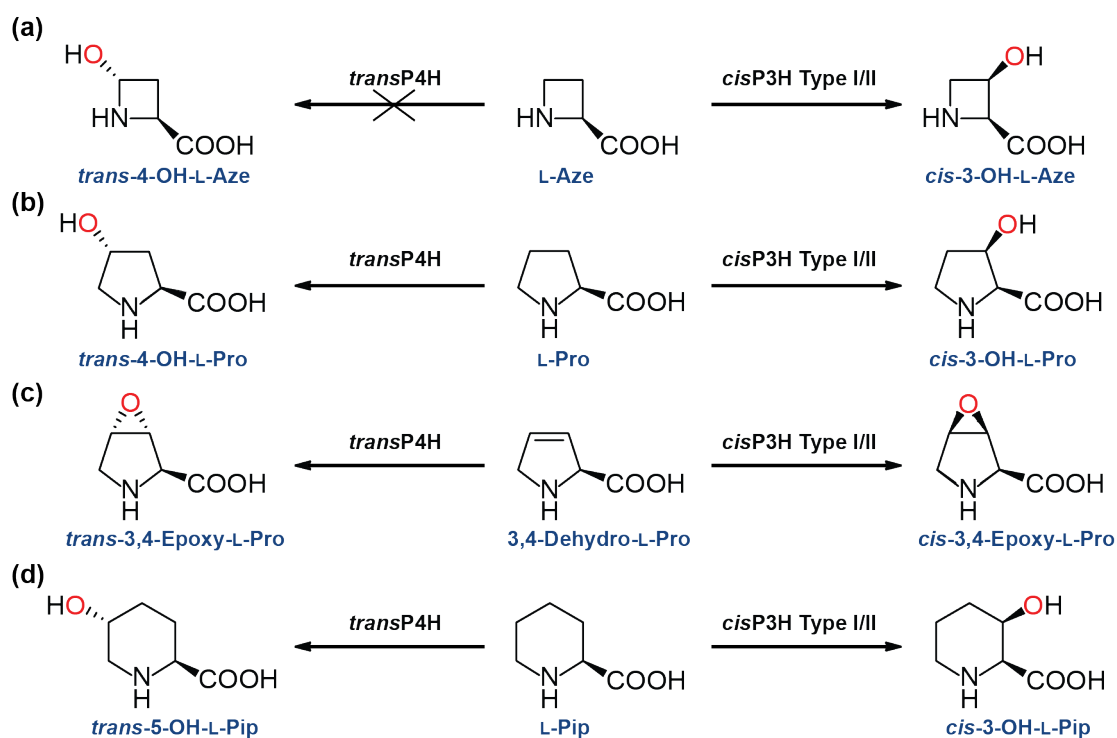
To investigate how effectively L-proline analogues are catalytically hydroxylated by proline hydroxylase enzymes, Shibasaki *et al.* tested various proline-like substrates (Scheme 3.9 and Table 3.2) for proline hydroxylase turnover using an HPLC assay, and compared the substrate turnover values with those measured for L-proline (the ‘optimal’ and ‘natural’ substrate for these enzymes);^{195,201} the specific activities of the purified enzymes [*trans*P4H (from *Dactylosporangium sp.* RH1), *cis*P3H Type I and Type II (from *Streptomyces sp.* strain TH1)] were 9,674 U mg⁻¹,¹⁹⁵ 7,150 U mg⁻¹,^{195,201} and 12,700 U mg⁻¹ (Table 3.2 b),^{195,201} respectively (where 1 U mg⁻¹ $\equiv$ 1 nmol min⁻¹ mg⁻¹).

Substrate	<i>trans</i> P4H			
	Products	K _{m, App} /mM	Activity/U mg ⁻¹	*Relative Activity/%
L-Aze	-	-	0	0
L-Pro	<i>trans</i> -4-OH-L-Pro	0.2	9,674	100
3,4-Dehydro-L-Pro	<i>trans</i> -3,4-epoxy-L-Pro	5.1	6,385	66
L-Pip	<i>trans</i> -5-OH-L-Pip	5.7	10,545	109

Substrate	<i>cis</i> P3H Type I/Type II			
	Products	K _{m, App} /mM	Activity/U mg ⁻¹	*Relative Activity/%
L-Aze	<i>cis</i> -3-OH-L-Aze	NT/2.1	214/5080	3/40
L-Pro	<i>cis</i> -3-OH-L-Pro	0.2/0.4	7,150/12,700	100/100
3,4-Dehydro-L-Pro	<i>cis</i> -3,4-epoxy-L-Pro	NT/8.4	2,145/7,239	30/57
L-Pip	<i>cis</i> -3-OH-L-Pip	3.8/>100	1,072/381	15/3

Table 3.2: Proline hydroxylase activity data for proline analogues: HPLC assays were used to measure substrate turnover for compounds such as L-azetidine-2-carboxylic acid (L-Aze), L-proline (L-Pro), 3,4-dehydro-L-proline (3,4-Dehydro-L-Pro), and L-pipecolic acid (L-Pip). 1 nmol min⁻¹ mg⁻¹ $\equiv$ 1 U mg⁻¹

*Relative activity values were referenced to L-proline turnover rates.



Scheme 3.9: Proline Hydroxylase (*trans*P4H, *cis*P3H Type I and II) proline-analogue substrate activity assays: (a) L-azetidine-2-carboxylic acid (L-Aze), (b) L-proline (L-Pro), (c) 3,4-dehydro-L-proline (3,4-Dehydro-L-Pro), and (d) L-pipecolic acid (L-Pip) were tested for proline hydroxylase turnover using a HPLC-based assay.^{195,201}

Although the reactivity of both *Streptomyces* sp. (strain TH1) *cis*P3H enzymes to proline analogues were reported to be different, they share the same regio- and stereoselectivity for hydroxylation.^{195,201} For example, whilst both *cis*P3H enzymes were shown to catalyse the hydroxylation of a 4-membered-ring analogue of L-proline, L-azetidine-2-carboxylic acid (L-Aze), to produce *cis*-3-hydroxy-L-azetidine-2-carboxylic acid (*cis*-3-OH-L-Aze) (Scheme 3.9 a),^{195,201} the *cis*P3H Type II enzyme has a ~24-fold higher activity for L-Aze hydroxylation than the *cis*P3H Type I isozyme (Table 3.2);^{195,201} L-Aze hydroxylation rates were 214 U mg⁻¹ and 5,080 U mg⁻¹ for *cis*P3H Type I and Type II (Table 3.2), respectively.^{195,201} The stereochemistry of the L-Aze-hydroxylation product (from *cis*P3H reactions) was largely determined by NMR analyses;¹⁹⁵ the product was *cis*-3-hydroxy-L-azetidine-2-carboxylic acid (*cis*-3-OH-L-Aze) (Scheme 3.9 a).¹⁹⁵

Both *cis*P3H Type I and Type II were shown to oxidise 3,4-dehydro-L-proline to produce 3,4-epoxy-L-proline by NMR spectral analyses (Scheme 3.9 c and Table 3.2).^{195,201} The turnover of 3,4-dehydro-L-proline by *Dactylosporangium* sp. RH1 *trans*P4H (6,385 U mg⁻¹)¹⁹⁵ is similar that of *cis*P3H Type II (7,239 U mg⁻¹), ~1.1-fold larger²⁰¹ (Table 3.2); HPLC and TLC analyses showed that *Dactylosporangium* sp. RH1 *trans*P4H converts 3,4-dehydro-L-proline into *trans*-3,4-epoxy-L-proline (Scheme 3.9 c),¹⁹⁵ the same reaction catalysed by *S. griseoviridis* P8648 *trans*P4H.²⁵⁶

Interestingly, although *cis*P3H Type II is generally more active than *cis*P3H Type I towards many substrates (*e.g.* L-proline, L-azetidine-2-carboxylic acid, 3,4-hydro-L-proline), *cis*P3H Type I displays a 2.8-fold higher activity towards L-pipecolic acid (a 6-membered-ring analogue of L-proline) than *cis*P3H Type II (Scheme 3.9 d and Table 3.2).²⁰¹ Shibasaki *et al.* reported that *Dactylosporangium* sp. RH1 *trans*P4H converts L-pipecolic acid to *trans*-5-hydroxy-L-pipecolic acid (*trans*-5-OH-L-Pip) (Scheme 3.9 d), based upon NMR spectral analyses and specific rotation data;¹⁹⁵ L-pipecolic acid turnover (10,545 U mg⁻¹) was reported to be slightly higher (1.09-fold) than that measured for *trans*P4H reactions with L-proline (Table 3.2).¹⁹⁵

3.4.3 Robust Methods for L-Proline-Derivative Hydroxylation

Amino acids are valuable building blocks for biological molecules (such as, peptides, proteins, and many other natural products)^{101–106} and organic synthesis (chiral starting materials, auxiliaries, and catalysts).¹⁰⁰ There is growing interest in developing methods of producing these compounds, including from proline.¹⁹⁸ Proline derivatives are useful for extension of the chiral pool. Hydroxyprolines are used in the synthesis of pharmaceutical drugs such as carbapenem antibiotics, angiotensin-converting enzyme (ACE) inhibitors, and antispastic agents.^{204,372} Proline's rigid

conformation gives it particularly strong chirality inducing properties, and proline is therefore widely used in stereoselective synthesis.^{107,108}

Relatively recently, Klein *et al.* set out to develop methods (with scalability) for producing hydroxyprolines and hydroxylated proline analogues with recombinant L-proline hydroxylase enzymes;¹⁹⁸ the group subcloned the gene-optimised sequence coding for *cis*P3H Type I (from *Streptomyces sp.* strain TH1) (Protein Data Bank: 1E5RA), *cis*P4H (from *Sinorhizobium meliloti*) (GenBank No. CAC47686), and *trans*P4H (from *Dactylosporangium sp.*) (GenBank No. BAA20094.1) into a high-expression, *E. coli* expression vector containing an *N*-terminal His₆-tag sequence for Ni-affinity protein purification (Takara pCOLD I).¹⁹⁸ Using these recombinantly expressed proline hydroxylase proteins (*cis*P3H type I, *cis*P4H, *trans*P4H), Klein *et al.* explored methods for the PH-catalysed hydroxylation of L-proline and L-proline analogues;^{198,295} the group explored the use of both *in vitro* enzyme reactions and biotechnology-based *in vivo* methods involving cell culture fermentation.^{198,295}

Microbiological (*In vivo*) production of hydroxyprolines from L-proline by proline hydroxylase enzymes involved ‘feeding’ L-proline (6.25 mM) to an exponentially growing culture of *E. coli* strains containing a particular PH enzyme (*cis*P3H Type I, *cis*P4H, or *trans*P4H) and incubation at 28 °C.^{198,295} HPLC analyses of the hydroxyproline-product mixtures, isolated from the PH reactions with two ion-exchange steps, provided quantification of the substrate conversions and final product concentrations (Table 3.3);^{198,295} under the growth condition used, the peak-product concentrations (and substrate turnovers) were found to be >3.75 mM (>60%) for *cis*P3H Type I and *trans*P4H reactions, and 2.56 mM (41%) for the *cis*P4H reactions.¹⁹⁸

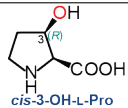
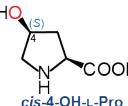
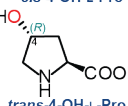
PH Enzyme	Product	Product Conversion Yield ^a	Final Isolation Yield ^b
<i>cis</i> P3H Type I	 <i>cis</i> -3-OH-L-Pro	69%	61% (305 mg)
<i>cis</i> P4H	 <i>cis</i> -4-OH-L-Pro	41%	35% (175 mg)
<i>trans</i> P4H	 <i>trans</i> -4-OH-L-Pro	63%	63% (255 mg)

Table 3.3: Hydroxylation of L-proline in *E. coli* cultures expressing L-proline *cis*-3-, *cis*-4-, and *trans*-4-hydroxylases (*cis*P3H, *cis*P4H, and *trans*P4H).¹⁹⁸

After 72 h of cell culture incubation, the hydroxyprolines were isolated from the culture broth by two steps of ion-exchange chromatography; by this method, 80-90% of the corresponding products were obtained.

^a 500 mg L-proline were converted in 800 mL culture for 3 days at 28 °C; the percentage conversion was determined by HPLC.

^b Since the product still contained inorganic salts after ion exchange chromatography, yields were determined by HPLC.

To demonstrate that their system was also applicable to non-natural PH substrates, Klein *et al.* carried out PH reactions (*in vitro*) with a 6-membered homologue of L-proline, L-pipecolic acid (L-Pip) (Table 3.4).¹⁹⁸ Hydroxylation of L-Pip by *trans*P4H was shown to produce *trans*-5-hydroxy-L-pipecolic acid (*trans*-5-OH-L-Pip) with a good yield (68%) (Table 3.4).¹⁹⁸ Contrastingly, hydroxylation of L-Pip by *cis*P3H Type I was shown to be poor (Table 3.4), with only 17% of the L-Pip being converted into *cis*-3-hydroxy-L-pipecolic acid (*cis*-3-OH-L-Pip).¹⁹⁸ With *cis*P4H reactions, the conversion of L-Pip was ~55% as effective as the reactions observed for L-Pro under identical conditions;¹⁹⁸ in addition, two products, namely *cis*-3-hydroxy-L-pipecolic acid (*cis*-3-OH-L-Pip) and *cis*-5-hydroxy-L-pipecolic acid (*cis*-5-OH-L-Pip), were obtained in an approximate 1:1 ratio (33% and 33%) (Table 3.4). HPLC and NMR analyses confirmed the proposed structures of these *cis*-3- and *cis*-5-hydroxylated products.¹⁹⁸

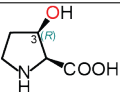
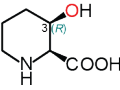
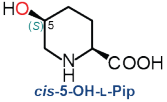
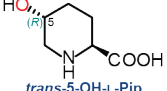
PH Enzyme	Product	Product Conversion Yield
<i>cis</i> P3H Type I	 <i>cis</i> -3-OH-L-Pro	17%
<i>cis</i> P4H	 <i>cis</i> -3-OH-L-Pip	33%
	 <i>cis</i> -5-OH-L-Pip	33%
<i>trans</i> P4H	 <i>trans</i> -5-OH-L-Pip	68%

Table 3.4: Hydroxylation of L-pipecolic acid (L-Pip) by FPLC-purified L-proline *cis*-3-, *cis*-4-, and *trans*-4-hydroxylases (*cis*P3H, *cis*P4H, and *trans*P4H).¹⁹⁸

8 mg (1 mM) pipecolic acid samples were incubated with fresh enzyme preparations for 14 h at 37 °C (*cis*P3H Type I, *trans*P4H) and 25 °C (*cis*P4H), respectively. No other products were detected in this assay (by HPLC or NMR analyses).

The hydroxypipecolic acids were produced on a small scale (few milligrams) using *in vitro* PH reactions, Klein *et al.* proposed that these reactions would have little potential for large-scale hydroxypipecolic acid production, and microbiological production (*in vivo*) methods were pursued;¹⁹⁸ a 4 mM solution of L-Pip was converted with a *trans*P4H-expressing *E. coli* strain under the same conditions used for their *in vivo* proline hydroxylation reactions.¹⁹⁸ Despite initial concerns that there may be issues with limiting L-Pip uptake by cells (since L-Pip is not proteinogenic amino acid¹⁰¹), a high product concentration (3.12 mM) was detected in the fermentation by HPLC analysis;¹⁹⁸ the *trans*-5-OH-L-Pip product was isolated with a 61% yield (Table 3.5).¹⁹⁸

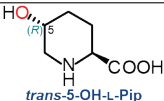
PH Enzyme	Product	Product Conversion ^a	Final Isolated Yield ^b
<i>trans</i> P4H	 <chem>OC[C@H]1CC[C@@H](C(=O)O)N1</chem> <i>trans</i> -5-OH-L-Pip	78%	61% (35 mg)

Table 3.5: Hydroxylation of L-proline in *E. coli* cultures expressing L-proline *trans*-4-hydroxylase (*trans*P4H).¹⁹⁸

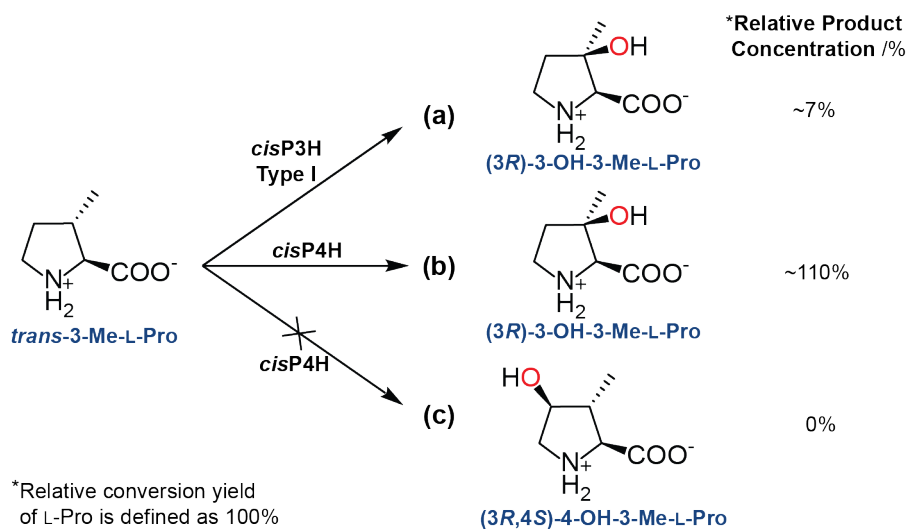
After 72 h of cell culture incubation, the hydroxyprolines were isolated from the culture broth by two steps of ion-exchange chromatography; by this method, 80-90% of the corresponding products were obtained.

^a 57 mg (4 mM) L-proline was converted in 800 mL culture for 3 days at 28 °C; the percentage conversion (78% - 3.12 mM) was determined by HPLC.

^b Since the product still contained inorganic salts after ion exchange chromatography, yields were determined by HPLC. Since endogenously produced L-proline was also hydroxylated by *E. coli*, *trans*-4-hydroxyproline was obtained in the same fraction as the product after chromatographic work-up (seen by HPLC and ¹H-NMR analyses).

Klein *et al.* also studied the hydroxylation of *trans*-3-methyl-L-proline (*trans*-3-Me-L-Pro) by *cis*P3H Type II (from *Streptomyces sp.* strain TH1) and *cis*P4H (from *Sinorhizobium meliloti*).²⁹⁵ A new compound was detected in low yield (Scheme 3.10 a).²⁹⁵ Based upon *cis*-3-hydroxylation selectivity of *cis*P3H Type II, the product was initially assumed to give the tertiary alcohol, (3*R*,4*S*)-4-hydroxy-3-methyl-L-proline [(3*R*,4*S*)-4-OH-3-Me-L-Pro]; surprisingly, the *cis*P4H reactions also gave a product with the same retention time as the *cis*P3H Type II reaction product,²⁹⁵ suggesting that both enzymes hydroxylated the C3-position.²⁹⁵ The hydroxylation yield was much higher for the *cis*P4H reaction than the *cis*P3H Type II reaction, with a relative yield of ~110% compared to the L-Pro control reactions (Scheme 3.10 b).²⁹⁵ 1D- and 2D-NOESY NMR-spectra measured for the products of these *trans*-3-Me-L-Pro PH reactions showed a correlation between the 3-methyl group and the proton at C-2, suggesting these two substituents were arranged in positions ‘*cis*’ to each other in the reaction product;²⁹⁵ this evidence is consistent with the PH-catalysed *cis*-3-hydroxylation of *trans*-3-Me-L-Pro (with a retention of stereochemistry), and the production (3*R*,4*S*)-4-hydroxy-3-methyl-L-proline.²⁹⁵ In addition, no

cis-4-hydroxylation product, (3*R*,4*S*)-4-hydroxy-3-methyl-L-proline [(3*R*,4*S*)-4-OH-3-Me-L-Pro], was observed in *cis*P4H reactions with *trans*-3-Me-L-Pro (Scheme 3.10 c).²⁹⁵



Scheme 3.10: Hydroxylation of *trans*-3-methyl-L-proline (*trans*-3-Me-L-Pro) by: (a) *cis*P3H Type II and (b) *cis*P4H.²⁹⁵

(a) and (b) Both *cis*P3H Type II and *cis*P4H were shown to catalyse the *cis*-3-hydroxylation of *trans*-3-Me-L-Pro (with a retention of stereochemistry) to produce (3*R*,4*S*)-4-hydroxy-3-methyl-L-proline [(3*R*,4*S*)-4-OH-3-Me-L-Pro]; relative yields were ~7% and ~110%, respectively, when compared with conversion yields of L-Pro control reactions ('100% yield').

(c) No *cis*-4-hydroxylation product, (3*R*,4*S*)-4-hydroxy-3-methyl-L-proline [(3*R*,4*S*)-4-OH-3-Me-L-Pro], was observed in *cis*P4H reactions with *trans*-3-Me-L-Pro.

3.4.4 Proline Analogue Substrate Screening

Although there had been some work carried out to determine alternative proline hydroxylase substrates,^{191,195,201,256} at the start of this project, there were no extensive studies on proline hydroxylase alternative substrates. Thus, the full extent of the catalytic potential of the proline hydroxylase enzymes for biocatalysis remained unclear. One aim of my work was to determine which classes of compounds the PH enzymes could be accepted as substrates, in order to reveal which chemical motifs are necessary for enzymatic turnover by proline hydroxylases; an effort was also

made, where possible, to characterise the products formed by these reactions, and improve our mechanistic understanding of the proline hydroxylases.

Both LC/MS and NMR methods were used to analyse substrate turnover by 3 proline hydroxylase enzymes with different hydroxylation stereoselectivities, namely *cis*P3H Type I (from *Streptomyces* sp. strain TH1) (Protein Data Bank: 1E5RA), *Sinorhizobium meliloti* (GenBank No. CAC47686), and *trans*P4H (from *Dactylosporangium* sp.) (GenBank No. BAA20094.1). *In vitro* reactions using each of the 3 purified PH enzymes were incubated in parallel for a set period time (typically 5-14 h) under the same conditions; reactions being analysed by LC/MS typically contained: 1 mM substrate, 10 mM 2-OG, 0.1-0.5 mM (NH₄)₂Fe(SO₄)₂, 1.0-1.5 mM sodium ascorbate in 50 mM MES-NaOH, pH 6.5 at 25 °C. If products were identified by LC/MS analysis or NMR-analysis of complete reaction mixtures, reactions were scaled up 10-20-fold, and products were purified by LC/MS methods to produce material for extensive NMR characterisation experiments (*e.g.* ¹H-spectra, ¹H-¹H-COSY, 2D-NOESY, ge-1D TOCSY, *etc.*); provided a sufficient quantity of a product sample was isolated (with sufficient purity), NMR characterisation experiments often provided unambiguous spectral assignments and allowed product stereochemistries to be determined.

3.4.4.1 Effects of Amino Acid α -Carbon-Stereochemistry on PH Catalysis

Amino acids are chiral compounds (with a few exceptions, *e.g.* glycine) with a stereocentre at carbon-2, the position bonded to both an amino group and a carboxylic acid group;¹⁰¹ most naturally occurring amino acids have '(2S)'-stereochemistry¹⁰¹ according to the Cahn-Ingold-Prelog rules for stereochemical naming conventions,^{373,374} with the exception of cysteine [which is '(2R)']. Biochemists often use L- and D-

nomenclature when discussing natural products, such as amino acids; all naturally occurring proteinogenic amino acids have L-stereochemistry (Figure 3.14 b), and are structurally related to the model compound, L-glyceraldehyde (Figure 3.14 a),¹⁰¹ a compound which was used to define the D/L-naming convention.¹⁰¹ Owing to the rigid, aliphatic ring of proline, the spatial differences between the D- and L-forms of proline are quite marked (*i.e.* the regions of space occupied by each molecule can be quite different) (Figure 3.14 c), and so a change in D/L-stereochemistry can, in some cases, have a large impact on the properties of a substance containing a proline-like motif.¹⁰¹

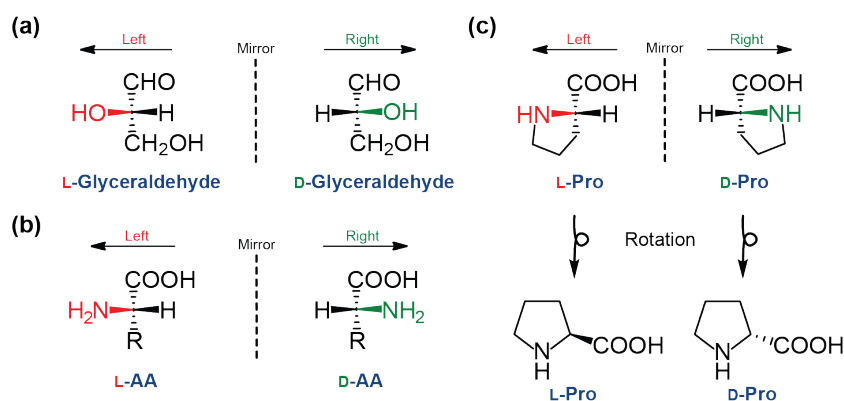


Figure 3.14: L- and D-stereochemistry nomenclature of biological molecules:

- (a) glyceraldehyde is the model compound used to define the D/L-naming conventions;
- (b) amino acid (AA) α -carbon stereochemistry can also be defined using D/L-labels, with D- and /L-AAs resembling D- and /L-glyceraldehyde;
- (c) the spatial arrangement of atoms in L- and D-Pro are particularly marked, as a result of the amino acid's ring 5-membered ring.

To test the effect of D/L-stereochemistry [*i.e.* '(2S)' versus '(2R)'] on proline hydroxylase activities, trial reactions were carried with *cis*P3H Type I, *cis*P4H, and *trans*P4H using D-form substrates; initially, D-proline (D-Pro), D-pipecolic acid (D-Pip - a 6-membered-ring D-proline homologue), and DL-Pip (Figure 3.15) were incubated with the PH enzymes under conditions used previously for L-Pro turnover. LC/MS-analysis of the completed reactions showed that no signs of D-Pro hydroxylation by *cis*P3H Type I, *cis*P4H, or *trans*P4H; the extracted-ion chromatograms (XICs)

showing 132.1 m/z ions (D-Hyp $[M+H]^+$), 154.1 m/z ions (*p*ASA $[M+H]^+$ - internal standard), and 116.1 m/z ions (D-Pro $[M+H]^+$) only contained peaks corresponding the starting materials (D-Pro) and the internal standard (*p*ASA) (Figures 3.16 a-c). In all three PH reactions with *cis*P3H Type I, *cis*P4H, and *trans*P4H, no *cis*-3-OH-D-Pro, *cis*-4-OH-D-Pro, or *trans*-4-OH-D-Pro were detected (Figures 3.16 a-c), respectively, suggesting that D-Pro is not a substrate; this is consistent with previous data collected by Shibasaki *et al.*,^{195,201} which suggested that D-Pro is not a substrate of *trans*P4H, *cis*P3H Type I, or *cis*P3H Type II.^{195,201} Similarly, PH reactions with a D-Pip substrate showed no signs of D-hydroxypipecolic acid (D-Hypip) products (namely, *cis*-3-OH-D-Pip, *cis*-4-OH-D-Pip, and *trans*-4-OH-D-Pip, respectively); the LC/MS chromatograms for the D-Pip PH reactions only contain peaks for D-Pip starting material and the internal standard (*p*ASA), and lacked peaks for any D-Hypip products. Consistent with this observation, PH reactions (*cis*P3H Type I, *cis*P4H, *trans*P4H) using a DL-Pip substrate (*i.e.* a racemic, 1:1 mixture of L- and D-Pip) showed reduced Pip turnover (~2-fold less) compared with PH reactions using a L-Pip substrate, implying that only the L-Pip portions of the substrate were hydroxylated; LC/MS analysis showed that the relative integrals of the Hypip peaks from the DL-Pip PH reaction chromatograms were roughly half of the corresponding relative integrals measured for Hypip peaks from L-Pip PH reactions. These observations are consistent with these proline hydroxylases only catalysing the hydroxylation of the L-Pip half of the racemic DL-Pip mixture, and not catalysing D-Pip turnover.

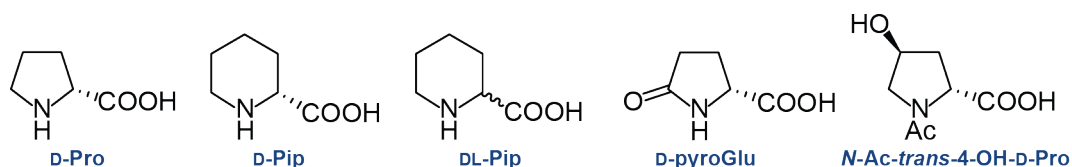


Figure 3.15: Proline analogues with D-stereochemistry tested for proline hydroxylase substrate activity: D-proline (D-Pro), D-pipecolic acid (D-Pip), D-pyrroglutamate (D-pyroGlu), *N*-acetyl-*trans*-4-hydroxy-D-proline (*N*-Ac-*trans*-4-OH-D-Pro) were not substrates. In the case of DL-Pip, only the L-form appeared to be converted by the PH enzymes.

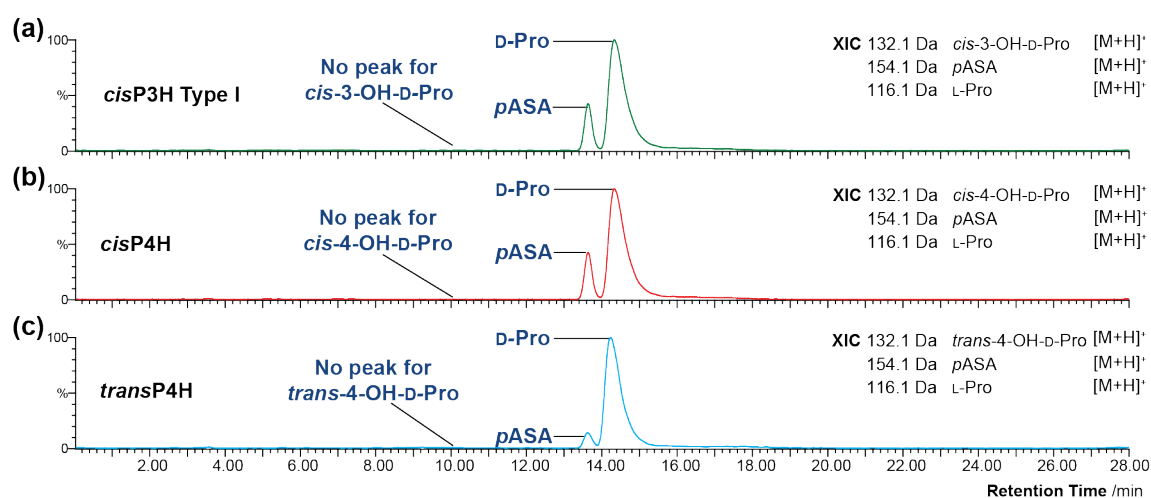


Figure 3.16: LC/MS-based analysis of proline hydroxylase reactions with D-Pro [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H].

Extracted-ion chromatograms (XICs) showing 132.1 m/z ions (D-Hyp [M+H]⁺), 154.1 m/z ions (pASA [M+H]⁺ internal standard), and 116.1 m/z ions (D-Pro [M+H]⁺) only contained peaks corresponding the starting materials (D-Pro) and the internal standard (pASA) (*i.e.* no *cis*-3-OH-D-Pro, *cis*-4-OH-D-Pro, or *trans*-4-OH-D-Pro products were observed, respectively).

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM D-Pro. 100 μ L reactions were carried out in at room temperature (25 $^{\circ}$ C) for 2 h in 50 mM MES-NaOH, pH 6.5 buffer.

Assays (with *cis*P3H Type I, *cis*P4H, and *trans*P4H) were also carried out with D-based proline analogue compounds, D-pyroglutamate (D-pyroGlu) and *N*-acetyl-*trans*-4-hydroxy-D-proline (*N*-Ac-*trans*-4-D-Pro) (Figure 3.15). As with D-Pro and D-Pip, these D-stereochemistry proline derivatives did not show any activity for proline hydroxylase turnover by LC/MS.

Based upon the results gathered in this project, and observations made by Shibasaki *et al.* (with *cis*P3Hs and *trans*P4H),^{195,201} it appears that there is a strict requirement for L-stereochemistry for PH substrates (based on assays with *cis*P3H Type I, *cis*P4H, and *trans*P4H); in most of my work, the D-enantiomer of the proline substrate-analogues were therefore excluded from activity screening tests.

3.4.4.1.1 Substrate Turnover Summary: D/L-Stereochemistry

The following table (Table 3.6) summarises the proline hydroxylase activities with the D/L-substrates screened for proline hydroxylase substrate activity in this section. See below for details of the reaction yields and products.

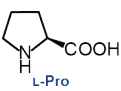
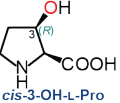
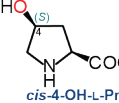
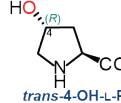
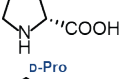
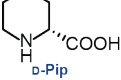
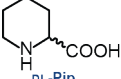
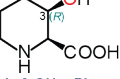
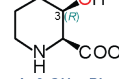
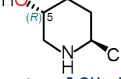
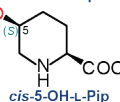
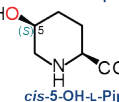
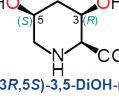
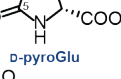
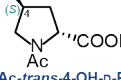
Substrate	Length/h	Product(s) [+ Yield(s)]					
		cisP3H Type I		cisP4H		transP4H	
 L-Pro	5	 <i>cis</i> -3-OH-L-Pro	100%	 <i>cis</i> -4-OH-L-Pro	100%	 <i>trans</i> -4-OH-L-Pro	100%
 D-Pro	14	-	0%	-	0%	-	0%
 D-Pip	14	-	0%	-	0%	-	0%
 DL-Pip	14	 <i>cis</i> -3-OH-L-Pip	50%	 <i>cis</i> -3-OH-L-Pip	22%	 <i>trans</i> -5-OH-L-Pip	50%
		 <i>cis</i> -5-OH-L-Pip		 <i>cis</i> -5-OH-L-Pip	20%		
		-		 (3 <i>R</i> ,5 <i>S</i>)-3,5-DiOH-L-Pip	5%		
 D-pyroGlu	14	-	0%	-	0%	-	0%
 <i>N</i> -Ac- <i>trans</i> -4-OH-D-Pro	14	-	0%	-	0%	-	0%

Table 3.6: Proline hydroxylase substrate analogue screening: the effects of substrate D/L-stereochemistry on proline hydroxylase activity.

D-Proline (D-Pro), D-pipecolic acid (D-Pip), DL-pipecolic acid (DL-Pip), D-pyrroglutamic acid (D-pyroGlu), and *N*-acetyl-*trans*-4-hydroxy-D-proline (*N*-Ac-*trans*-4-OH-D-Pro) were screening for proline hydroxylase substrate activity and compared to the natural substrate, L-Pro.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM Substrate.

3.4.4.2 Effects of Amino- & Carboxylic-Functionalisation on PH Catalysis

To investigate the effects of functionalising the amino and carboxylic acid groups on proline hydroxylase substrate acceptance, a few different amino-modified/carboxylic acid-modified substrate analogues (Figure 3.17) were tested for turnover by the three proline hydroxylase enzymes (*cis*P3H Type I, *cis*P4H, and *trans*P4H).

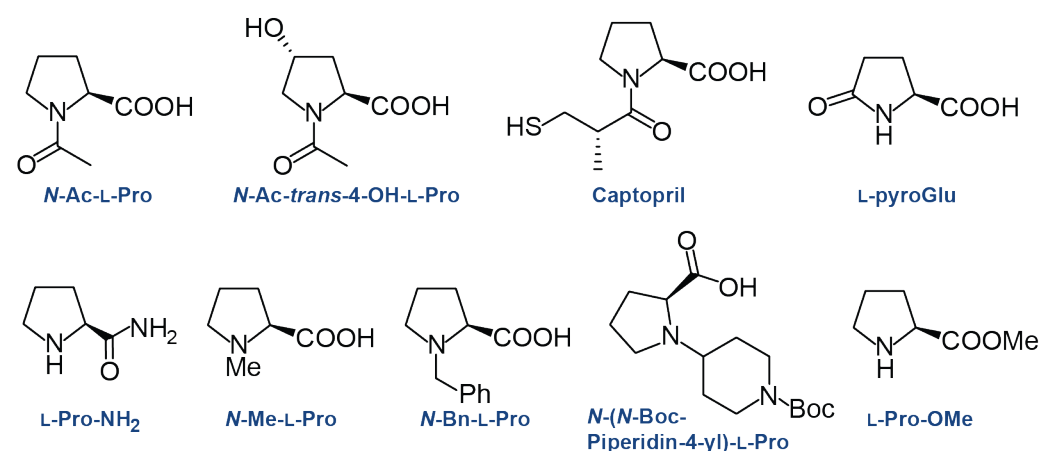


Figure 3.17: Amino- and carboxylic-acid-functionalised L-proline test substrate: the substrates tested for substrate activity by LC/MS and/or NMR assay.

Substrates included:

N-carbonylised compounds - *N*-acetyl-L-proline (*N*-Ac-L-Pro), *N*-acetyl-*trans*-4-hydroxy-L-proline (*N*-Ac-*trans*-4-OH-L-Pro), and captopril;

Other amides - L-pyrroglutamic acid (L-pyroGlu) and L-prolineamide (L-Pro-NH₂);

N-alkylated substances - *N*-methyl-L-proline (*N*-Me-L-Pro), *N*-benzyl-L-proline (*N*-Bn-L-Pro), and *N*-[(*N*-*tert*-butoxycarbonyl)piperidin-4-yl]-L-proline [*N*-(*N*-Boc-Piperidin-4-yl)-L-Pro];

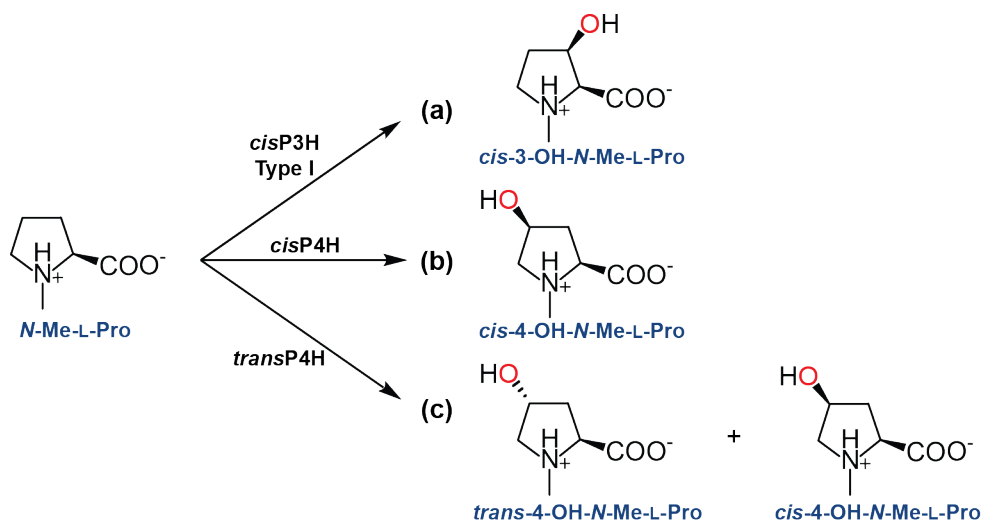
A carboxylic ester - L-proline methyl ester (L-Pro-OMe)

Three proline derivatives with *N*-carbonyl groups (*i.e.* amides), namely *N*-acetyl-L-proline (*N*-Ac-L-Pro), *N*-acetyl-*trans*-4-hydroxy-L-proline (*N*-Ac-*trans*-4-OH-L-Pro), and captopril (Figure 3.17), were tested for proline hydroxylase activity by LC/MS and NMR assays; captopril was a compound of particular interest, because it is an angiotensin-converting enzyme (ACE) inhibitor used for the treatment of hyper-

tension and some types of congestive heart failure.^{375–377} Modified (e.g. hydroxylated) versions of captopril may have beneficial pharmaceutical properties, and therefore may be a useful biosynthesis target. The proline hydroxylase (*cis*P3H Type I, *cis*P4H, and *trans*P4H) reactions with *N*-Ac-L-Pro, *N*-Ac-*trans*-4-OH-L-Pro, and captopril showed that all of these compounds are not substrates, with no products being detected, only starting materials and internal standards.

Two other compounds (Figure 3.17) with an amide functionality were also tested as substrates for the three proline hydroxylase enzymes; L-pyroglutamic acid (L-pyroGlu), a compound with C5-carbonyl group functionality, and L-prolineamide (L-Pro-NH₂), a compounds with C-terminal amide functionality, were also found to be inactive as proline hydroxylase substrates. These results suggest that an amide-functionality is incompatible with proline hydroxylase activity; it is likely that an amino nitrogen and/or carboxyl lone pairs are required for substrate recognition by the PH enzymes.

Of the *N*-alkylated L-proline derivatives, *N*-benzyl-L-proline (*N*-Bn-L-Pro) and *N*-[*N*-(*tert*-butoxycarbonyl)piperidin-4-yl]-L-proline [*N*-(*N*-Boc-Piperidin-4-yl)-L-Pro] (Figure 3.17) were not accepted as proline hydroxylase substrates for *cis*P3H Type I, *cis*P4H, or *trans*P4H, suggesting that the *N*-terminal functionalities make these compounds too ‘bulky’ for the PH enzymes to bind them and efficiently catalyse their hydroxylation; however, *N*-methyl-L-proline (*N*-Me-L-Pro) was accepted as a substrate by all three tested PH enzymes (Scheme 3.11), implying that the PH enzymes can accommodate substrates with short alkyl chains on the ring nitrogen.



Scheme 3.11: Proposed hydroxylation reactions catalysed by the proline hydroxylases using an *N*-methyl-L-proline (*N*-Me-L-Pro) substrate [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H]:

(a) *cis*P3H Type I produces *cis*-3-hydroxy-*N*-methyl-L-proline (*cis*-3-OH-*N*-Me-L-Pro);

(b) *cis*P4H produces *cis*-4-hydroxy-*N*-methyl-L-proline (*cis*-4-OH-*N*-Me-L-Pro);

(c) *trans*P4H produces *trans*-4-hydroxy-*N*-methyl-L-proline (*trans*-4-OH-*N*-Me-L-Pro) and *cis*-4-hydroxy-*N*-methyl-L-proline (*cis*-4-OH-*N*-Me-L-Pro) in an approximate 5:1 ratio.

LC/MS analysis of the *N*-Me-L-Pro PH reactions showed clear hydroxylation products (Figure 3.18); extracted-ion chromatograms displaying 146.1 *m/z* ions showed intense peaks corresponding to L-hydroxyprolines, proposed to be *cis*-3-hydroxy-*N*-methyl-L-proline (*cis*-3-OH-*N*-Me-L-Pro - retention time = ~19 min), *cis*-4-hydroxy-*N*-methyl-L-proline (*cis*-4-OH-*N*-Me-L-Pro - retention time = ~9.5 min), *trans*-4-hydroxy-*N*-methyl-L-proline (*trans*-4-OH-*N*-Me-L-Pro - retention time = ~8 min) for the *cis*P3H Type I, *cis*P4H, and *trans*P4H reactions, respectively. Interestingly, analysis of the *trans*P4H reaction with *N*-Me-L-Pro also showed a second, smaller peak in the 146.1 *m/z* extracted-ion chromatogram (XIC) with a retention time of ~9.5 min, the same retention time as seen for the product of the *N*-Me-L-Pro *cis*P4H reaction (presumably *cis*-4-OH-*N*-Me-L-Pro); this is similar to what was observed for the wild-type *trans*P4H L-Pro reaction, where it appears (by LC/MS) that a very small amount of *cis*-4-hydroxylation; however, in the case of the *N*-Me-L-Pro

reaction, the side reaction is even more prominent (peak integration showed that major and minor products were generated in an approximate ratio of 5:1).

Hydroxylation of *N*-Me-L-Pro by *cis*P3H Type I and *cis*P4H gave very low yields over 14 h of incubation (for maximal turnover); LC/MS peak integration was used to calculate the yields, which were 100% ('complete conversion') and 97% for *cis*P3H Type I and *cis*P4H, respectively. The hydroxylation yields for the *N*-Me-L-Pro *trans*P4H reactions were poor in comparison; however, peak integration was used to show that only ~6% total conversion was seen (to give ~5% *trans*-4-OH-*N*-Me-L-Pro and ~1% *cis*-4-OH-*N*-Me-L-Pro).

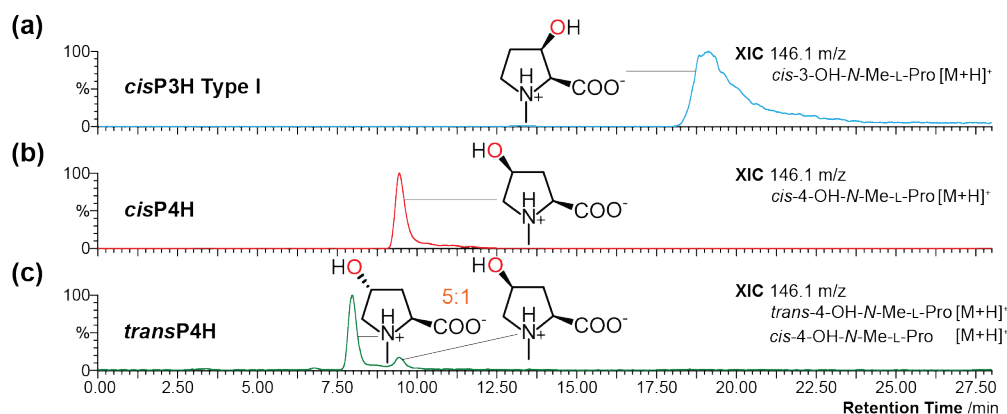


Figure 3.18: LC/MS-based analysis of proline hydroxylase reactions with *N*-methyl-L-proline (*N*-Me-L-Pro) [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H].

Extracted-ion chromatograms (XICs) (showing 146.1 *m/z* ions) contain peaks thought to correspond to (a) *cis*-3-OH-*N*-Me-L-Pro, (b) *cis*-4-OH-*N*-Me-L-Pro, and (c) *trans*-4-OH-*N*-Me-L-Pro (and *cis*-4-OH-*N*-Me-L-Pro), respectively.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 10 mM 2-OG, 1 mM *N*-Me-L-Pro. 100 μ L reactions were carried out in at room temperature (25 $^\circ$ C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Complete reactions were quenched with 100 μ L methanol (1:1) containing a *para*-aminosalicylic acid (0.25 mM *p*ASA) standard, and then microcentrifuged to remove precipitate. Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 $\text{\AA}$, 5 μ m) [4.6 x 250mm].

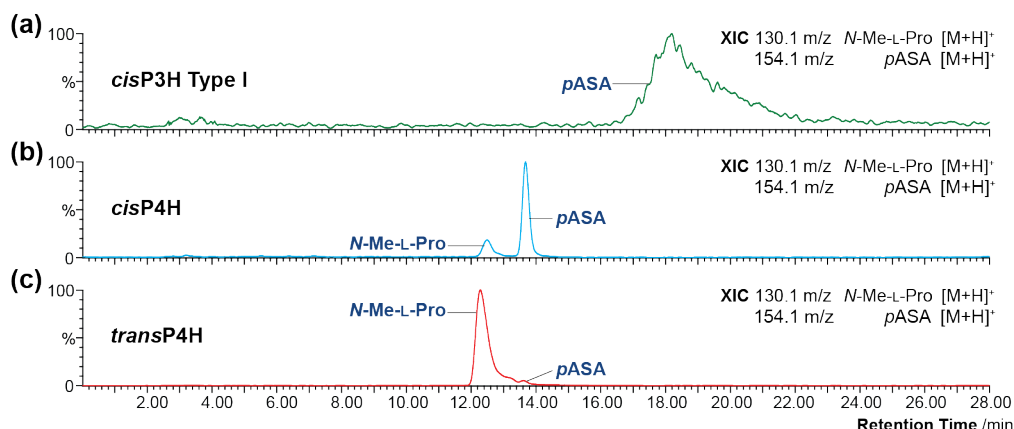


Figure 3.19: LC/MS-based analysis of proline hydroxylase reactions with *N*-methyl-L-proline (*N*-Me-L-Pro) [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H]. Extracted-ion chromatograms (XICs) (showing 130.1 m/z ions and 154.1 m/z ions) contain peaks corresponding to (a) *para*-amino-salicylic acid (*p*ASA - internal standard), (b) *N*-Me-L-Pro and *p*ASA, and (c) *N*-Me-L-Pro and *p*ASA, respectively.

The *N*-Me-L-Pro proline hydroxylase reactions were scaled up 2×10 -fold (2×1 mL) and the products were purified by LC/MS methods developed for larger scale chromatography (see Section 5.2.2.3); NMR characterisation experiments were carried out to confirm the stereochemistry of the products. For example, the COSY and NOESY spectra measured for the *cis*P3H Type I product, *cis*-3-hydroxy-L-methyl-L-proline (*cis*-3-OH-*N*-Me-L-Pro) (Figures 3.20 and 3.21).

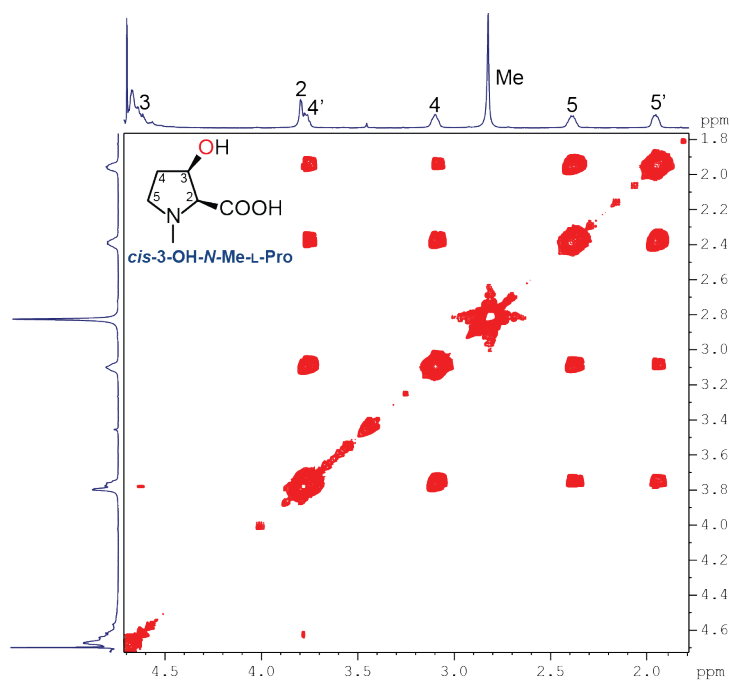


Figure 3.20: COSY-NMR analysis of the product from the *cis*P3H Type I reaction with *N*-methyl-L-proline (*N*-Me-L-Pro): NMR assignments confirmed that the product is *cis*-3-hydroxy-*N*-methyl-L-proline (*cis*-3-OH-*N*-Me-L-Pro).

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D_2O containing 0.05% (w/v) TSP- d_4 , NMR chemical shift reference/internal standard.

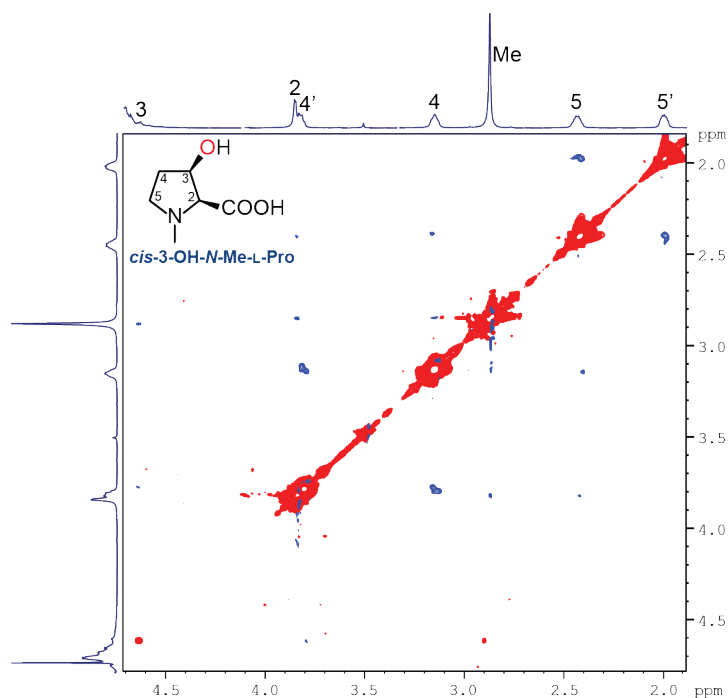


Figure 3.21: NOESY-NMR analyses of the product from the *cis*P3H Type I reaction with *N*-methyl-L-proline (*N*-Me-L-Pro): NMR assignments confirmed that the product is *cis*-3-hydroxy-*N*-methyl-L-proline (*cis*-3-OH-*N*-Me-L-Pro).

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D_2O containing 0.05% (w/v) TSP- d_4 , NMR chemical shift reference/internal standard.

To further test the effect of carboxyl functionalisation on substrate turnover by the PH enzymes, a carboxylic methyl ester, L-proline methyl ester (L-Pro-OMe) (Figure 3.17), was tested as a substrate for *cis*P3H Type I, *cis*P4H, and *trans*P4H; given that L-Pro-OMe is L-Pro with a methyl group attached to the carboxyl oxygen atom, it was thought that it may be similar enough to L-Pro to be hydroxylated. LC/MS analysis of 1 h incubations of L-Pro-OMe with the three PH enzymes showed the reactions contained L-Pro, as well as *cis*-3-OH-L-Pro (Figure 3.22 a - ~96% yield), *cis*-4-OH-L-Pro (Figure 3.22 b - ~85% yield), and *trans*-4-OH-L-Pro (Figure 3.22 c - ~17% yield), respectively; no L-Pro-OMe or hydroxy-L-Pro-OMe products were detected, however, only their hydrolysis products (Figure 3.22). It is unclear whether the hydrolysis occurs prior to PH hydroxylation or after PH hydroxylation. Under

the reaction conditions used (pH 6.5), the amino group was probably protonated (with a net positive charge), attracting water-solvent molecules close to the carbonyl group of the methyl ester, which could potentially promote ester hydrolysis [i.e. $R_{Ring}-COOMe \xrightleftharpoons{\pm H^+} R_{Ring}-COOH + MeOH$]; this hydrolysis is also a plausible occurrence in the HPLC solvents used [which contain 0.1% (v/v) formic acid (FA)], since the hydrolysis reaction would be acid catalysed [i.e. protonation of the carbonyl oxygen atom increases nucleophilic attack from water molecules, since the carbonyl carbon partial-positive charge ($\delta+$) increases]. If L-Pro-OMe and L-Hyp-OMe compounds are stable for long enough (towards hydrolysis) under reaction conditions, it may be possible to observe these species in a short-incubation reaction using an NMR method; direct observation of hydroxylated L-proline methyl esters would suggest L-Pro-OMe is a PH substrate, whereas detection of only L-hydroxyprolines may suggest that L-Pro-OMe must be hydrolysed first before any PH hydroxylation can occur.

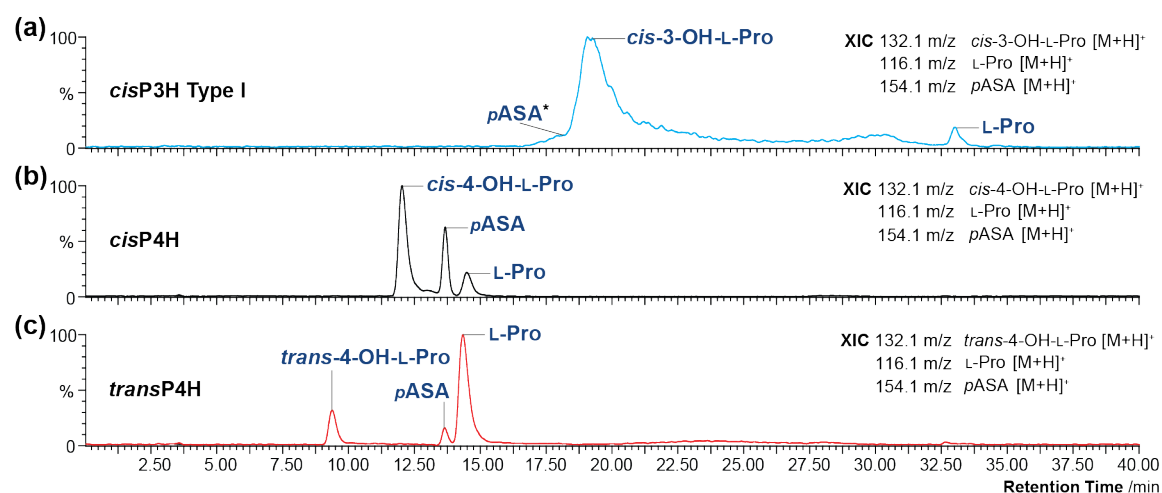


Figure 3.22: LC/MS-based analysis of proline hydroxylase reactions with L-proline methyl ester (L-Pro-OMe) [(a) *cisP3H Type I*, (b) *cisP4H*, and (c) *transP4H*].

Extracted-ion chromatograms (XICs) (showing 116.1 m/z, 130.1 m/z, and 154.1 m/z ions) contain peaks corresponding to the hydroxyprolines (all 130.1 m/z [M+H]⁺ ions), as well as *para*-aminosalicylic acid (*pASA* - internal standard) (154.1 m/z [M+H]⁺ ion) and L-Pro (116.1 m/z [M+H]⁺ ion).

PH reactions used 15 μ M *cisP3H Type I/cisP4H/transP4H*, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM L-Pro-OMe.

3.4.4.2.1 Substrate Turnover Summary: Amino- & Carboxylic-Functionalisation

The following table (Tables 3.7) summarises the proline hydroxylase activities with the amino- and carboxylically functionalised substrates screened for proline hydroxylase substrate activity in this section. See below for details of the reaction yields and products.

CHAPTER 3. THE L-PROLINE HYDROXYLASES (PHS)

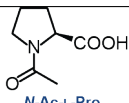
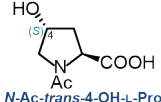
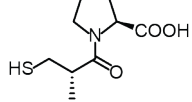
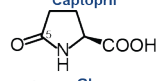
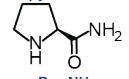
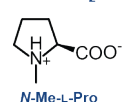
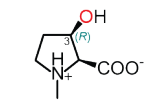
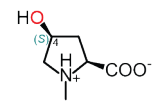
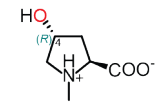
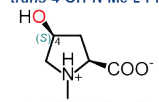
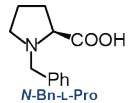
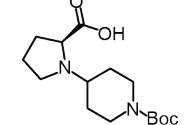
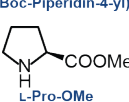
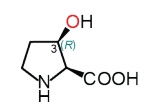
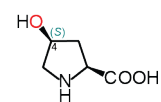
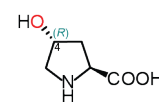
Substrate	Length/h	Product(s) [+ Yield(s)]					
		cisP3H Type I		cisP4H		transP4H	
 <i>N</i> -Ac-L-Pro	14	-	0%	-	0%	-	0%
 <i>N</i> -Ac- <i>trans</i> -4-OH-L-Pro	14	-	0%	-	0%	-	0%
 Captopril	14	-	0%	-	0%	-	0%
 L-pyroGlu	14	-	0%	-	0%	-	0%
 L-Pro-NH ₂	14	-	0%	-	0%	-	0%
 <i>N</i> -Me-L-Pro	14	 <i>cis</i> -3-OH- <i>N</i> -Me-L-Pro	100%	 <i>cis</i> -4-OH- <i>N</i> -Me-L-Pro	97%	 <i>trans</i> -4-OH- <i>N</i> -Me-L-Pro	5%
		-	-			 <i>cis</i> -4-OH- <i>N</i> -Me-L-Pro	1%
 <i>N</i> -Bn-L-Pro	14	-	0%	-	0%	-	0%
 <i>N</i> -(<i>N</i> -Boc-Piperidin-4-yl)-L-Pro	14	-	0%	-	0%	-	0%
 L-Pro-OMe	14	 <i>cis</i> -3-OH-L-Pro	96%	 <i>cis</i> -4-OH-L-Pro	85%	 <i>trans</i> -4-OH-L-Pro	17%

Table 3.7: Proline hydroxylase substrate analogue screening: the effects of substrate amino- and carboxylic functionalisation on proline hydroxylase activity.

D-Proline (D-Pro), D-pipecolic acid (D-Pip), DL-pipecolic acid (DL-Pip), D-pyrroglutamic acid (D-pyroGlu), and *N*-acetyl-*trans*-4-hydroxy-D-proline (*N*-Ac-*trans*-4-OH-D-Pro) were screening for proline hydroxylase substrate activity and compared to the natural substrate, L-Pro.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM Substrate.

3.4.4.3 Effects of Carbon Backbone Substitutions

To investigate the effects of carbon-backbone functionality (*i.e.* substitutions at ring positions C2-C5) on proline hydroxylase activity, a number of proline analogues (Figure 3.23) were screened for catalytic turnover by the three PH enzymes (*cis*P3H Type I, *cis*P4H, and *trans*P4H); these compounds included proline derivatives with alkylation (some with additional groups on the alkyl chain), halogenation (specifically fluorination), hydroxylation, carbonylation, and carboxymethylation (and derivatives thereof), as well as some compounds with a combination of modifications (Figure 3.23).

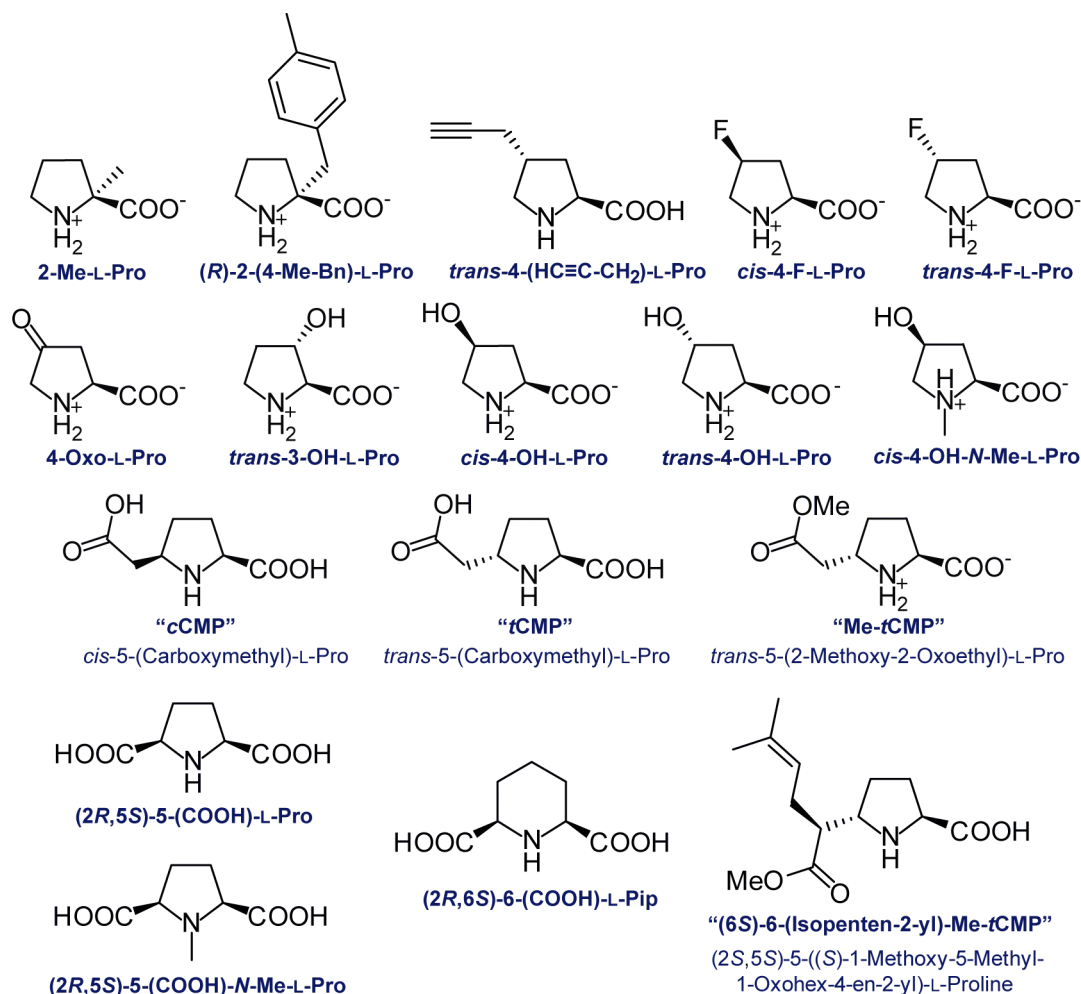


Figure 3.23: Proline analogues with carbon-backbone functionality: substitutions occupy positions C2-C5, or a combination there of.

Compounds include: 2-methyl-L-proline (2-Me-L-Pro), (*R*)-2-(4-methylbenzyl)-L-proline [(*R*)-2-(4-Me-Bn)-L-Pro], *cis*-4-fluoro-L-proline (*cis*-4-F-L-Pro), *trans*-4-fluoro-L-proline (*cis*-4-F-L-Pro), 4-oxo-L-proline (4-Oxo-L-Pro), *trans*-3-hydroxy-L-proline (*trans*-3-OH-L-Pro), *cis*-4-hydroxy-L-proline (*cis*-4-OH-L-Pro), *trans*-4-hydroxy-L-proline (*trans*-4-OH-L-Pro), *cis*-4-hydroxy-*N*-methyl-L-proline (*cis*-4-OH-*N*-Me-L-Pro), (2*S*,5*S*)-5-(carboxymethyl)-L-proline ('*cis*-carboxymethylproline' - cCMP), (2*S*,5*R*)-5-(carboxymethyl)-L-proline ('*trans*-carboxymethylproline' - tCMP), (2*S*,5*R*)-5-(2-methoxy-2-oxoethyl)-L-proline ('*O*-methyl-*trans*-carboxymethylproline' - Me-tCMP), (2*S*,5*S*)-5-[(*S*)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline ['(6*S*)-isopenten-2-yl-*trans*-carboxymethylproline' - (6*S*)-isopenten-2-yl-tCMP], (2*S*,5*S*)-5-carboxyl-L-proline [(2*S*,5*S*)-5-(COOH)-L-Pro], (2*S*,5*S*)-5-carboxyl-*N*-methyl-L-proline [(2*S*,5*S*)-5-(COOH)-*N*-Me-L-Pro], (2*S*,6*S*)-6-carboxyl-L-pipecolic acid [(2*S*,5*S*)-6-(COOH)-L-Pip], and *trans*-4-(propyn-1-yl)-L-proline [*trans*-4-(HC≡CH₂)-L-Pro].

3.4.4.3.1 Alkylated and Isoprenylated Proline Analogues

In a recent study carried out during my work, Klein *et al.* reported catalytic *cis*-3-hydroxylation of *trans*-3-methyl-L-proline by both *cis*P3Hs and *cis*P4Hs, reactions which produce a tertiary alcohol compound, (3*R*)-3-hydroxy-3-methyl-L-proline (discussed in Section 3.4.3); based on this evidence, and the broad selectivity we observed for these enzymes, we thought that other proline derivatives with alkyl functionality/isoprenylation may be proline hydroxylase substrates. Initially, two structurally related compounds (Figure 3.24), 2-methyl-L-proline (2-Me-L-Pro) and (*R*)-2-(4-methylbenzyl)-L-proline [(*R*)-2-(4-Me-Bn)-L-Pro], were tested for proline hydroxylase substrate activity; prior to this study, no proline derivatives with additional C-2 functionality had been tested as proline hydroxylase substrates.

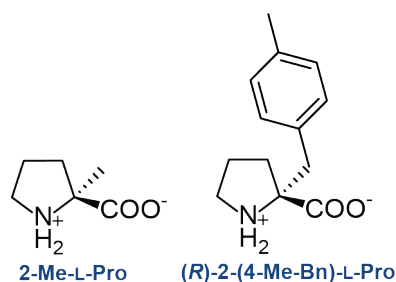
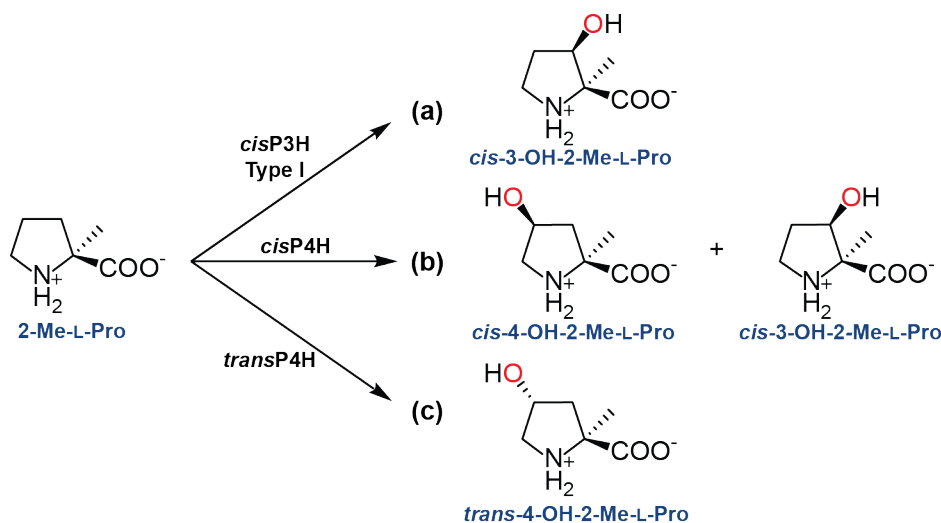


Figure 3.24: 2-Methyl-L-proline (2-Me-L-Pro) and (*R*)-2-(4-methylbenzyl)-L-proline [(*R*)-2-(4-Me-Bn)-L-Pro] were tested for proline hydroxylase substrate activity.

Proline hydroxylase reactions with (*R*)-2-(4-methylbenzyl)-L-proline [(*R*)-2-(4-Me-Bn)-L-Pro] contained no LC/MS-detectable products, suggesting that (*R*)-2-(4-Me-Bn)-L-Pro is not a proline hydroxylase substrate; the large 4-methylbenzyl group at carbon-2 is probably too large to be accommodated by the active sites of the PH enzymes. By contrast, LC/MS-analysis of PH reactions using a 2-Me-L-Pro substrate showed that 2-Me-L-Pro is a substrate of all three proline hydroxylases (*cis*P3H Type I, *cis*P4H, *trans*P4H) (Schemes 3.12 a-c); extracted-ion chromatograms (XICs) showing 146.1 *m/z* ions contained peaks which correspond the presence of hydroxy-2-Me-L-Pros (Figures 3.25 a-c). A 2-methyl group must be small enough to allow

2-Me-L-Pro to bind to proline hydroxylase active sites.



Scheme 3.12: Proposed hydroxylation reactions catalysed by the proline hydroxylases using an 2-methyl-L-proline (2-Me-L-Pro) substrate [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H]:

- (a) *cis*P3H Type I produces *cis*-3-hydroxy-2-methyl-L-proline (*cis*-3-OH-2-Me-L-Pro);
- (b) *cis*P4H produces *cis*-4-hydroxy-2-methyl-L-proline (*cis*-4-OH-2-Me-L-Pro) and *cis*-3-hydroxy-2-methyl-L-proline (*cis*-3-OH-2-Me-L-Pro) in an approximate 3:1 ratio;
- (c) *trans*P4H produces *trans*-4-hydroxy-2-methyl-L-proline (*trans*-4-OH-2-Me-L-Pro).

The *cis*P3H Type I reaction contained a single product peak (retention time = ~15.75 min), which was proposed to correspond to *cis*-3-hydroxy-2-methyl-L-proline (*cis*-3-OH-2-methyl-L-Pro) (Figure 3.25 a); yield calculations (based on peak-integration values) showed that the yield of this reaction was only ~1%. The *cis*P4H reactions contained two products with retention times of ~13.75 min and ~15.75 min (Figure 3.25 b), respectively; the ‘first’, major peak is proposed to be *cis*-4-OH-2-Me-L-Pro the ‘second’ peak has the same retention time as the product of the *cis*P3H Type I, suggesting these products are the same substance (*i.e.* *cis*-3-OH-2-Me-L-Pro). Peak integration showed that the ‘major’ and ‘minor’ products, *cis*-4-OH-2-Me-L-Pro and *cis*-3-OH-2-Me-L-Pro, are generated in an approximate 3:1 ratio by the *cis*P4H reaction, with yields of ~3% and ~1%, respectively (total yield = ~4%). The LC/MS chromatogram for the *trans*P4H 2-Me-L-Pro reaction contained a single

product peak (retention time = ~9.75 min) (Figure 3.25 c); the product is thought to be *trans*-4-hydroxy-2-methyl-L-proline (*trans*-4-OH-2-Me-L-Pro), and is produced with a yield of ~5% under the conditions used. Interestingly, the *trans*P4H reaction gave the highest yield (and *cis*P3H Type I the lowest), which is the opposite trend usually observed for proline hydroxylase reactions with many other proline analogues; in a number of cases, *trans*P4H is the least promiscuous enzyme of these 3 enzymes, having the most strict substrate selectivity, and often produces hydroxylation products of slightly lower yields (if any at all) than *cis*P3H Type I and *cis*P4H.

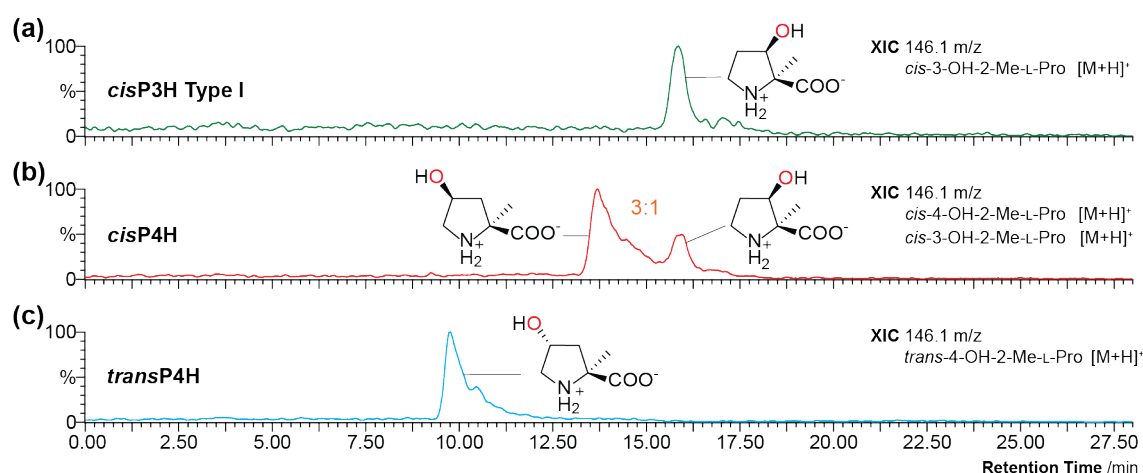


Figure 3.25: LC/MS-based analysis of proline hydroxylase reactions with 2-methyl-L-proline (2-Me-L-Pro) [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H].

Extracted-ion chromatograms (XICs) (showing 146.1 m/z ions) contain peaks thought to correspond to (a) *cis*-3-OH-2-Me-L-Pro, (b) *cis*-4-OH-2-Me-L-Pro and *cis*-3-OH-2-Me-L-Pro, and (c) *trans*-4-OH-2-Me-L-Pro, respectively.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 10 mM 2-OG, 1 mM 2-Me-L-Pro. 100 μ L reactions were carried out in at room temperature (25 $^\circ$ C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 $\text{\AA}$, 5 μ m) [4.6 x 250mm].

The 2-Me-L-Pro hydroxylation yields are poor for all three PH enzymes, revealing that 2-Me-L-Pro is a poor PH enzyme substrate. Perhaps, the 2-methyl group of 2-Me-L-Pro weakens PH substrate binding (e.g. by steric hindrance), and/or alters the alignment of the substrate's 5-membered ring in a way which makes hydroxylation

tion less favourable; what it is clear, however, is that catalysis is ‘hindered’ the least for *trans*P4H reactions, which gives better yields than the corresponding *cis*P3H Type I and *cis*P4H reactions. These results suggest that *trans*P4H may be ‘more accepting’ of C2-functionalised substrates than the other two PH enzymes.

The products of the 2-Me-L-Pro proline hydroxylase reactions were purified for NMR characterisation, in order to confirm the product stereochemistries. For example, Figure 3.26 (below) shows the COSY-spectrum of the major product of the reaction of *cis*P4H with 2-Me-L-Pro; NMR assignments are consistent with the product being *cis*-4-hydroxy-2-methyl-L-proline (*cis*-4-OH-2-Me-L-Pro).

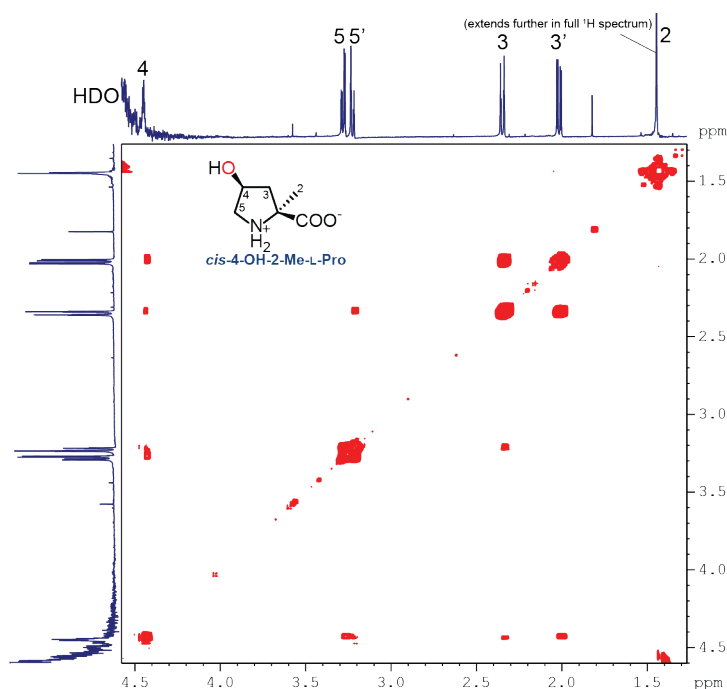


Figure 3.26: COSY-NMR analysis of the product from the *cis*P4H reaction with 2-methyl-L-proline (2-Me-L-Pro): NMR assignments confirmed that the product is *cis*-4-hydroxy-2-methyl-L-proline (*cis*-4-OH-2-Me-L-Pro).

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a ¹H/¹³C/¹⁵N TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D₂O containing 0.05% (w/v) TSP-*d*₄, NMR chemical shift reference/internal standard.

A *trans*-4-alkynyl compound, *trans*-4-(2-propynyl)-L-proline [*trans*-4-(HC≡C-CH₂)-L-Pro] (Figure 3.27), was also tested as a substrate of the proline hydroxylases (*cis*P3H Type I, *cis*P4H, and *trans*P4H), in order to determine the effect of isoprenylation of other proline ring positions on proline hydroxylase activity.

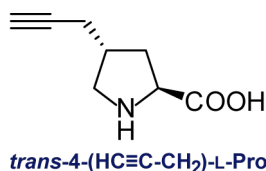


Figure 3.27: A *trans*-4-(2-propynyl)-L-proline [*trans*-4-(HC≡C-CH₂)-L-Pro] substrate with tested for proline hydroxylase turnover

LC/MS-analysis of *cis*P3H Type I and *cis*P4H reactions using a *trans*-4-(2-propynyl)-L-proline [*trans*-4-(HC≡C-CH₂)-L-Pro] substrate contained a hydroxylation product thought be (3*R*,4*S*)-3-hydroxy-4-(2-propynyl)-L-proline [(3*R*,4*S*)-3-OH-4-(HC≡C-CH₂)-L-Pro] (Figures 3.28 a and b); LC/MS extracted-ion chromatograms (XICs) showing 170.1 m/z [M+H]⁺ ions contained a single peak with a retention time of ~15 min (Figures 3.28 a and b). Peak integrals indicated that the yields of these *cis*P3H Type I and *cis*P4H reactions were ~26% and ~21%, respectively. Conversely, *trans*-4-(HC≡C-CH₂)-L-Pro was not a substrate of *trans*P4H, presumably because the site of propynylation is blocking the preferred *trans*-4-hydroxylation selectivity of *trans*P4H.

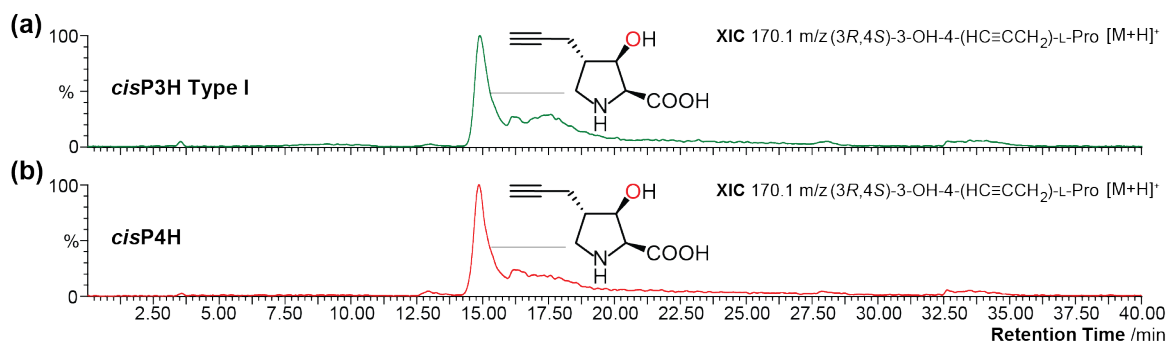


Figure 3.28: LC/MS-based analysis of proline hydroxylase reactions with *trans*-4-(2-propynyl)-L-proline [*trans*-4-(HC≡C-CH₂)-L-Pro].

Extracted-ion chromatograms (XICs) (showing 170.1 m/z ions) contain peaks thought to correspond to $(3R,4S)$ -3-hydroxy-4-(2-propynyl)-L-proline [$(3R,4S)$ -3-OH-4-(HC≡C-CH₂)-L-Pro].

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM *trans*-4-(HC≡C-CH₂)-L-Pro. 100 μ L reactions were carried out in at room temperature (25 °C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μ m) [4.6 x 250mm].

The *cis*P3H Type I reaction with *trans*-4-(HC≡C-CH₂)-L-Pro was scaled up 2 × 10-fold (2 × 1 mL), and the $(3R,4S)$ -3-hydroxy-4-(2-propynyl)-L-proline [$(3R,4S)$ -3-OH-4-(HC≡C-CH₂)-L-Pro] product was purified using an LC/MS (HPLC) method; NMR characterisation experiments were carried out, in order to confirm the identity of the product. For example, Figure 3.29 (below) shows a COSY spectrum, which supports the identity of the product.

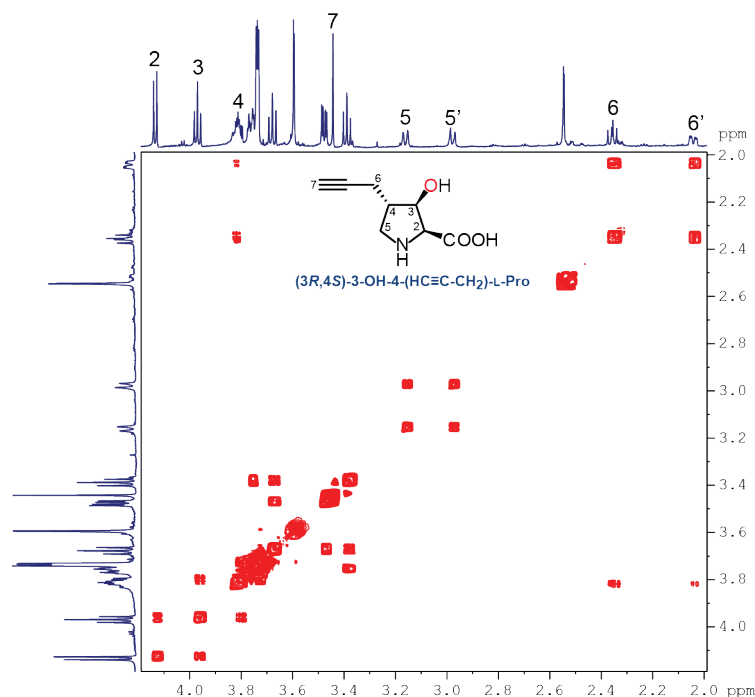
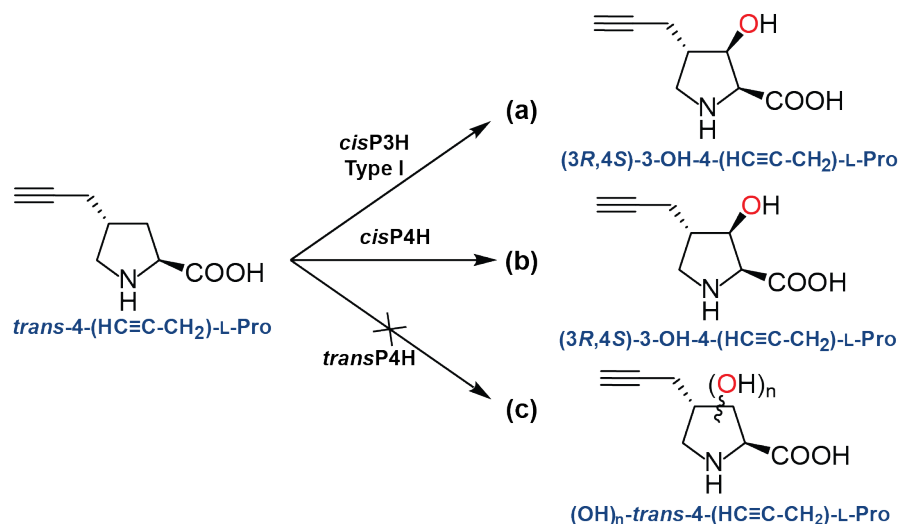


Figure 3.29: COSY-NMR analysis of the product from the *cis*P3H Type I/*cis*P4H reactions with *trans*-4-(2-propynyl)-L-proline [*trans*-4-(HC≡C-CH₂)-L-Pro]: NMR assignments confirmed that the product is (3*R*,4*S*)-3-hydroxy-4-(2-propynyl)-L-proline [(3*R*,4*S*)-3-OH-4-(HC≡C-CH₂)-L-Pro].

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a ¹H/¹³C/¹⁵N TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D₂O containing 0.05% (w/v) TSP-*d*₄, NMR chemical shift reference/internal standard.

In summary, the following reactions are proposed to occur when the proline hydroxylase enzymes use a *trans*-4-(2-propynyl)-L-proline [*trans*-4-(HC≡C-CH₂)-L-Pro] substrate.



Scheme 3.13: Proline hydroxylase reactions using a *trans*-4-(2-propynyl)-L-proline [*trans*-4-(HC≡C-CH₂)-L-Pro] substrate [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H]: *trans*-4-(HC≡C-CH₂)-L-Pro is thought to be converted to (3*R*,4*S*)-3-hydroxy-4-(2-propynyl)-L-proline [(3*R*,4*S*)-3-OH-4-(HC≡C-CH₂)-L-Pro], whereas *trans*-4-(HC≡C-CH₂)-L-Pro is not a *trans*P4H substrate.

3.4.4.3.2 Substrate Turnover Summary: Alkyl & Other Non-Polar Substituents

The following table (Table 3.8) summarises the proline hydroxylase activities with substrates with alkyl and other non-polar functionalities. See below for details of the reaction yields and products.

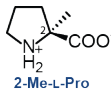
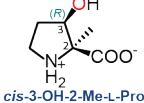
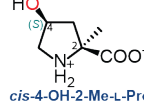
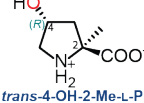
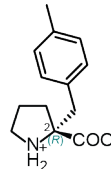
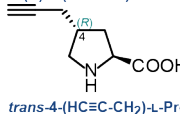
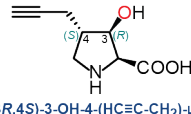
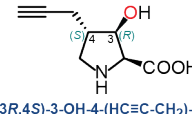
Substrate	Length/h	Product(s) [+ Yield(s)]					
		<i>cis</i> P3H Type I	<i>cis</i> P4H	<i>trans</i> P4H			
 2-Me-L-Pro	14	 <i>cis</i> -3-OH-2-Me-L-Pro	1%	 <i>cis</i> -4-OH-2-Me-L-Pro	4%	 <i>trans</i> -4-OH-2-Me-L-Pro	100%
 (<i>R</i>)-2-(4-Me-Bn)-L-Pro	14	-	0%	-	0%	-	0%
 <i>trans</i> -4-(HC≡C-CH ₂)-L-Pro	14	 (3 <i>R</i> ,4 <i>S</i>)-3-OH-4-(HC≡C-CH ₂)-L-Pro	26%	 (3 <i>R</i> ,4 <i>S</i>)-3-OH-4-(HC≡C-CH ₂)-L-Pro	21%	-	0%

Table 3.8: Proline hydroxylase substrate analogue screening: the effects of substrate hydroxyl functionalities on proline hydroxylase activity.

2-Methyl-L-proline (2-Me-L-Pro), (R)-2-(4-methylbenzyl)-L-proline [(R)-2-(4-Me-Bn)-L-Pro], and *trans*-4-(2-propynyl)-L-proline [*trans*-4-(HC≡C-CH₂)-L-Pro] were screened as proline hydroxylase substrates.

PH reactions used 15 μM *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μM (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM Substrate.

3.4.4.3.3 Halogenated Proline Analogues

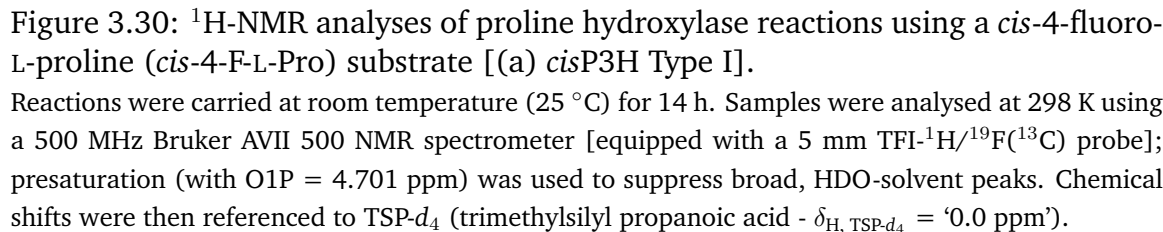
In order to test the effect of halogen functionalities on proline hydroxylase activity, fluorinated L-prolines were also tested as substrates of the three proline hydroxylases (*cis*P3H Type I, *cis*P4H, and *trans*P4H); *cis*-4-fluoro-L-proline (*cis*-4-F-L-Pro) and *trans*-4-fluoro-L-proline (*trans*-4-F-L-Pro) were selected as proline hydroxylase test substrates. Fluorine atoms are small (not much bigger than a hydrogen atom), and have been known to sterically ‘mimic’ hydrogen atoms; therefore, these C4-fluorinated substrates were thought to be possible PH substrates. Despite the high polarity of C—F bonds (owing to the high electronegativity of fluorine), fluorinated prolyl substrates have been reported to be active substrates of other ODD enzymes; for example, the HIF prolyl hydroxylase, PHD2, has been reported to hydroxylate a *cis*-4-fluoro-prolyl derivative of its CODD substrate.³⁷⁸

Early proline hydroxylase reactions using *cis*-4-F-L-Pro and *trans*-4-F-L-Pro substrates were carried out in D₂O containing deuterated buffers (50 mM Tris-*d*₁₁-DCl, pD 7.5) suitable for NMR analyses; the reactions were then analysed using both ¹H- and ¹⁹F-NMR experiments to determine if any catalytic turnover had taken place (by looking for a decrease in starting material peak intensities and the appearance of new resonance peaks).

Proton NMR analyses of proline hydroxylase reactions using a *cis*-4-F-L-Pro substrate showed that *cis*P3H Type I consumed the starting material and produced at least one product (Figures 3.30 a); NMR spectra of the *cis*P3H Type I reactions (Figures 3.30 a) contained new peaks that were not present in the negative control (‘no enzyme’) reaction (Figure 3.30 b). For example, there are prominent double-multiplet peaks centred around ~5.3 ppm and ~4.65 ppm in *cis*P3H Type I reaction spectra, suggesting both enzymes produce the same, *cis*-3-hydroxylation product {thought to be (3*S*,4*R*)-4-fluoro-3-hydroxy-L-proline [(3*S*,4*R*)-4-F-3-OH-L-Pro]}. Based upon chemical shifts and peak multiplicity, the ~5.3 ppm double-

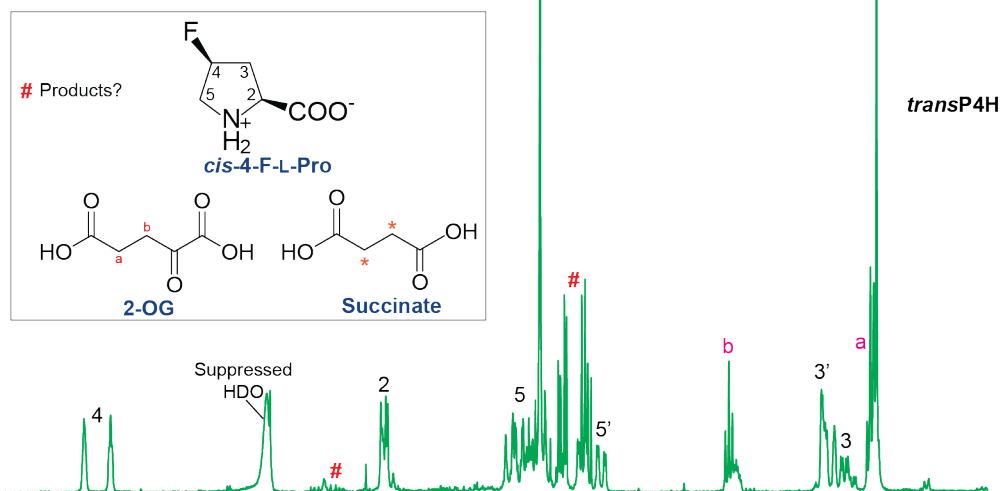
multiplet peak was thought to be produced by a C4-proton (with a large geminal coupling (2J) to a C4- ^{19}F atom), and the ~ 4.65 ppm peak was thought to be from a proton attached to the hydroxylated carbon (C3).

Proton NMR analyses of *trans*P4H reactions using a *cis*-4-F-L-Pro substrate (Figure 3.31 a) showed new peaks but they appeared to be small or overlapping with much larger starting material peaks, making it difficult to draw any meaningful conclusions from these spectra; however, 2-OG turnover by *trans*P4H appeared to be stimulated by the presence of *cis*-4-F-L-Pro, since the *trans*P4H reaction spectra contain an intense singlet (~ 2.4 ppm) (Figure 3.31 a), corresponding to the formation of succinate, as well as a marked reduction in the 2-OG peak intensities (~ 2.8 ppm and ~ 2.35 ppm) compared to the negative control (Figure 3.31 b).

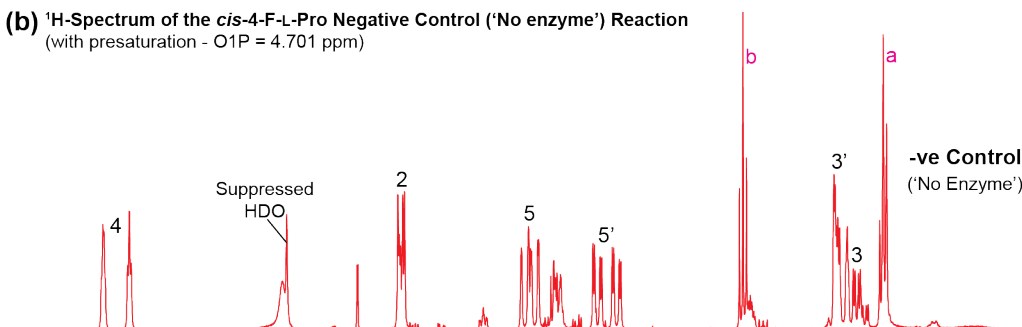


(a) ^1H -Spectrum of the *trans*P4H Reaction with *cis*-4-fluoro-L-proline (*cis*-4-F-L-Pro)
(with presaturation - O1P = 4.701 ppm)

1 mM *cis*-4-F-L-Pro
500 μM Sodium 2-OG
250 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$
1 mM Sodium Ascorbate
15 μM (~0.5mg/ml) *trans*P4H
50 mM Tris pD 7.5, D_2O



(b) ^1H -Spectrum of the *cis*-4-F-L-Pro Negative Control ('No enzyme') Reaction
(with presaturation - O1P = 4.701 ppm)



(c) ^1H -Spectrum of the 5mM *cis*-4-F-L-Pro in 50 mM Tris- d_{11} -DCl, D_2O
(with presaturation - O1P = 4.701 ppm)

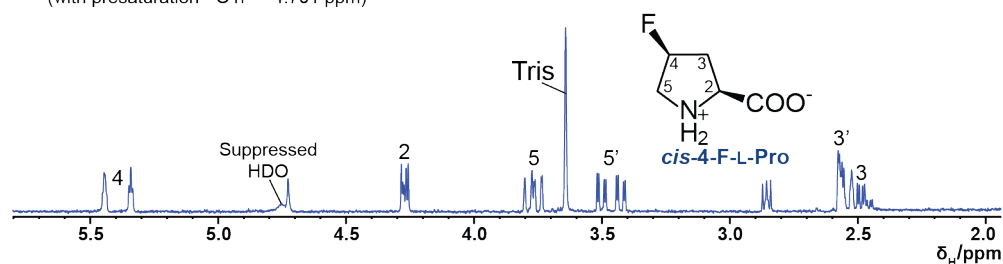


Figure 3.31: ^1H -NMR analyses of proline hydroxylase reactions using a *cis*-4-fluoro-L-proline (*cis*-4-F-L-Pro) substrate [(a) *trans*P4H].

Reactions were carried at room temperature (25 °C) for 14 h. Samples in 5 mm tubes were analysed at 298 K using a 500 MHz Bruker AVII 500 NMR spectrometer [equipped with a 5 mm TFI- $^1\text{H}/^{19}\text{F}$ (^{13}C) probe]; presaturation (with O1P = 4.701 ppm) was used to suppress broad, HDO-solvent peaks. Chemical shifts were then referenced to TSP- d_4 (trimethylsilyl propanoic acid - $\delta_{\text{H, TSP-d}_4}$ = '0.0 ppm').

^{19}F -NMR analysis of the same samples analysed in Figures 3.30 and 3.31 showed that a fluorine-containing product was generated from *cis*-4-F-L-Pro by *cis*P3H Type I (Figure 3.32 a); in addition to *cis*-4-F-L-Pro starting material peaks (~ 175.5 ppm), these spectra also contained additional peaks at ~ 202 ppm, consistent with a (3*S*,4*R*)-4-F-3-OH-L-Pro product being generated. Interestingly, proton-decoupled ^{19}F -NMR spectra for the *trans*P4H reaction with *cis*-4-F-L-Pro showed a small peak at ~ 122 ppm, suggesting that *cis*-4-F-L-Pro is also a substrate of *trans*P4H. The *trans*P4H product peak has roughly the same chemical shift value as aqueous potassium fluoride (previous measurements by Prof. Tim Claridge showed $\delta_{\text{H, TSP-}d_4} = -125.3$ ppm), which suggests that hydroxylation of *cis*-4-F-L-Pro by *trans*P4H results in F^- release; the *trans*-4-hydroxylation of *cis*-4-F-L-Pro would produce (4*S*)-4-fluoro-4-hydroxy-L-proline [(4*S*)-4-F-4-OH-L-Pro] (Scheme 3.14), an unstable ‘tetrahedral intermediate’, which fragments to produce 4-oxo-L-proline (4-Oxo-L-Pro) and fluoride. A similar reaction is seen when the HIF prolyl hydroxylase, PHD2, *trans*-4-hydroxylates a *cis*-4-fluoro-prolyl derivative of its CODD substrate;³⁷⁸ matrix-assisted desorption/ionisation time-of-flight mass spectrometry (MALDI-TOF-MS) analyses of these PHD2 reactions showed a product 4 *m/z* units smaller than the starting material (consistent with 4-oxo-prolyl CODD formation), and 2,4-dinitrophenylhydrazine (2,4-DNP) derivatisation experiments confirmed the product was a ketone.

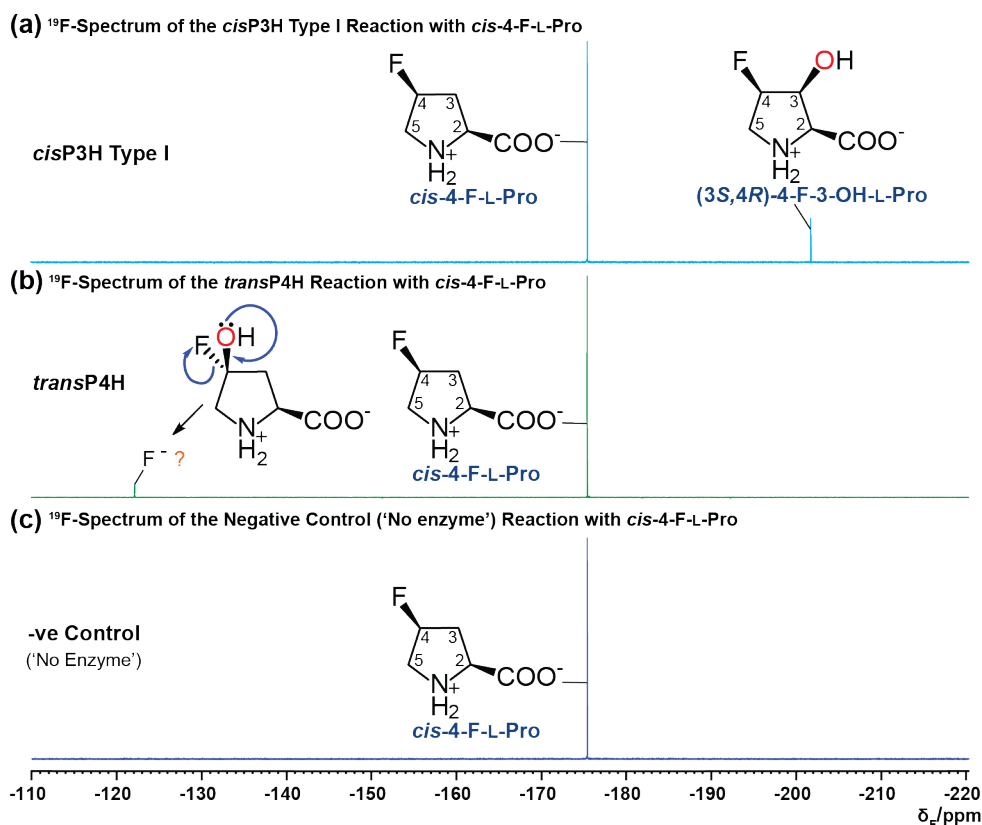
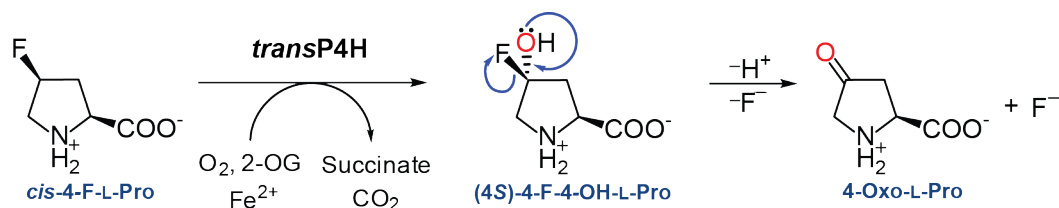


Figure 3.32: Proton-decoupled ^{19}F -NMR analyses of proline hydroxylase reactions using a *cis*-4-fluoro-L-proline (*cis*-4-F-L-Pro) substrate [(a) *cis*P3H Type I and (b) *trans*P4H].

Proton-decoupled ^{19}F -spectra were measured at 298 K for samples in 5 mm tubes using a 470.4 MHz (^{19}F) Bruker AVII 500 NMR spectrometer [equipped with a 5 mm TFI- $^1\text{H}/^{19}\text{F}(^{13}\text{C})$ probe]. Chemical shift values are referenced to those of CFCl_3 (where $\delta_{\text{F}, \text{CFCl}_3} = '0.0 \text{ ppm}'$).



Scheme 3.14: Turnover of *cis*-4-fluoro-L-proline (*cis*-4-F-L-Pro) by *trans*P4H: *trans*-4-hydroxylation of *cis*-4-F-L-Pro produces a (4*S*)-4-fluoro-4-hydroxy-L-proline [(4*S*)-4-F-4-OH-L-Pro] 'tetrahedral' intermediate, which fragments to produce 4-oxo-L-proline (4-Oxo-L-Pro) and fluoride.

When an LC/MS assay had been developed, the proline hydroxylase reactions were carried out again using a *cis*-4-F-L-Pro substrate; these samples were then anal-

ysed by LC/MS to confirm conclusions previously made about proline hydroxylase substrate selectivity.

LC/MS chromatograms measured for *cis*P3H Type I reactions contained peaks showing the presence of 150.1 m/z ($[M+H]^+$) ions, which correspond to a (3*S*,4*R*)-4-fluoro-3-hydroxy-L-proline [(3*S*,4*R*)-4-F-3-OH-L-Pro] product (Figures 3.33 a); the LC/MS results showed that catalytic hydroxylation of *cis*-4-F-L-Pro by *cis*P3H Type I had a yield of ~35%. The enzymatic production of this product by *cis*P3H Type I (using a *cis*-4-F-L-Pro substrate) is not surprising given that proline *cis*-3-hydroxylation is the ‘natural’ stereoselectivity for *cis*P3H Type I.^{195,201} LC/MS-analysis showed that *trans*P4H reactions contained a different product with 130.1 m/z $[M+H]^+$ ions (Figure 3.33 b); this peak corresponds to a 4-oxo-L-proline (4-Oxo-L-Pro) product, the product previously indicated by ‘fluoride release’ peaks in proton-decoupled ¹⁹F-spectra (Figure 3.32 b). Unlike the *cis*-hydroxylating PH enzyme, *cis*P3H Type I, the turnover of *cis*-4-F-L-Pro by *trans*P4H was comparatively poor, only giving a yield of ~6% turnover.

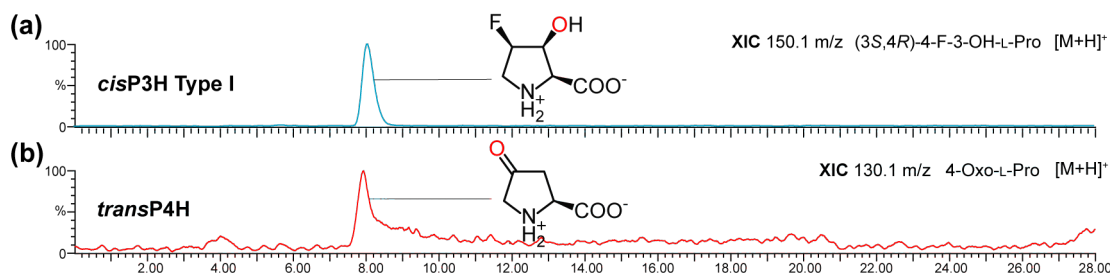


Figure 3.33: LC/MS-based analysis of proline hydroxylase reactions with *cis*-4-fluoro-L-proline (*cis*-4-F-L-Pro) [(a) *cis*P3H Type I and (b) *trans*P4H].

Extracted-ion chromatograms (XICs) contain peaks thought to correspond to (a) (3*S*,4*R*)-4-fluoro-3-hydroxy-L-proline [(3*S*,4*R*)-4-F-3-OH-L-Pro] (150.1 m/z $[M+H]^+$ ions) and (b) 4-oxo-L-proline (4-Oxo-L-Pro) (130.1 m/z $[M+H]^+$ ions), respectively.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 10 mM 2-OG, 1 mM *cis*-4-F-L-Pro. 100 μ L reactions were carried out in at room temperature (25 °C) for 4 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μ m) [4.6 x 250mm].

The *cis*-4-F-L-Pro *cis*P3H Type I reactions was scaled up 2×10 -fold (2×1 mL

reactions), and the (3*S*,4*R*)-4-F-3-OH-L-Pro product was purified using an LC/MS method for further NMR characterisation experiments. If we compare the *cis*-4-F-L-Pro *cis*P3H Type I reaction ¹H-spectrum (Figure 3.34 a) with the ¹H-spectrum of the purified (3*S*,4*R*)-4-F-3-OH-L-Pro (Figure 3.34 c), it is clear that the ‘new’ peaks produced by the reaction belong to those of the (3*S*,4*R*)-4-F-3-OH-L-Pro product. Selective ge-1D TOCSY experiments show clearly which peaks belong the (3*S*,4*R*)-4-F-3-OH-L-Pro product molecule in both reaction mixtures (Figure 3.34 b) and purified products (Figure 3.34 d), and support the NMR assignment.

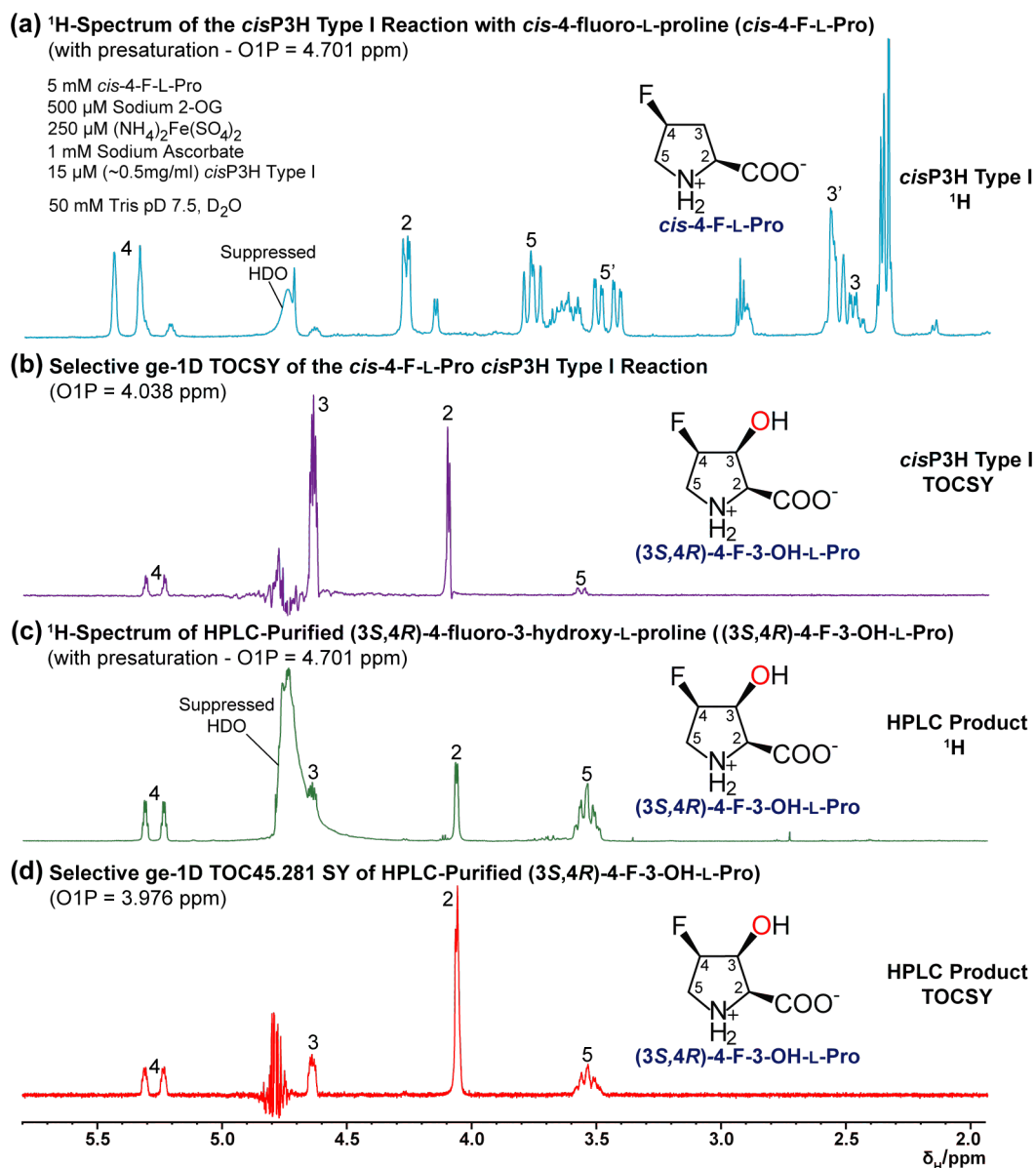


Figure 3.34: ^1H - and TOCSY-NMR analyses of *cis*-4-F-L-Pro *cis*P3H Type I reaction products in both (a) and (b) complete reaction mixtures and in (c) and (d) LC/MS-purified forms.

Spectrum (a) was measured on a 500 MHz Bruker AVII 500 NMR spectrometer [equipped with a 5 mm TFI- $^1\text{H}/^{19}\text{F}$ (^{13}C) probe]; spectra (b), (c), and (d) were measured on a 700 MHz Bruker AV III 700 spectrometer (equipped with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe).

Further NMR characterisation experiments were also carried out to confirm the identity of the (3*S*,4*R*)-4-F-3-OH-L-Pro product. For example, Figure 3.35 (below) is a COSY spectrum which supports the structure of the (3*S*,4*R*)-4-F-3-OH-L-Pro product.

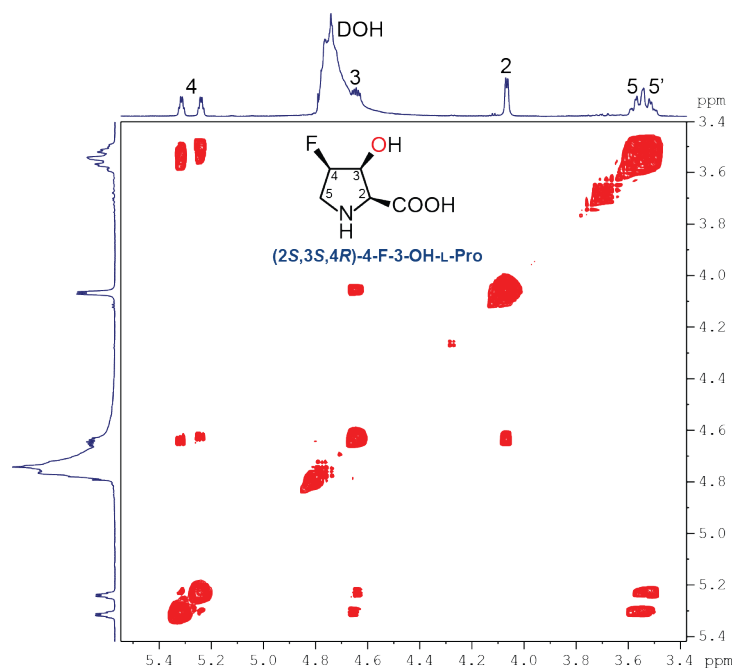
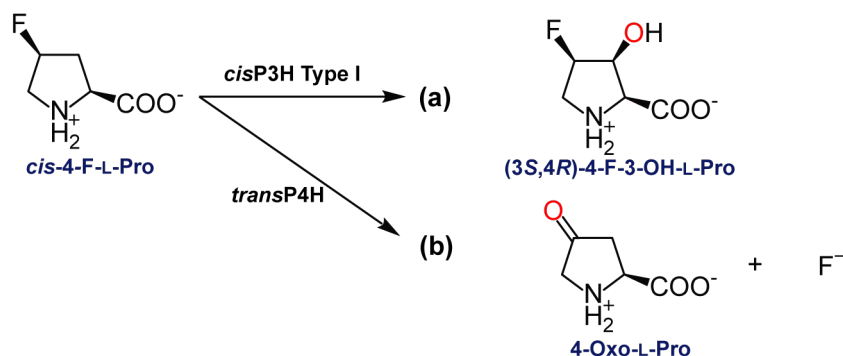


Figure 3.35: COSY-NMR analysis of the product from the *cis*P3H Type I/*cis*P4H reactions with *cis*-4-fluoro-L-proline (*cis*-4-F-L-Pro): NMR assignments confirmed that the product is (3*S*,4*R*)-4-fluoro-3-hydroxy-L-proline [(3*S*,4*R*)-4-F-3-OH-L-Pro]. Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a ¹H/¹³C/¹⁵N TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μ L) D₂O containing 0.05% (w/v) TSP-*d*₄, NMR chemical shift reference/internal standard.

In summary, the following reactions using a *cis*-4-F-Pro substrate have been found to be catalysed by the proline hydroxylase enzymes (Schemes 3.18 a and b).



Scheme 3.15: Proposed reactions catalysed by the proline hydroxylases using a *cis*-4-fluoro-L-proline (*cis*-4-F-L-Pro) substrate [(a) *cis*P3H Type I and (b) *trans*P4H]: (a) *cis*P3H Type I catalyses the production of (3*S*,4*R*)-4-fluoro-3-hydroxy-L-proline [(3*S*,4*R*)-4-F-3-OH-L-Pro], and (b) *trans*P4H produces 4-oxo-L-proline (4-Oxo-L-Pro).

Proton NMR analyses of proline hydroxylase reactions using a *trans*-4-F-L-Pro substrate showed that *cis*P3H Type I and *cis*P4H consumed the starting material and produced at least one product (Figures 3.36 a and b); both reactions (Figures 3.36 a and b) contained new peaks that were not present in the negative control ('no enzyme') reaction (Figure 3.36 c). For example, there are prominent double-multiplet peaks centered around ~5.2 ppm and ~4.7 ppm in both *cis*P3H Type I and *cis*P4H reaction spectra, suggesting both enzymes produce the same, *cis*-3-hydroxylation product, thought to be (3*S*,4*S*)-4-fluoro-3-hydroxy-L-proline [(3*S*,4*S*)-4-F-3-OH-L-Pro]. Based upon chemical shifts and peak multiplicity, the ~5.2 ppm double-multiplet peak was thought to be produced by a C4-proton [with a large geminal coupling (²*J*) to a C4-¹⁹F atom], and the ~4.7 ppm peak was thought to be from a proton attached to the hydroxylated carbon (C3). Given that the new peaks are much larger in the *cis*P3H Type I reactions (Figure 3.36 a) than the *cis*P4H reactions (Figure 3.30 b), it appears that *cis*-4-F-L-Pro turnover is greater for *cis*P3H

Type I than *cis*P4H; moreover, the depletion of *cis*-4-F-L-Pro starting material peaks is also much greater for *cis*P3H Type I than *cis*P4H (most of the *trans*-4-F-L-Pro was consumed), and 2-OG turnover to produce succinate (~2.4 ppm) is also greater for *cis*P3H Type I than *cis*P4H (Figures 3.36 a and b). These initial results suggested that *cis*-4-F-L-Pro may be a better substrate of *cis*P3H Type I than *cis*P4H.

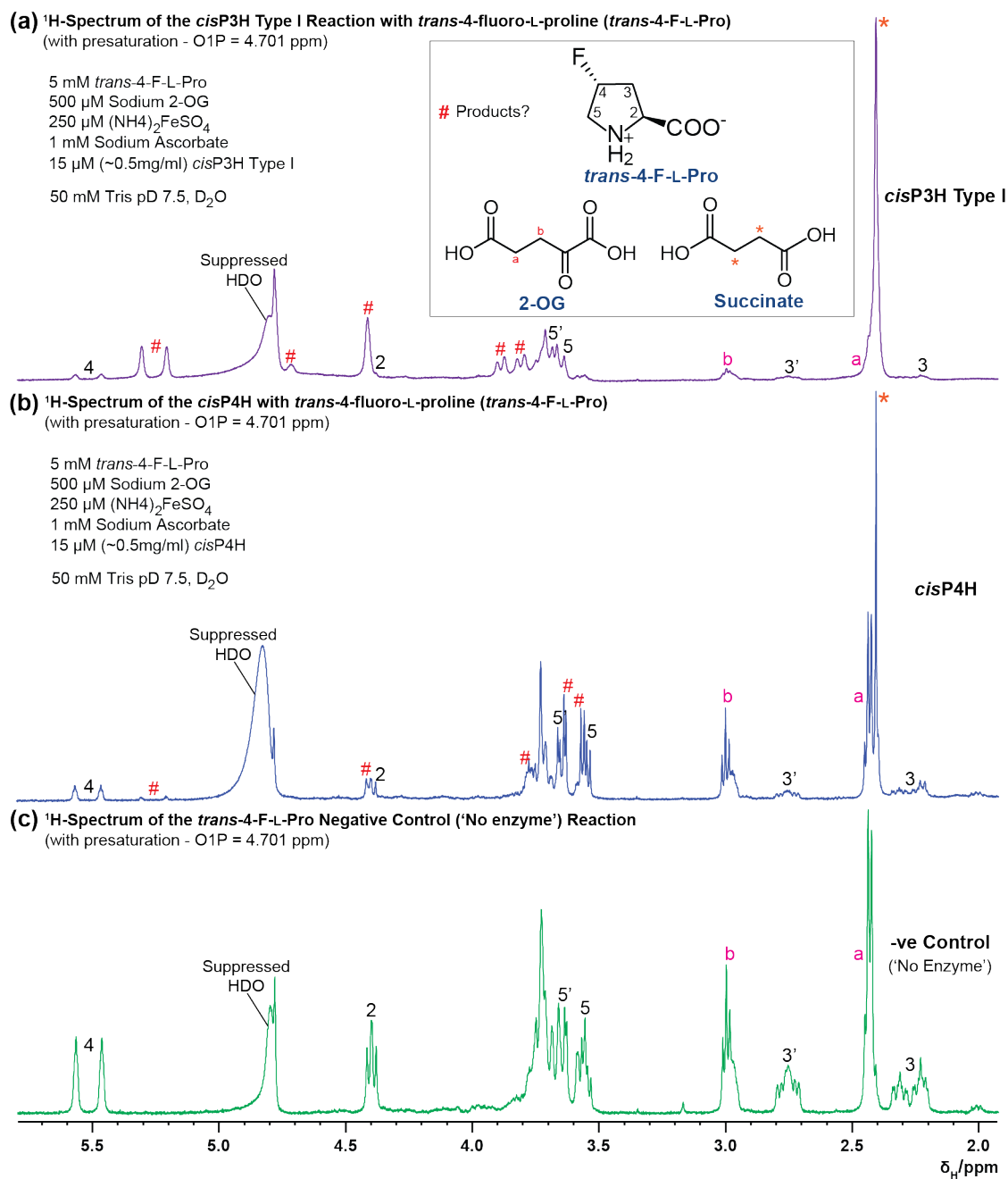
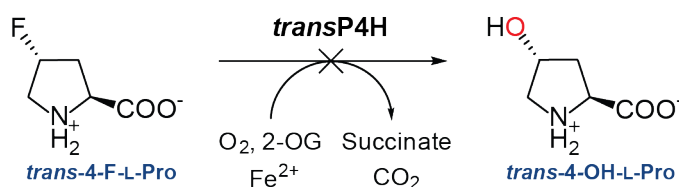


Figure 3.36: ^1H -NMR analyses of proline hydroxylase reactions using a *trans*-4-fluoro-L-proline (*trans*-4-F-L-Pro) substrate [(a) *cis*P3H Type I and (b) *cis*P4H].

Reactions were carried at room temperature (25 °C) for 14 h. Samples were analysed at 298 K using a 500 MHz Bruker AVII 500 NMR spectrometer [equipped with a 5 mm TFI- $^1\text{H}/^{19}\text{F}$ (^{13}C) probe]; presaturation (with O1P = 4.701 ppm) was used to suppress broad, HDO-solvent peaks. Chemical shifts were then referenced to TSP- d_4 (trimethylsilyl propanoic acid - $\delta_{\text{H, TSP-}d_4}$ = '0.0 ppm').

Proton-NMR analyses of *trans*P4H reactions using a *trans*-4-F-L-Pro substrate did not contain any additional resonance peaks compared with the negative control ('no enzyme') reaction, suggesting *trans*-4-F-L-Pro is not a substrate of *trans*P4H; this contradicts preliminary GC data, which Lawrence *et al.* had collected in the 1990s,¹⁹¹ which suggested that the *trans*-4-fluorine atom of *trans*-4-F-L-Pro may be replaced by a *trans*-4-hydroxyl group in reactions catalysed by *S. griseoviridis* P8648 *trans*P4H.¹⁹¹ However, the oxidation of fluorine is highly energetically unfavourable (owing to F's high electronegativity), and therefore Lawrence's observations¹⁹¹ may be an artifact of the GC method used, and not a 'true' indication of hydroxylation of the *trans*-4-fluorine site of *trans*-4-F-L-Pro (Scheme 3.16).



Scheme 3.16: *trans*-4-fluoro-L-proline (*trans*-4-F-L-Pro) is not a substrate of *trans*P4H: *Dactylosporangium sp.* RH1 *trans*P4H does not catalyse direct *trans*-4-hydroxylation of (*trans*-4-F-L-Pro).

Fluorine-19 NMR analysis of the same samples analysed in Figure 3.36 showed that a fluorine-containing product was generated from *trans*-4-F-L-Pro by *cis*P3H Type I (Figure 3.37 a) and *cis*P4H (Figure 3.37 b); in addition to *trans*-4-F-L-Pro starting material peaks (~ 177.5 ppm), these spectra also contained additional peaks at ~ 183 ppm, consistent with a (3*S*,4*S*)-4-F-3-OH-L-Pro product being generated. The product peak at ~ 183 ppm in the *cis*P3H Type I reaction spectra (Figure 3.37 a) is much larger than the corresponding peak from the *cis*P4H spectra (Figure 3.37 b), suggesting that reaction yields of this product are higher for *cis*P3H Type I than *cis*P4H (which is consistent with the results from ^1H -spectra measured for these samples). Interestingly, proton-decoupled ^{19}F -NMR spectra for the *cis*P4H reaction with *trans*-4-F-L-Pro also showed a peak at ~ 122 ppm ('F⁻ peak'), suggesting that

trans-4-F-Pro is also converted to fluoride and 4-oxo-L-proline (4-Oxo-L-Pro) via an unstable ‘tetrahedral intermediate’, (4*R*)-4-fluoro-4-hydroxy-L-proline [(4*R*)-4-F-4-OH-L-Pro] (Scheme 3.17), which spontaneously fragments to generate the reaction products; this reaction pathway is analogous to the *trans*-4-hydroxylations of *cis*-4-fluoro-proline/prolyl derivatives catalysed by *trans*P4H and PHD2.³⁷⁸

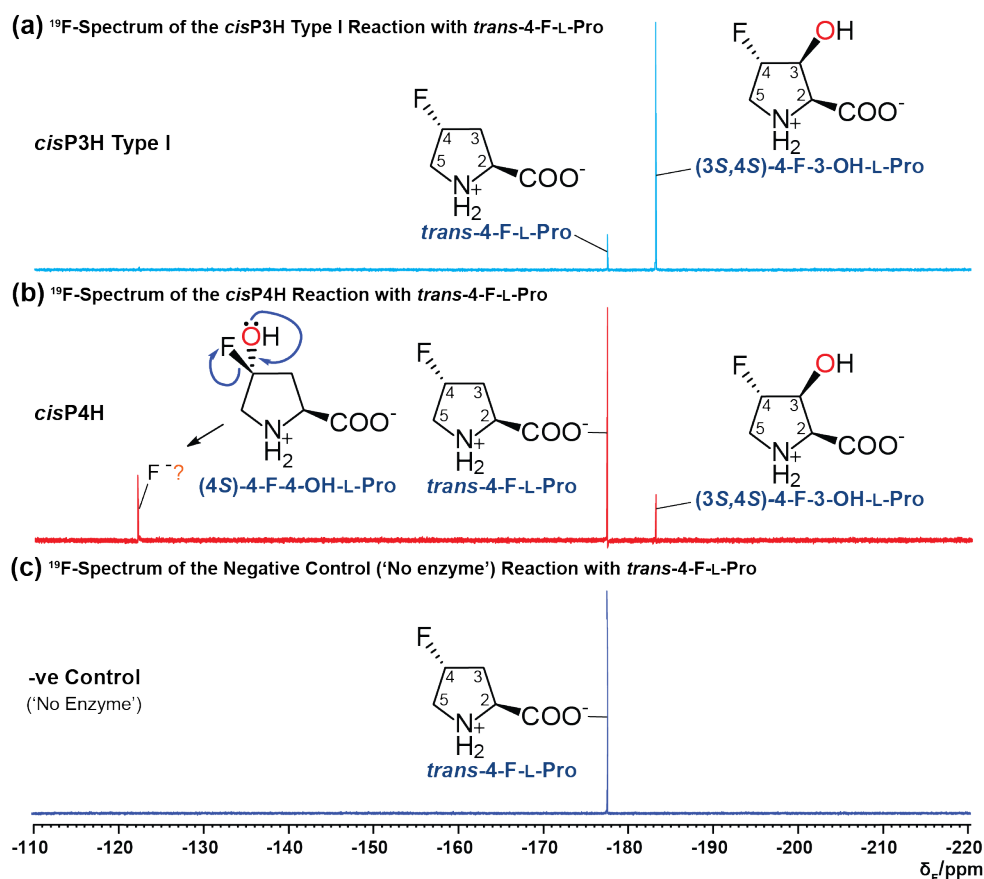
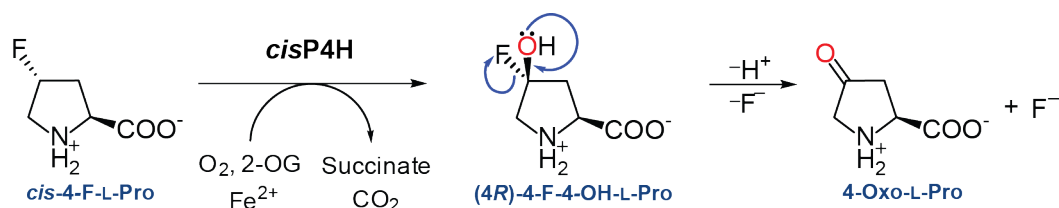


Figure 3.37: Proton-decoupled ¹⁹F-NMR analyses of proline hydroxylase reactions using a *trans*-4-fluoro-L-proline (*trans*-4-F-L-Pro) substrate [(a) *cis*P3H Type I, (b) *cis*P4H, and 'no enzyme' negative control]. Proton-decoupled ¹⁹F-spectra were measured at 298 K for samples in 5 mm tubes using a 470.4 MHz (¹⁹F) Bruker AVII 500 NMR spectrometer [equipped with a 5 mm TFI-¹H/¹⁹F(¹³C) probe]. Chemical shift values are referenced to those of CFCl₃ (where $\delta_F, \text{CFCl}_3 = '0.0 \text{ ppm}'$).



Scheme 3.17: Turnover of *trans*-4-fluoro-L-proline (*trans*-4-F-L-Pro) by *cisP4H*: *cis*-4-hydroxylation of *trans*-4-F-L-Pro produces a (4*R*)-4-fluoro-4-hydroxy-L-proline [(4*R*)-4-F-4-OH-L-Pro] 'tetrahedral' intermediate, which fragments to produce 4-oxo-L-proline (4-Oxo-L-Pro) and fluoride.

Consistent with the proton-NMR experiments, no reaction products were detected by ^{19}F -NMR in *transP4H* reactions which used a *trans*-4-F-L-Pro substrate, suggesting that *trans*-4-F-L-Pro is not a substrate of *transP4H* (Scheme 3.16).

LC/MS chromatograms measured for *cisP3H* Type I (Figures 3.33 a) and *cisP4H* (Figures 3.33 b) reactions contained an 150.1 m/z ($[M+H]^+$) ion peak, which corresponds to a (3*S*,4*S*)-4-fluoro-3-hydroxy-L-proline [(3*S*,4*S*)-4-F-3-OH-L-Pro] product. Consistent with previous NMR data, the LC/MS results showed that catalytic *cis*-3-hydroxylation of *trans*-4-F-L-Pro is greater for *cisP3H* Type I (yield = ~29%) than for *cisP4H* (yield = ~25.5%). The superior *cis*-3-hydroxylation activity of *cisP3H* Type I (over *cisP4H*) when using a *trans*-4-F-L-Pro substrate is not surprising given that proline *cis*-3-hydroxylation is the 'natural' stereoselectivity for *cisP3H* Type I.^{195,201} LC/MS-analysis also showed another peak that *cisP4H* reactions contained a different product with 130.1 m/z $[M+H]^+$ ions (Figure 3.38 c); this peak corresponds to a 4-oxo-L-proline (4-Oxo-L-Pro) product, the product previously indicated by 'fluoride release' peaks in proton-decoupled ^{19}F -spectra (Figure 3.37 b). For the shorter proline hydroxylase reactions (5 h compared with 14 h for the NMR assays), it appears that the 4-Oxo-L-Pro yields are much lower [~2.5%, total yield (both products) = ~28%], suggesting that the *cis*-3-hydroxylating reaction of *cisP4H* may be faster than the *cis*-4-hydroxylating pathway when a *trans*-4-F-L-Pro substrate is used; alternatively, there could be issues with stability of 4-Oxo-L-Pro under

the LC/MS conditions used, and some 4-Oxo-L-Pro product may have been undetected. 4-Oxo-L-Pro was also tested as a proline hydroxylase substrate to determine whether it could be being consumed also, but was not found to be substrate of the proline hydroxylase enzymes.

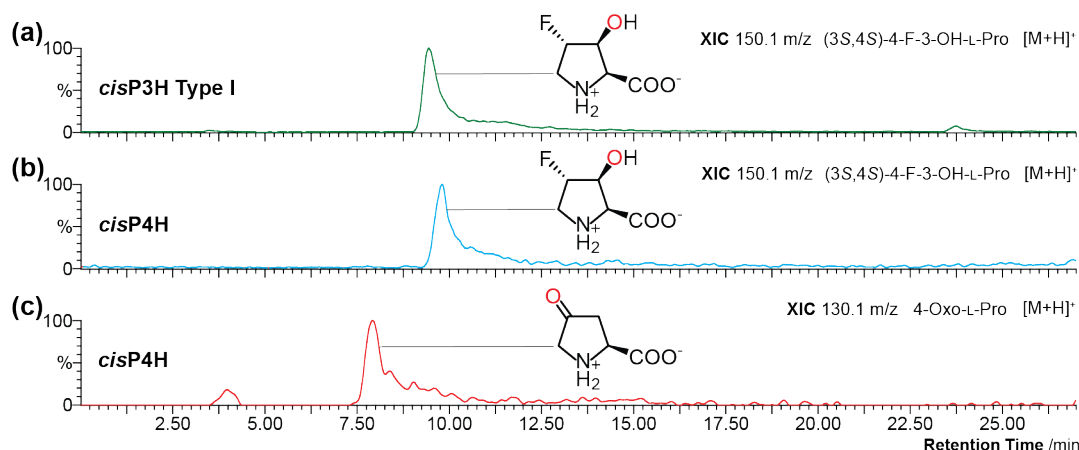


Figure 3.38: LC/MS-based analysis of proline hydroxylase reactions with *trans*-4-fluoro-L-proline (*trans*-4-F-L-Pro) [(a) *cis*P3H Type I, (b) and (c) *cis*P4H].

Extracted-ion chromatograms (XICs) contain peaks thought to correspond to (a) and (b) (3*S*,4*S*)-4-fluoro-3-hydroxy-L-proline [(3*S*,4*S*)-4-F-3-OH-L-Pro] (150.1 *m/z* [*M*+*H*]⁺ ions) and (c) 4-oxo-L-proline (4-Oxo-L-Pro) (130.1 *m/z* [*M*+*H*]⁺ ions), respectively.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM *trans*-4-F-L-Pro. 100 μ L reactions were carried out in at room temperature (25 °C) for 4 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μ m) [4.6 x 250mm].

The *trans*-4-F-L-Pro *cis*P3H Type I reactions was scaled up 2 $\times$ 10-fold (2 $\times$ 1 mL reactions), and the (3*S*,4*S*)-4-F-3-OH-L-Pro product was purified using an LC/MS method for further NMR characterisation experiments. NMR characterisation data is consistent with the product being (3*S*,4*R*)-4-F-3-OH-L-Pro. For example, Figure 3.39 (below) shows an assigned COSY spectrum, which supports the structure of the (3*S*,4*R*)-4-F-3-OH-L-Pro product.

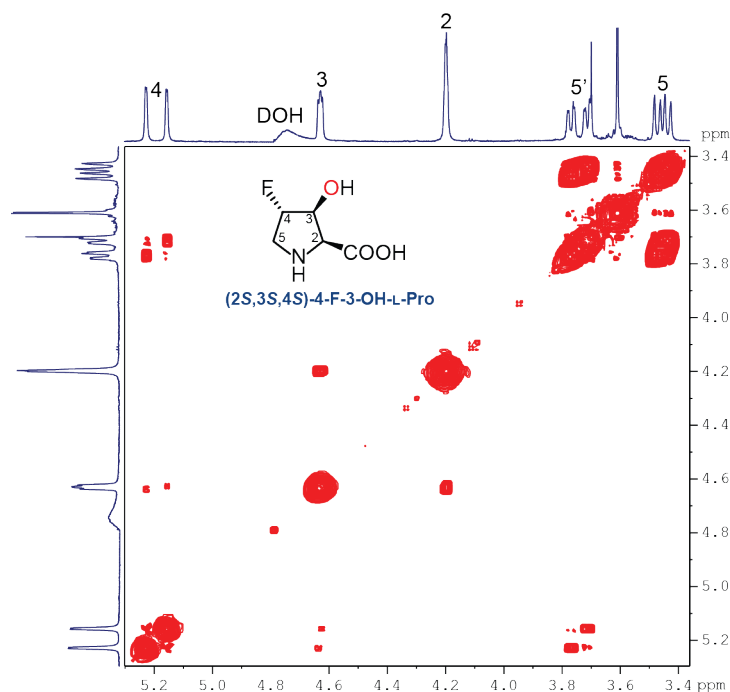
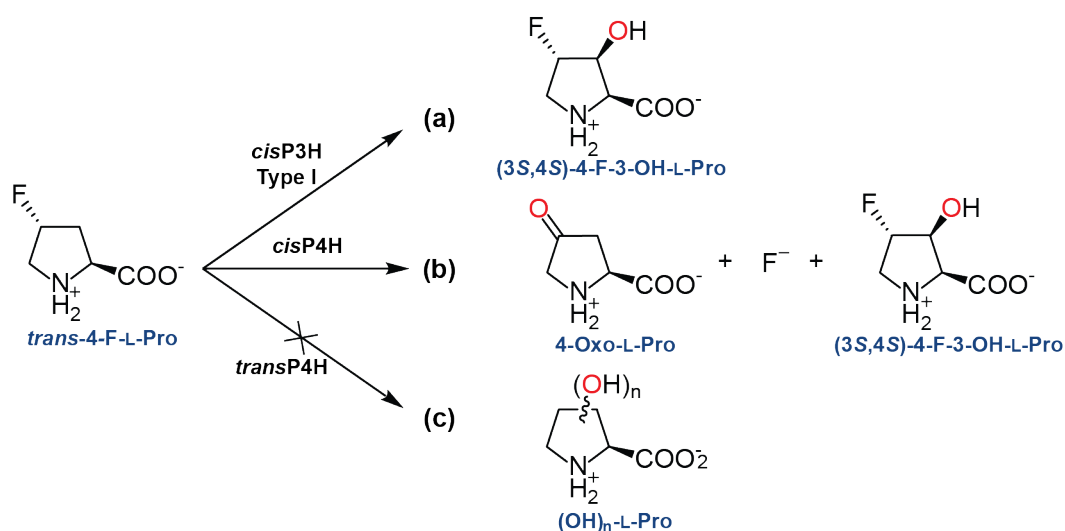


Figure 3.39: COSY-NMR analysis of the product from the *cis*P3H Type I/*cis*P4H reactions with *trans*-4-fluoro-L-proline (*trans*-4-F-L-Pro): NMR assignments confirmed that the product is (3*S*,4*S*)-4-fluoro-3-hydroxy-L-proline [(3*S*,4*S*)-4-F-3-OH-L-Pro]. Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D_2O containing 0.05% (w/v) TSP- d_4 , NMR chemical shift reference/internal standard.

In summary, the following reactions using a *trans*-4-F-Pro substrate have been found to be catalysed by the proline hydroxylase enzymes (Schemes 3.18 a-c).



Scheme 3.18: Proposed reactions catalysed by the proline hydroxylases using a *cis*-4-fluoro-L-proline (*cis*-4-F-L-Pro) substrate [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H]: (a) *cis*P3H Type I and (b) *cis*P4H both catalyse the production of (3*S*,4*S*)-4-fluoro-3-hydroxy-L-proline [(3*S*,4*S*)-4-F-3-OH-L-Pro], and (b) *cis*P4H also produces 4-oxo-L-proline (4-Oxo-L-Pro); *trans*-4-F-L-Pro is not a *trans*P4H substrate.

3.4.4.3.4 Substrate Turnover Summary: Halogenated Proline Derivatives

The following table (Table 3.9) summarises the proline hydroxylase activities with the 4-fluoroproline substrates screened for proline hydroxylase substrate activity in this section. See below for details of the reaction yields and products.

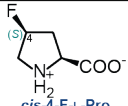
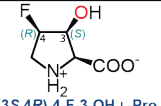
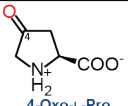
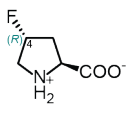
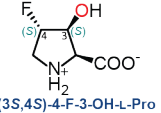
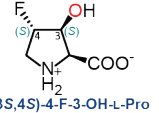
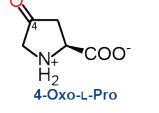
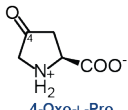
Substrate	Length/h	Product(s) [+ Yield(s)]					
		<i>cis</i> P3H Type I		<i>cis</i> P4H	<i>trans</i> P4H		
 <i>cis</i> -4-F-L-Pro	5	 (3 <i>S</i> ,4 <i>R</i>)-4-F-3-OH-L-Pro	35%	-	0%	 4-Oxo-L-Pro	6%
 <i>trans</i> -4-F-L-Pro	5	 (3 <i>S</i> ,4 <i>S</i>)-4-F-3-OH-L-Pro	29%	 (3 <i>S</i> ,4 <i>S</i>)-4-F-3-OH-L-Pro	25.5%	-	
				 4-Oxo-L-Pro	2.5%		
 4-Oxo-L-Pro	14	-	0%	-	0%	-	0%

Table 3.9: Proline hydroxylase substrate analogue screening: the effects of substrate halogenation on proline hydroxylase activity.

cis-4-Fluoro-L-proline (*cis*-4-F-L-Pro), *trans*-4-Fluoro-L-proline (*trans*-4-F-L-Pro), and 4-oxo-L-proline (4-Oxo-L-Pro) were screened for proline hydroxylase activity.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM Substrate.

3.4.4.3.5 Hydroxylated Proline Analogues

To test the effect of hydroxyl functionalities on proline hydroxylase activity, proline derivatives with hydroxyl groups at various ring positions (Figure 3.40) were tested for substrate activity.

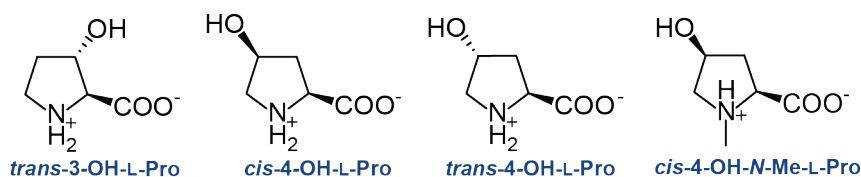
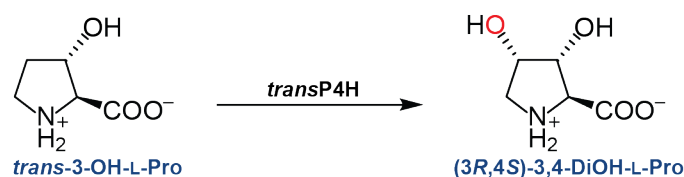


Figure 3.40: Hydroxylated proline derivatives tested for proline hydroxylase substrate activity

LC/MS-analyses of proline hydroxylase reactions (*cis*P3H Type I, *cis*P4H, and *trans*P4H) using a *trans*-3-hydroxy-L-proline (*trans*-3-OH-L-Pro) substrate showed that *trans*-3-OH-L-Pro is a substrate of *trans*P4H (Scheme 3.19); extracted-ion chromatograms (XICs) showing 148.1 m/z [M+H]⁺ ions contained a peak with a re-

tention time of ~12 min (Figure 3.41), thought to correspond with a (3*R*,4*S*)-3,4-dihydroxy-L-proline [(3*R*,4*S*)-3,4-DiOH-L-Pro] (*i.e.* the *trans*-4-hydroxylation product). Calculations using LC/MS peak integration showed that the yield of the *trans*P4H reaction with *trans*-3-OH-L-Pro was ~16%. However, no hydroxylation activity was detected in reactions using *cis*P3H Type I and *cis*P4H with a *trans*-3-OH-L-Pro substrate.



Scheme 3.19: *trans*P4H catalyses the *trans*-4-hydroxylation of *trans*-3-hydroxy-L-proline (*trans*-3-OH-L-Pro) to produce (3*R*,4*S*)-3,4-dihydroxy-L-proline [(3*R*,4*S*)-3,4-DiOH-L-Pro].

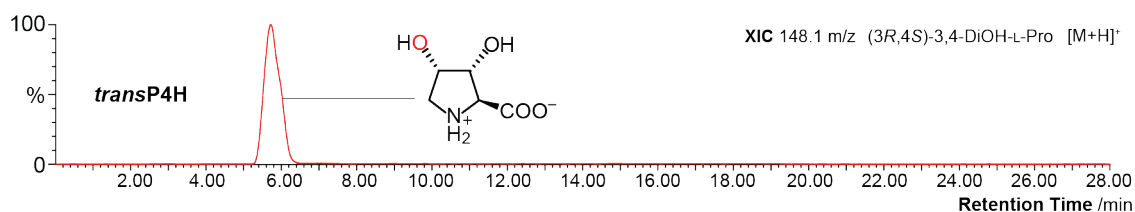


Figure 3.41: LC/MS-based analysis of proline hydroxylase (*trans*P4H) reactions with *trans*-3-hydroxy-L-proline (*trans*-3-OH-L-Pro).

Extracted-ion chromatograms (XICs) showing 148.1 m/z $[M+H]^+$ ions contained a single peak corresponding to (3*R*,4*S*)-3,4-dihydroxy-L-proline [(3*R*,4*S*)-3,4-DiOH-L-Pro].

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 10 mM 2-OG, 1 mM *trans*-3-OH-L-Pro. 100 μ L reactions were carried out in at room temperature (25 $^\circ$ C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 $\text{\AA}$, 5 μ m) [4.6 x 250mm].

The *trans*P4H reactions using a *trans*-3-OH-L-Pro substrate were scaled up 2×10 -fold (2×1 mL) for LC/MS-purification of the (3*R*,4*S*)-3,4-DiOH-L-Pro product; given that both the *trans*-3-OH-L-Pro starting material and the (3*R*,4*S*)-3,4-DiOH-L-Pro product are both very polar compounds with similar retention times (even in analytical experiments) (Figure 3.41), the product HPLC purification was challenging, and the HPLC method required some optimisation (e.g. mobile phase gradients, sample quantities injected per run, *etc.*). However, a pure sample was isolated, and a number of NMR-characterisation experiments were carried out to confirm the stereochemistry of the product. For example, Figure 3.42 (below) shows the assigned ^1H - and COSY spectra of the (3*R*,4*S*)-3,4-DiOH-L-Pro product.

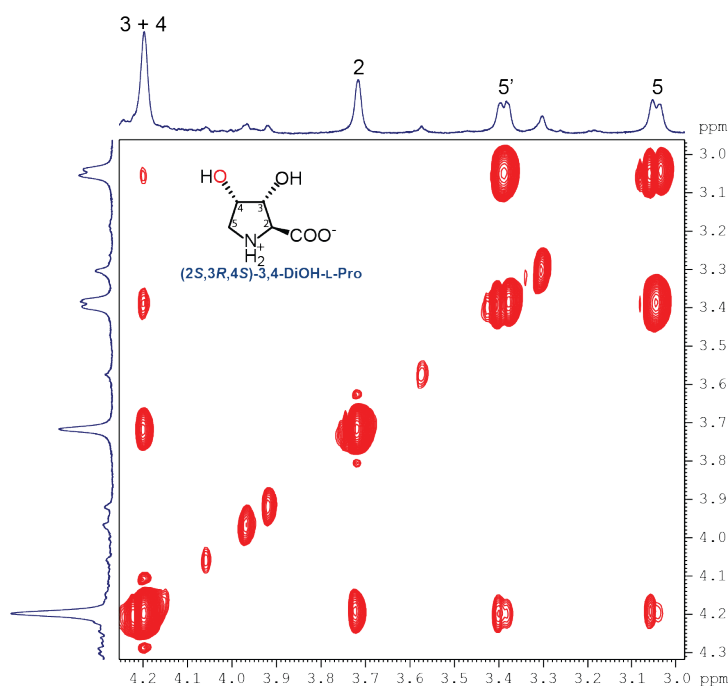
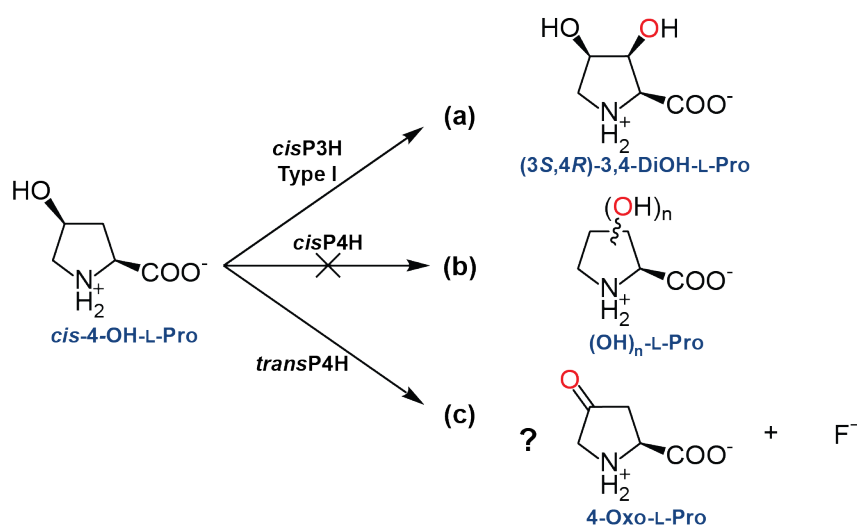


Figure 3.42: COSY-NMR analysis of a product from the *trans*P4H reactions with *trans*-3-hydroxy-L-proline (*trans*-3-OH-L-Pro): NMR assignments confirmed that the product is (3*R*,4*S*)-3,4-dihydroxy-L-proline [(3*R*,4*S*)-3,4-DiOH-L-Pro].

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D_2O containing 0.05% (w/v) TSP- d_4 , NMR chemical shift reference/internal standard.

Proline hydroxylase (*cis*P3H Type I, *cis*P4H, and *trans*P4H) reactions using a *cis*-4-hydroxy-L-proline (*cis*-4-OH-L-Pro) substrate showed that *cis*-4-OH-L-Pro is a substrate of *cis*P3H Type I (Scheme 3.20); LC/MS-analysis showed that the reactions contained a 148.1 m/z ($[M+H]^+$) product, consistent with the *cis*-3-hydroxylation of *cis*-4-OH-L-Pro to produce (3*S*,4*R*)-3,4-dihydroxy-L-proline [(3*S*,4*R*)-3,4-DiOH-L-Pro]. The (3*S*,4*R*)-3,4-DiOH-L-Pro product had a retention time of ~12 min (Figure 3.45), and peak integration indicated that the reaction yield was ~22%.



Scheme 3.20: *cis*P3H Type I catalyses the *cis*-4-hydroxylation of *cis*-4-hydroxy-L-proline (*cis*-4-OH-L-Pro) to produce (3*S*,4*R*)-3,4-dihydroxy-L-proline [(3*S*,4*R*)-3,4-DiOH-L-Pro].

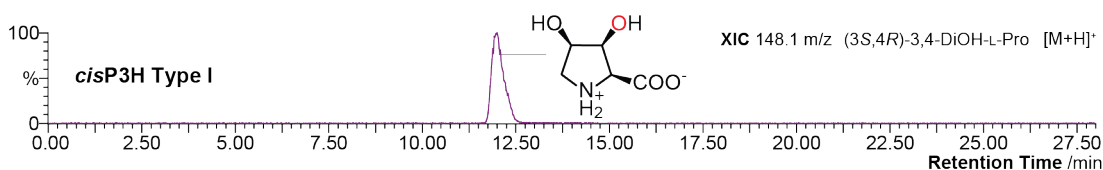


Figure 3.43: LC/MS-based analysis of proline hydroxylase (*cis*P3H Type I) reactions with *cis*-4-hydroxy-L-proline (*cis*-4-OH-L-Pro).

Extracted-ion chromatogram showing 148.1 m/z $[M+H]^+$ ions contained a single peak corresponding to (3*S*,4*R*)-3,4-dihydroxy-L-proline [(3*S*,4*R*)-3,4-DiOH-L-Pro].

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM *cis*-4-OH-L-Pro. 100 μ L reactions were carried out in at room temperature (25 °C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μ m) [4.6 x 250mm].

The *cis*P3H Type I reaction with *cis*-4-OH-L-Pro was then scaled up 2×10 -fold (2×1 mL) for LC/MS-purification of the (3*S*,4*R*)-3,4-DiOH-L-Pro product; a similar HPLC method was used to purify (3*S*,4*R*)-3,4-DiOH-L-Pro as was used for (3*S*,4*R*)-3,4-DiOH-L-Pro (from the *trans*P4H reaction with *trans*-3-OH-L-Pro - Scheme 3.19), since both dihydroxyprolines have similar retention properties [*n.b.* on a larger scale, (3*S*,4*R*)-3,4-DiOH-L-Pro is retained by the HPLC column slightly better than (3*R*,4*S*)-3,4-DiOH-L-Pro though, making its purification slightly easier]. Various NMR characterisation experiments were carried out in attempt to confirm the stereochemistry of the (3*S*,4*R*)-3,4-DiOH-L-Pro product; for example, Figure 3.44 (below) shows the assigned ^1H - and COSY spectra of the (3*S*,4*R*)-3,4-DiOH-L-Pro product.

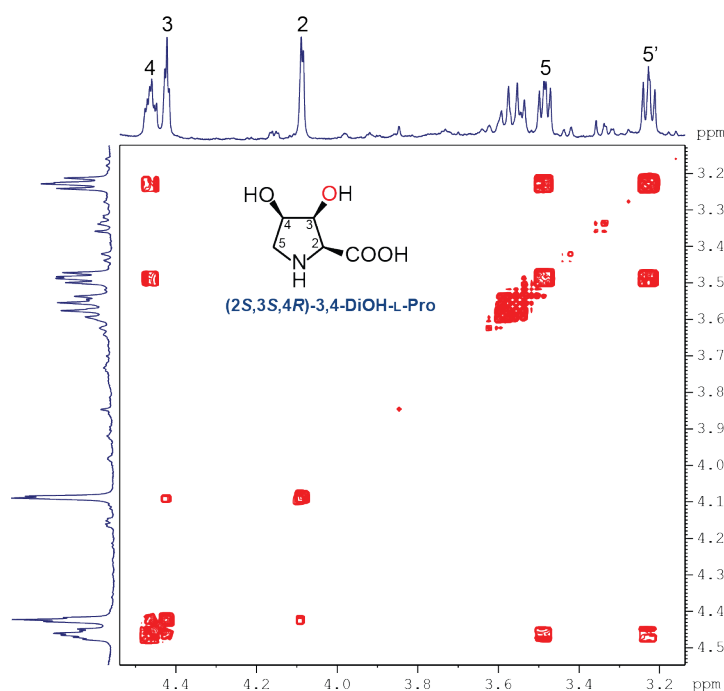
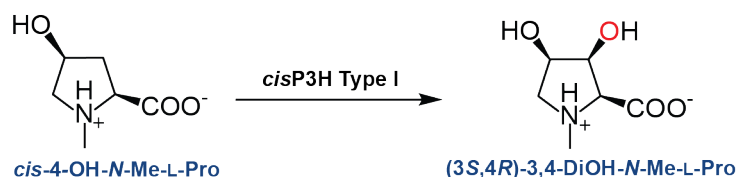


Figure 3.44: COSY-NMR analysis of a product from the *cis*P3H Type I reactions with *cis*-4-hydroxy-L-proline (*cis*-4-OH-L-Pro): NMR assignments confirmed that the product is (3*S*,4*R*)-3,4-dihydroxy-L-proline [(3*S*,4*R*)-3,4-DiOH-L-Pro].

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D_2O containing 0.05% (w/v) TSP- d_4 , NMR chemical shift reference/internal standard.

The *cis*-4-hydroxy-*N*-methyl-L-proline (*cis*-4-OH-*N*-Me-L-Pro) substrate was also tested as a substrate of the proline hydroxylases (*cis*P3H Type I, *cis*P4H, and *trans*P4H). Like in the case of *cis*-4-hydroxy-L-proline (*cis*-4-OH-L-Pro), LC/MS-analysis showed *cis*-4-OH-*N*-Me-L-Pro is a substrate of *cis*P3H Type I (Scheme 3.21), and no products were detected with *cis*P4H and *trans*P4H; extracted-ion chromatograms (XICs) showing 162.1 *m/z* $[M+H]^+$ ions contained a single peak with a retention time of ~8 min {presumably corresponding to (3*S*,4*R*)-3,4-dihydroxy-*N*-methyl-L-proline [(3*S*,4*R*)-3,4-DiOH-*N*-Me-L-Pro]}. Peak integration values suggested the reaction yield was ~1%, roughly twenty times less than the corresponding reaction with *cis*-4-OH-L-Pro; the effect of a combination of two different modifications appears to have a combined, negative effect on proline hydroxylase turnover. The (3*S*,4*R*)-3,4-DiOH-*N*-Me-L-Pro product has been purified, but, as of writing this thesis, no NMR characterisation has been measured to confirm the product stereochemistry.



Scheme 3.21: *cis*P3H Type I catalyses the *cis*-4-hydroxylation of *cis*-4-hydroxy-*N*-methyl-L-proline (*cis*-4-OH-*N*-Me-L-Pro) to produce (3*S*,4*R*)-3,4-dihydroxy-*N*-methyl-L-proline [(3*S*,4*R*)-3,4-DiOH-*N*-Me-L-Pro].

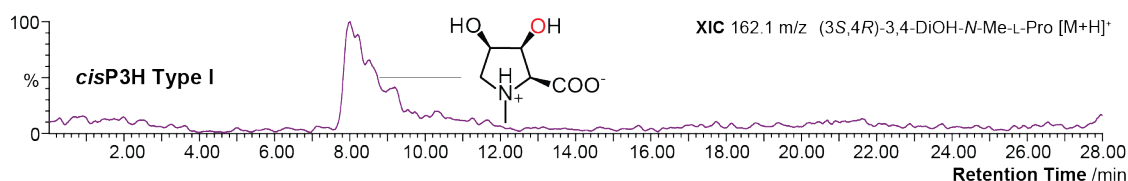


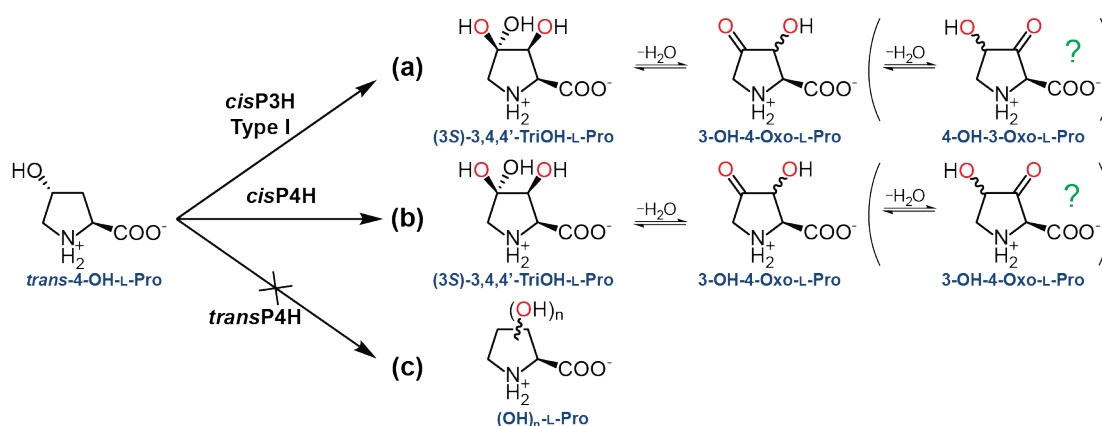
Figure 3.45: LC/MS-based analysis of proline hydroxylase (*cis*P3H Type I) reactions with *cis*-4-hydroxy-*N*-methyl-L-proline (*cis*-4-OH-*N*-Me-L-Pro).

Extracted-ion chromatogram showing 162.1 m/z $[M+H]^+$ ions contained a single peak corresponding to (3*S*,4*R*)-3,4-dihydroxy-*N*-methyl-L-proline [(3*S*,4*R*)-3,4-DiOH-*N*-Me-L-Pro].

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 10 mM 2-OG, 1 mM *cis*-4-OH-*N*-Me-L-Pro. 100 μ L reactions were carried out in at room temperature (25 °C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μ m) [4.6 x 250mm].

When the proline hydroxylase reactions using a *trans*-4-hydroxy-L-proline were analysed, 2 products were observed for the *cis*P3H Type I and *cis*P4H reactions (Schemes 3.22 a and b); conversely, no products were observed in the corresponding *trans*P4H reactions, suggesting that *trans*-4-OH-L-Pro is not a *trans*P4H substrate. LC/MS analysis of the *cis*P3H Type I (Figure 3.46 a) and *cis*P4H (Figure 3.46 b) reactions each showed single peaks corresponding to 162.1 m/z $[M+H]^+$ ions; these peaks were thought to be produced by (3*S*,4*S*)-3,4,4'-trihydroxy-L-proline [(3*S*)-3,4,4'-TriOH-L-Pro], the *cis*-3,4-dihydroxylation product. Peak integration suggested that the yields of this product were ~7% and ~15% for *cis*P3H Type I and *cis*P4H, respectively. In addition, extracted-ion chromatograms (XICs) for analysed *cis*P3H Type I (Figure 3.46 a) and *cis*P4H (Figure 3.46 b) reactions both contained two 146.1 m/z ($[M+H]^+$) peaks, suggesting that the 'keto' form of (3*S*)-3,4,4'-TriOH-L-Pro was in equilibrium with the trihydroxyproline product. Peak integration suggested that the yields of the hydroxyketo products were ~3.5% and ~2.5% in the *cis*P3H Type I reaction, and ~7% and ~5% for the *cis*P4H reaction.



Scheme 3.22: Proposed reactions catalysed by the proline hydroxylases using a *trans*-4-hydroxy-L-proline (*trans*-4-OH-L-Pro) substrate [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H]:

(a) *cis*P3H Type I and (b) *cis*P4H *cis*-3,4-dihydroxylate *trans*-4-OH-L-Pro to produce (3*S*)-3,4,4'-trihydroxy-L-proline [(3*S*)-3,4,4'-TriOH-L-Pro], a product which is in equilibrium with 3-hydroxy-4-oxo-L-proline (3-OH-4-Oxo-L-Pro) (which has mixed stereochemistry at carbon-3, owing to enol-keto tautomerisation); there is also some chance that enol-keto tautomerisation could also lead to the formation of 4-hydroxy-3-oxo-L-proline (4-OH-3-Oxo-L-Pro);

(c) *trans*-4-hydroxy-L-proline (*trans*-4-OH-L-Pro) is not a substrate of *trans*P4H.

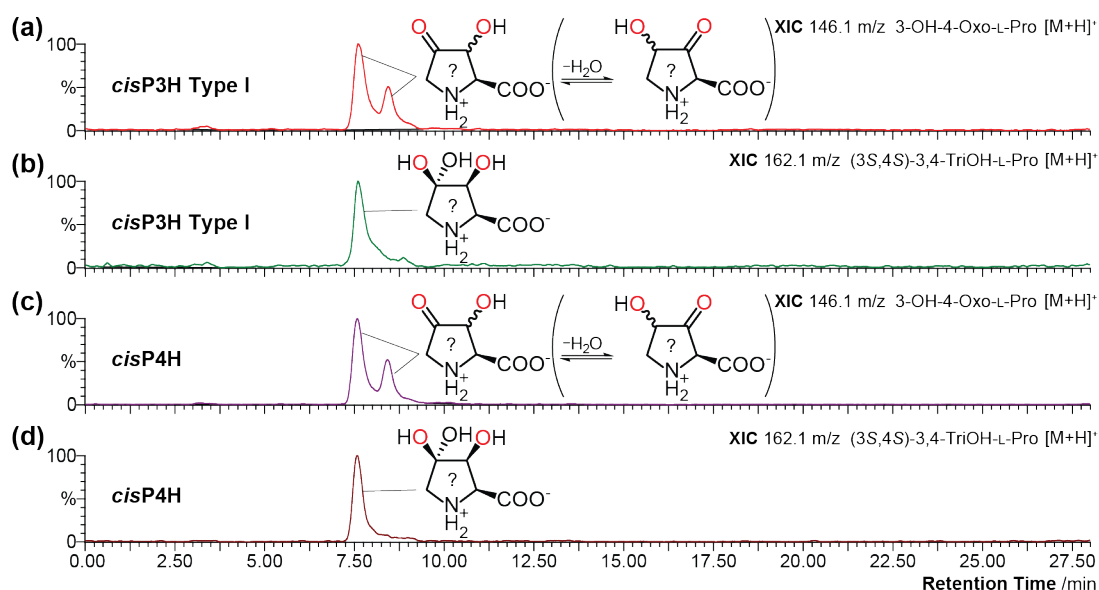


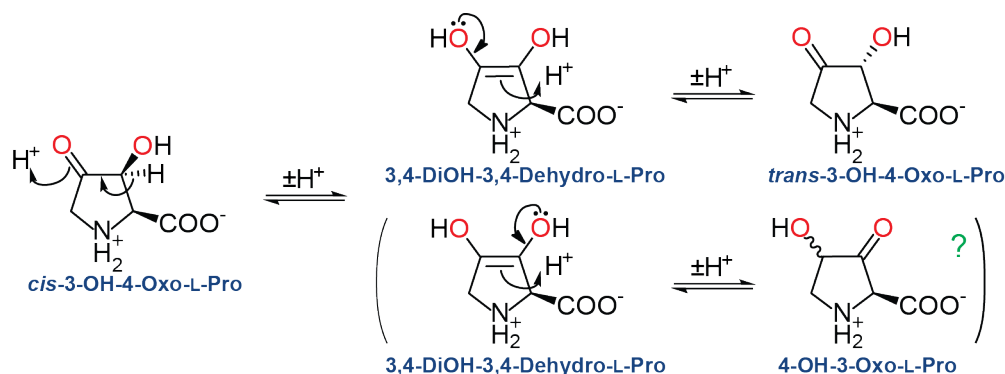
Figure 3.46: LC/MS-based analysis of proline hydroxylase reactions with *trans*-4-hydroxy-L-proline (*trans*-4-OH-L-Pro) [(a) *cis*P3H Type I and (b) *cis*P4H].

Extracted-ion chromatograms (XICs) (showing 162.1 m/z ions) contain peaks thought to correspond to (3*S*,4*S*)-3,4,4'-trihydroxy-L-proline [(3*S*)-3,4,4'-TriOH-L-Pro], and XICs showing 146.1 m/z ions contain two peaks thought to be isomers of 3-hydroxy-4-oxo-L-proline (3-OH-4-Oxo-L-Pro).

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM *trans*-4-OH-L-Pro. 100 μ L reactions were carried out in at room temperature (25 °C) for 5 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μ m) [4.6 x 250mm].

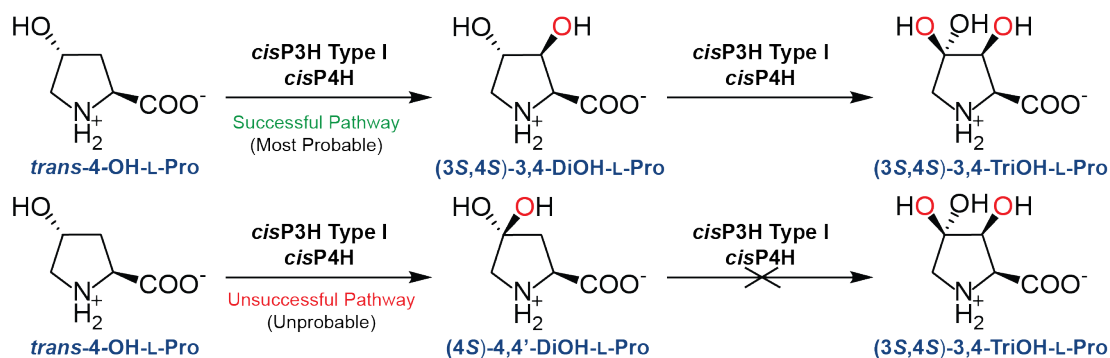
LC/MS-analysis indicated the presence of two (likely) hydroxy-keto-proline products in the *cis*P3H Type I and *cis*P4H reactions (Figure 3.46); it is possible that the enol-keto tautomerisation results in mixed stereochemistry of the 3-hydroxy group (Scheme 3.23 a). It is possible that some 3-hydroxy-4-oxo-L-proline (3-OH-4-Oxo-L-Pro) product could convert to 4-hydroxy-3-oxo-L-proline (4-OH-3-Oxo-L-Pro) via a dihydroxyenol intermediate, 3,4-dihydroxy-3,4-dehydro-L-proline (3,4-DiOH-3,4-Dehydro-L-Pro) (Scheme 3.23 a). Therefore, it is not entirely clear whether the two hydroxy-keto-proline peaks in the *cis*P3H Type I/*cis*P4H LC/MS chromatograms (Figure 3.46) are the two isomers of 3-OH-4-Oxo-L-Pro (*i.e.* *cis*-3-OH-4-Oxo-L-Pro and *trans*-3-OH-4-Oxo-L-Pro), or interconverting 3-OH-4-Oxo-L-Pro and 4-OH-3-Oxo-L-Pro.



Scheme 3.23: Potential product rearrangement pathways when *cis*P3H Type I and *cis*P4H use a *trans*-4-hydroxy-L-proline (*trans*-4-OH-L-Pro).

Enol-keto tautomerisation of *cis*-3-hydroxy-L-proline (*cis*-4-OH-4-Oxo-L-Pro) products via a dihydroxyketo intermediate, 3,4-dihydroxy-3,4-dehydro-L-proline (3,4-DiOH-3,4-Dehydro-L-Pro), could lead conversion to *trans*-3-OH-Oxo-L-Pro (and potentially racemisation); additionally, enol-keto tautomerisation could even lead to the formation of 4-hydroxy-3-oxo-L-proline (4-OH-3-Oxo-L-Pro).

Interestingly, in both *cis*P3H Type I and *cis*P4H reactions dihydroxylation products were observed; no monohydroxylation product was observed. Given that 4-oxo-L-proline (4-Oxo-L-Pro) (and its hydrate form) is not a substrate of the proline hydroxylases, in order for *cis*P3H Type I/*cis*P4H-catalysed dihydroxylation of *trans*-4-OH-L-Pro to occur, the order of hydroxylation must be *cis*-3-hydroxylation followed by *cis*-4-hydroxylation (Scheme 3.23 b) (by process of elimination); if the PH enzymes were to *cis*-4-hydroxylate first (before *cis*-3-hydroxylation), this would form 4-Oxo-L-Pro (via its hydrate), and further hydroxylation would not occur (*i.e.* the reaction would stop at 4-Oxo-L-Pro).



Scheme 3.24: Potential product formation pathways when *cis*P3H Type I and *cis*P4H use a *trans*-4-hydroxy-L-proline (*trans*-4-OH-L-Pro).

Dihydroxylation of *trans*-4-OH-L-Pro must involve *cis*-3-hydroxylation followed by *cis*-4-hydroxylation by process of elimination, since initial *cis*-4-hydroxylation forms (4*S*)-4,4'-dihydroxy-L-proline [(4*S*)-4,4'-DiOH-L-Pro], the hydrate form of 4-Oxo-L-Pro, which is not a proline hydroxylase substrate (*i.e.* no second hydroxylation can thus occur by this pathway).

3.4.4.3.6 Substrate Turnover Summary: Hydroxylated Proline Derivatives

The following table (Table 3.10) summarises the proline hydroxylase activities with the hydroxyproline substrates screened for proline hydroxylase substrate activity in this section. See below for details of the reaction yields and products.

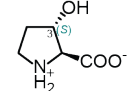
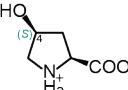
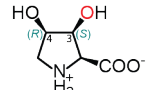
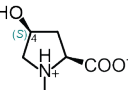
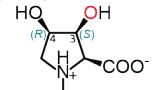
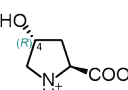
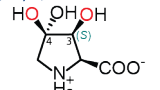
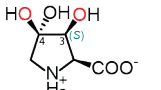
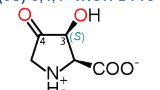
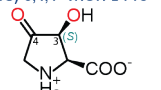
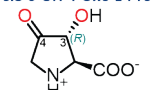
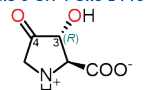
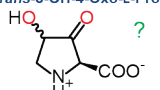
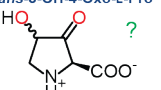
Substrate	Length/h	Product(s) [+ Yield(s)]			
		<i>cis</i> P3H Type I	<i>cis</i> P4H	<i>trans</i> P4H	
 <i>trans</i> -3-OH-L-Pro	14	-	0%	-	0%
 <i>cis</i> -4-OH-L-Pro	14	 (3 <i>S</i> ,4 <i>R</i>)-3,4-DiOH-L-Pro	22%	-	0%
 <i>cis</i> -4-OH-N-Me-L-Pro	14	 (3 <i>S</i> ,4 <i>R</i>)-3,4-DiOH-N-Me-L-Pro	1%	-	0%
 <i>trans</i> -4-OH-L-Pro	5	 (3 <i>S</i>)-3,4,4'-TriOH-L-Pro	7%	 (3 <i>S</i>)-3,4,4'-TriOH-L-Pro	15%
		 <i>cis</i> -3-OH-4-Oxo-L-Pro	3.5%	 <i>cis</i> -3-OH-4-Oxo-L-Pro	7%
		 <i>trans</i> -3-OH-4-Oxo-L-Pro	3.5%	 <i>trans</i> -3-OH-4-Oxo-L-Pro	2.7%
		 4-OH-3-Oxo-L-Pro	?	 4-OH-3-Oxo-L-Pro	?

Table 3.10: Proline hydroxylase substrate analogue screening: the effects of substrate hydroxyl functionality on proline hydroxylase activity.

trans-3-hydroxy-L-proline (*trans*-3-OH-L-Pro), *cis*-4-hydroxy-L-proline (*cis*-4-OH-L-Pro), *cis*-4-hydroxy-N-methyl-L-proline (*cis*-4-OH-N-Me-L-Pro), and *trans*-4-hydroxy-L-proline (*trans*-4-OH-L-Pro) were screened for proline hydroxylase activity.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM Substrate.

3.4.4.3.7 Carboxymethylproline (CMP) Derivatives

The carboxymethylprolines (CMPs) are an interesting class of proline derivatives (Figure 3.47) which are involved in the biosynthesis of carbapenem antibiotics (Scheme 3.25);^{29,362,363} they originate from L- Δ^1 -pyrroline-5-carboxylic acid (L-P5C),^{29,362,363} which, itself, originates from L-proline.^{29,362,379} These CMP compounds were tested as substrates of the 3 proline hydroxylase enzymes (*cis*P3H Type I, *cis*P4H, and *trans*P4H); these compounds were kindly provided for testing by Dr Refaat Hamed, and reactions were carried out in collaboration with Dr Hamed.

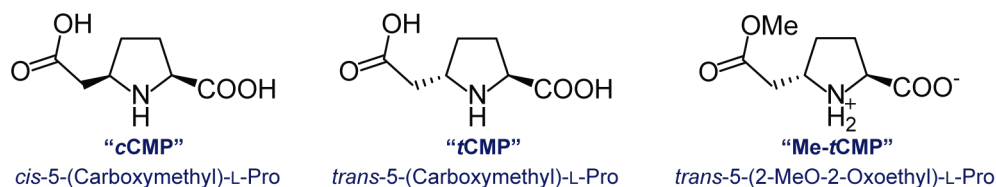
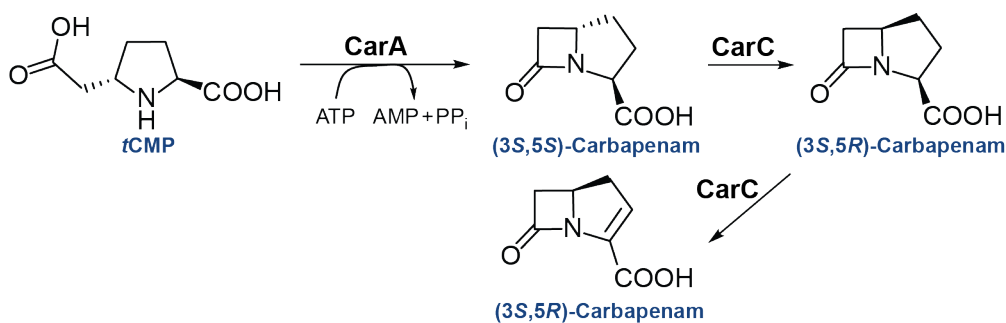


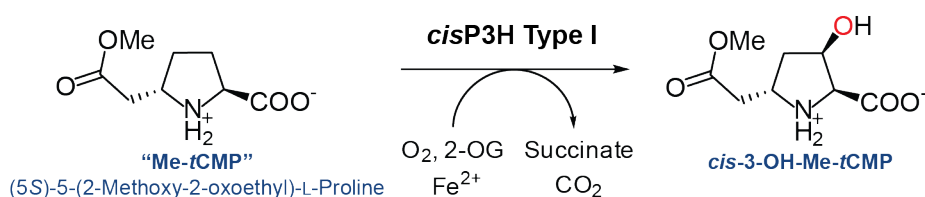
Figure 3.47: Carboxymethylproline (CMPs) derivatives tested as proline hydroxylase substrates.



Scheme 3.25: Carboxymethylprolines (CMPs) are involved in carbapenem antibiotic biosynthesis.

Interestingly, although both *cis*- and *trans*-5-(carboxymethyl)-L-proline (cCMP and tCMP) were not substrates of the proline hydroxylase enzymes (*cis*P3H Type I, *cis*P4H, and *trans*P4H), the side-chain methyl ester of tCMP, *trans*-5-(2-methoxy-2-oxoethyl)-L-proline ("Me-tCMP"), was found to be a substrate of *cis*P3H Type I (Scheme 3.26).

LC/MS-analysis of proline hydroxylase reactions using a Me-*t*CMP substrate showed that there was a hydroxylation product (204.1 m.z [M+H]⁺ ions) in the *cis*P3H Type I reaction using a Me-*t*CMP substrate (Figure 3.51); peak integration showed that turnover was good with this substrate, reactions occurring with a yield of ~65%.



Scheme 3.26: *trans*-5-(2-methoxy-2-oxoethyl)-L-proline ("Me-*t*CMP") is a substrate of *cis*P3H Type I, and is *cis*-3-hydroxylated to produce (3*R*,5*S*)-3-hydroxy-5-(2-methoxy-2-oxoethyl)-L-proline ["(3*R*)-3-OH-Me-*t*CMP"].

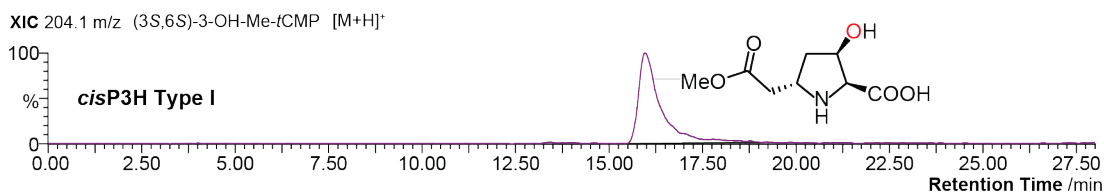


Figure 3.48: LC/MS-based analysis of proline hydroxylase (cisP3H Type I) reactions with *trans*-5-(2-methoxy-2-oxoethyl)-L-proline ("Me-*t*CMP").

Extracted-ion chromatogram showing 204.1 m/z [M+H]⁺ ions contained a single peak corresponding to (3R,5S)-3-hydroxy-5-(2-methoxy-2-oxoethyl)-L-proline ["(3R)-3-OH-Me-tCMP"].

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM Me-*t*CMP. 100 μ L reactions were carried out in at room temperature (25 °C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 µm) [4.6 x 250mm].

The methylation of the carboxymethyl sidechain (by forming the methyl ester) removes the net negative charge of the carboxyl oxygen; this may stop the sidechain carboxyl group from ‘interfering’ with the recognition of the α -carboxyl group of *t*CMP’s the L-proline centre, and therefore ‘activate’ the *t*CMP as a *cis*P3H Type I substrate.

The *cis*P3H Type I reaction with Me-*t*CMP was scaled up and the (3*R*)-3-OH-Me-*t*CMP product was HPLC-purified by Dr Refaat Hamed. The (3*R*)-3-OH-Me-*t*CMP product was characterised using NMR experiments; Figure 3.49 shows COSY and NOESY spectra for the product of Me-*t*CMP hydroxylation by *cis*P3H Type I, (3*R*,5*S*)-3-hydroxy-5-(2-methoxy-2-oxoethyl)-L-proline ("*cis*-3-OH-Me-*t*CMP") (as well as proton and ge-1D-selective TOCSY spectra).

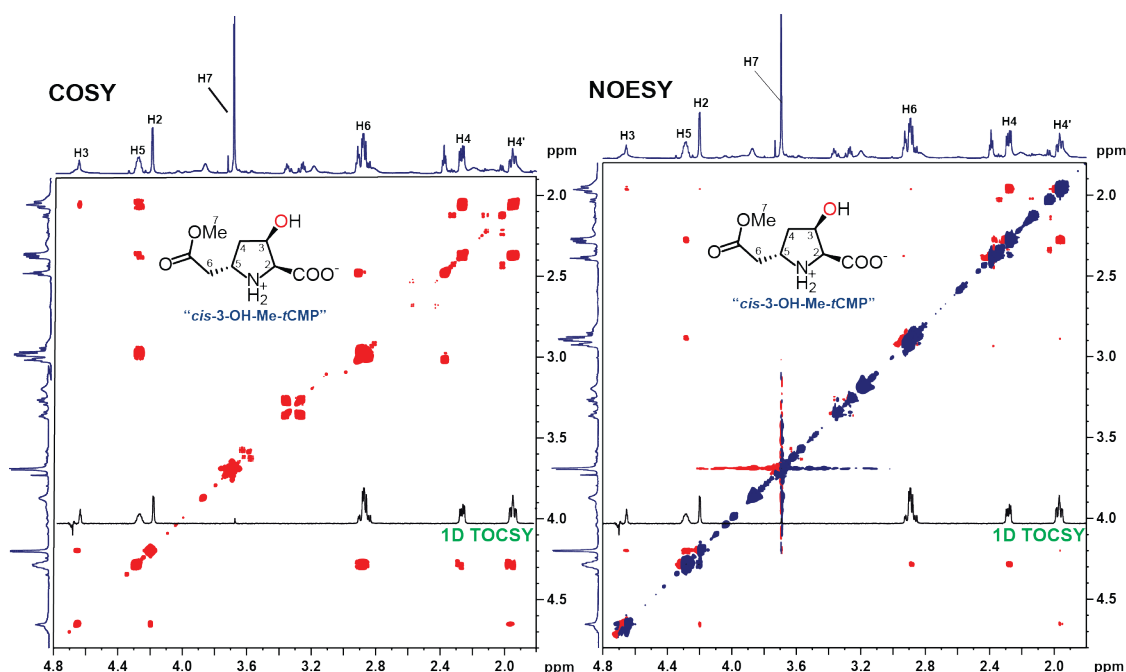


Figure 3.49: COSY and NOESY spectra of the HPLC-purified product of *cis*P3H-catalysed hydroxylation of *trans*-5-(2-methoxy-2-oxoethyl)-L-proline ("Me-*t*CMP") (as well as its ^1H -, ge-1D-selective TOCSY-spectra); the product was unambiguously assigned and found to be the *cis*-3-hydroxylation product, (3*R*,5*S*)-3-hydroxy-5-(2-methoxy-2-oxoethyl)-L-proline ("*cis*-3-OH-Me-*t*CMP").

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D_2O containing 0.05% (w/v) TSP- d_4 , NMR chemical shift reference/internal standard. This particular spectrum was measured by Dr Refaat Hamed.

Another *t*CMP derivative, with an isopenten-2-yl group attached to carbon-6 of *t*CMP (*i.e.* on the methylene carbon of the carboxymethyl group), was converted to a carboxymethyl ester (Scheme 3.27), and tested as a *cis*P3H Type I substrate (Figure 3.50); if this ester compound is a proline hydroxylase substrate, it may suggest that selectively ‘protecting’ the sidechain carboxylic acid group (by converting it to a methyl ester) is a general method for ‘activating’ *t*CMP-like compounds to become *cis*P3H Type I substrates. Moreover, this compound has a relatively large isopenten-2-yl group attached to the carboxylmethyl sidechain (Figure 3.50), and any conversion of this compound by *cis*P3H Type I may suggest that the enzyme can accommodate relatively bulky substrates; if *cis*P3H Type I were to accept a wide range of substrates, of various shapes and sizes, the enzyme may be a useful biosynthetic tool. The biosynthesis of (6*S*)-6-(Isopenten-2-yl)-Me-*t*CMP initially involves converting the (6*S*)-6-(isopenten-2-yl)-*t*CMP into the corresponding pantetheinyl thioester using the enzyme malonyl-CoA synthetase (MatB) (from *Streptomyces coelicolor*) (Scheme 3.27);^{360,380} quenching of the MatB reaction with methanol then produces the (6*S*)-6-(isopenten-2-yl)-Me-*t*CMP target methyl ester, when methanol replaces the pantetheine thiol by nucleophilic substitution (Scheme 3.27). This method for thioester-conversion of *t*CMPs was previously optimised by Hamed *et al.*³⁶⁰

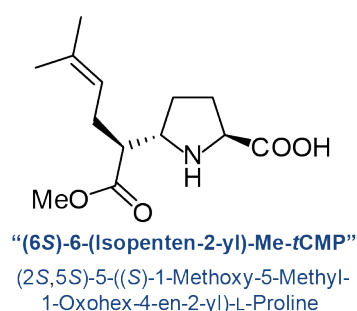
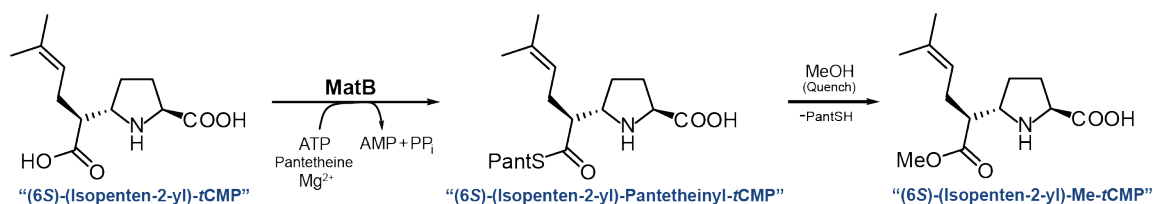
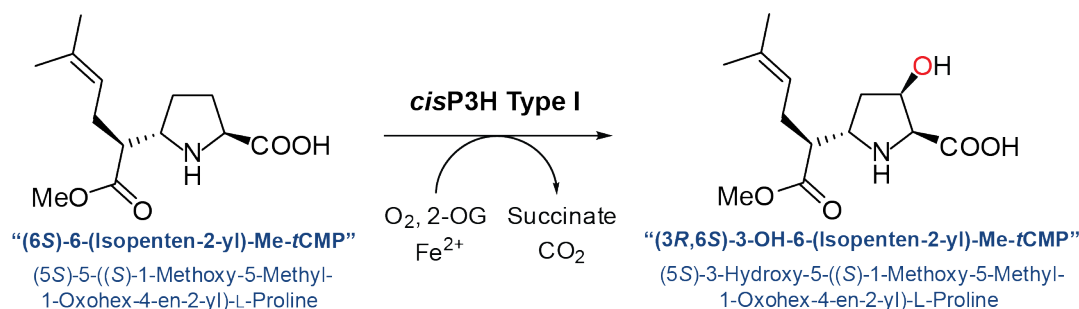


Figure 3.50: (2*S*,5*S*)-5-[(*S*)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline-["(6*S*)-6-(isopenten-2-yl)-Me-*t*CMP"] was tested as a proline hydroxylase substrate.



Scheme 3.27: The *in vitro* biosynthesis of (5S)-5-[(S)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline ["(6S)-6-(isopenten-2-yl)-Me-*t*CMP"]: malonyl CoA synthetase (MatB) was used to convert (6S)-6-(isopenten-2-yl)-*t*CMP into (6S)-6-(Isopenten-2-yl)-Pantetheinyl-*t*CMP (using pantetheine in replacement of CoA-SH), which is then spontaneously converted by nucleophilic substitution to the methyl ester, (6S)-6-(isopenten-2-yl)-Me-*t*CMP, when the MatB reaction is quenched with methanol.

Like the Me-*t*CMP, the (6S)-6-(isopenten-2-yl)-Me-*t*CMP substrate was a substrate of *cis*P3H Type I, despite having a large (isopenten-2-yl) group on the carboxymethyl sidechain, and the reactions using a slightly lower substrate concentration (250 μ M) than normal used for substrate testing (since the starting material was isolated from enzymatic biosynthesis reactions and material was in fairly scarce supply); it was thought that this reaction occurred in a fashion analogous to that seen for the Me-*t*CMP substrate, where the substrate undergoes *cis*-3-hydroxylation (Scheme 3.28). LC/MS-analysis of proline hydroxylase reactions using a (6S)-6-(isopenten-2-yl)-Me-*t*CMP substrate showed that, along with the starting material (255.1 m/z $[M+H]^+$ ions), there was a small amount of hydroxylation product (272.1 m/z $[M+H]^+$ ions) (Figure 3.51); like with Me-*t*CMP, the product was thought to be the *cis*-3-hydroxylation product, (3R,5S)-3-hydroxy-5-[(S)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline ["(3R,6S)-3-OH-6-(isopenten-2-yl)-Me-*t*CMP"]. However, hydroxylation product yields were much poorer using a (6S)-6-(isopenten-2-yl)-Me-*t*CMP substrate than a less-bulky Me-*t*CMP substrate, and yields were only ~3.5%.



Scheme 3.28: (5S)-5-[(S)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline [“(6S)-6-(isopenten-2-yl)-Me-*t*CMP”] is a substrate of *cis*P3H Type I, and is *cis*-3-hydroxylated to produce (3R,5S)-3-hydroxy-5-[(S)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline [“(3R,6S)-3-OH-6-(isopenten-2-yl)-Me-*t*CMP”].

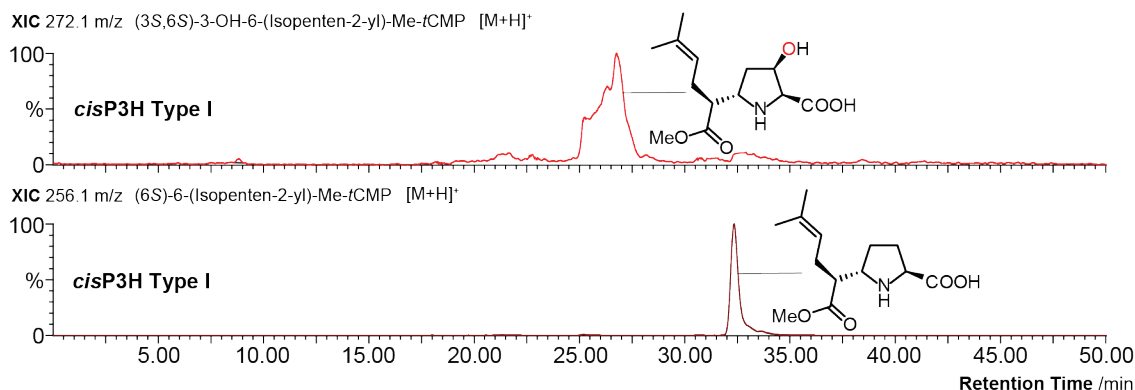


Figure 3.51: LC/MS-based analysis of proline hydroxylase (*cis*P3H Type I) reactions with (5S)-5-[(S)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline [“(6S)-6-(isopenten-2-yl)-Me-*t*CMP”]: extracted-ion chromatogram showing 272.1 m/z [M+H]⁺ ions contained a single peak corresponding to (3R,5S)-3-hydroxy-5-[(S)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline [“(3R,6S)-3-OH-6-(isopenten-2-yl)-Me-*t*CMP”].

PH reactions used 15 μM *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μM (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM (6S)-6-(isopenten-2-yl)-Me-*t*CMP. 100 μL reactions were carried out in at room temperature (25 °C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μm) [4.6 x 250mm].

3.4.4.3.8 Substrate Turnover Summary: Carboxymethylprolines (CMPs)

The following table (Table 3.11) summarises the proline hydroxylase activities with the carboxymethylproline substrates screened for proline hydroxylase substrate activity in this section.

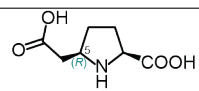
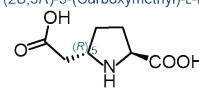
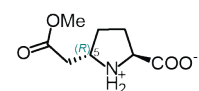
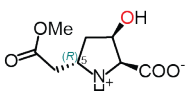
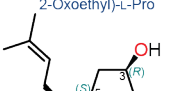
Substrate	Length/h	Product(s) [+ Yield(s)]			
		<i>cis</i> P3H Type I	<i>cis</i> P4H	<i>trans</i> P4H	
 (2 <i>S</i> ,5 <i>R</i>)-5-(Carboxymethyl)-L-Pro "cCMP"	14	-	0%	-	0%
 (2 <i>S</i> ,5 <i>R</i>)-5-(Carboxymethyl)-L-Pro "tCMP"	14	-	0%	-	0%
 <i>trans</i> -5-(2-Methoxy-2-Oxoethyl)-L-Pro "Me-tCMP"	14	 (3 <i>R</i> ,5 <i>S</i>)-3-Hydroxy-5-(2-Methoxy-2-Oxoethyl)-L-Pro "cis-3-OH-Me-tCMP"	63%	-	0%
 (6 <i>S</i>)-6-(Isopenten-2-yl)-Me-tCMP (2 <i>S</i> ,5 <i>S</i>)-5-[(<i>S</i>)-1-Methoxy-5-Methyl-1-Oxohex-4-en-2-yl]-L-Proline	14	 (3 <i>R</i> ,6 <i>S</i>)-3-OH-6-(Isopenten-2-yl)-Me-tCMP (3 <i>R</i> ,5 <i>S</i>)-3-Hydroxy-5-[(<i>S</i>)-1-Methoxy-5-Methyl-1-Oxohex-4-en-2-yl]-L-Proline	3.5%	-	0%

Table 3.11: Proline hydroxylase substrate analogue screening: proline hydroxylase activity with carboxymethylproline (CMP) substrates.

cis-5-(Carboxymethyl)-L-proline (*c*CMP), *trans*-5-(Carboxymethyl)-L-proline (*t*CMP), *trans*-5-(2-methoxy-2-oxoethyl)-L-proline (Me-*t*CMP), and (2*S*,5*S*)-5-[(*S*)-1-methoxy-5-methyl-1-oxohex-4-en-2-yl]-L-proline [(6*S*)-6-(isopenten-2-yl)-Me-*t*CMP] were screened for proline hydroxylase activity. PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM Substrate [and 250 μ M (6*S*)-6-(isopenten-2-yl)-Me-*t*CMP].

3.4.4.3.9 α,α' -Dicarboxy Proline Derivatives

To investigate the effect of additional substrate carboxylic acid groups on proline hydroxylase turnover, dicarboxy proline analogues (Figures 3.52) were tested as substrates of the proline hydroxylases (*cis*P3H Type I, *cis*P4H, and *trans*P4H); the compounds included (5*S*)-5-carboxy-L-proline [(2*R*,5*S*)-5-(COOH)-L-Pro], (5*S*)-5-carboxy-*N*-methyl-L-proline [(2*R*,5*S*)-5-(COOH)-*N*-Me-L-Pro], and (6*S*)-6-carboxy-L-pipecolic acid [(2*R*,6*S*)-6-(COOH)-L-Pip].

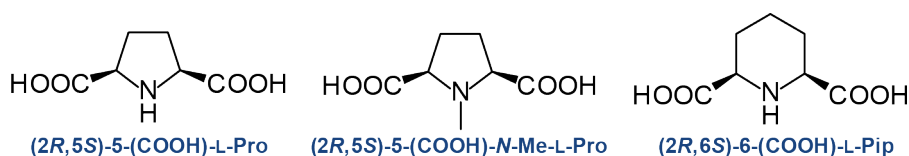


Figure 3.52: α,α' -Dicarboxy proline derivatives were tested as proline hydroxylases; compounds included (5*S*)-5-carboxy-L-proline [(2*R*,5*S*)-5-(COOH)-L-Pro], (5*S*)-5-carboxy-*N*-methyl-L-proline [(2*R*,5*S*)-5-(COOH)-*N*-Me-L-Pro], and (6*S*)-6-carboxy-L-pipecolic acid [(2*R*,6*S*)-6-(COOH)-L-Pip].

Much like *c*CMP and *t*CMP, none of the dicarboxy prolines were substrates of the proline hydroxylase enzymes; presumably the additional carboxyl group interferes with proline α -carboxy substrate recognition by the proline hydroxylases.

3.4.4.4 Proline analogues of Varying Ring Size

The aim of the work detailed in this section was to explore the selectivity of the PHs towards substrate ring size. Proline analogues of varying ring sizes (3-7 carbons) were tested as proline hydroxylase (*cis*P3H Type I, *cis*P4H, *trans*P4H substrates (Figure 3.53) to investigate the effect of substrate ring size on proline hydroxylase activity; compounds included: L-aziridine-2-carboxylic acid (L-Azy) (3-carbons), L-azetidine-2-carboxylic acid (L-Aze) (4-carbons), L-pipecolic acid (L-Pip) (6-carbons), *N*-methyl-L-pipecolic acid (*N*-Me-L-Pip) (6-carbons), and L-azepane-2-carboxylic acid (L-Azp) (7-carbons).

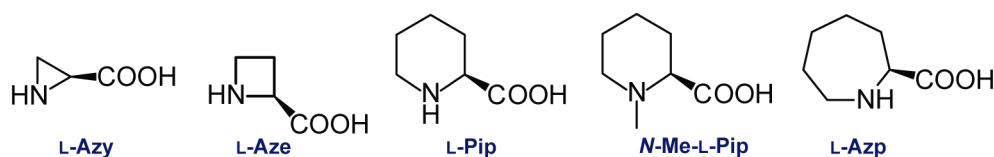
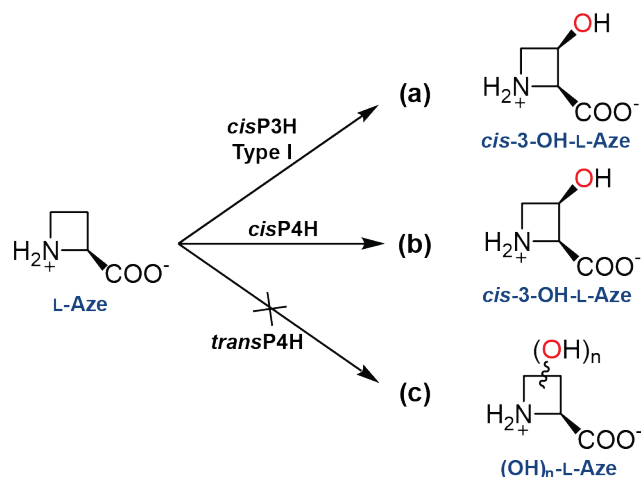


Figure 3.53: Proline analogues with different ring sizes were tested as proline hydroxylase (*cis*P3H Type I, *cis*P4H, and *trans*P4H) substrates. Compounds included: L-aziridine-2-carboxylic acid (L-Azy) (3-carbons), L-azetidine-2-carboxylic acid (L-Aze) (4-carbons), L-pipecolic acid (L-Pip) (6-carbons), *N*-methyl-L-pipecolic acid (*N*-Me-L-Pip) (6-carbons), and L-azepane-2-carboxylic acid (L-Azp) (7-carbons)

Analysis of proline hydroxylase (*cis*P3H Type I, *cis*P4H, and *trans*P4H) reactions

The smallest compound, L-aziridine-2-carboxylic acid (L-Azy), showed no indication of products; it is possible that the PH enzymes do not show activity with a L-Azy substrate, since the three-membered ring system of L-Azy is not oriented productively with respect to the proline hydroxylase ferryl [Fe(IV)=O] species for hydroxylation to occur. It is also plausible that substrate recognition (binding) of L-Azy is too poor for catalytic turnover, since it lacks vital features of proline's substrate pharmacophore.

L-Azetidine-2-carboxylic acid (L-Aze) has previously been reported as a substrate of *cis*P3H Type I and *cis*P3H Type II,^{195,201} as well as being tested as a substrate of *trans*P4H (for which it was reported as a non-substrate).^{191,195} In this study, L-Aze was tested as a substrate of *cis*P3H Type I, *cis*P4H, and *trans*P4H; LC/MS-analysis of the reactions showed that *cis*P3H Type I and *cis*P4H converted L-Aze into a hydroxylation product (Scheme 3.29) with a retention time of ~10.5 min (Figures 3.54 a and b); extracted-ion chromatograms (XICs) showing 118.1 m/z [M+H]⁺ ions contained single peaks thought to correspond to *cis*-3-hydroxy-L-azetidine-2-carboxylic acid (*cis*-3-OH-L-Aze) (as reported previously for *cis*P3H Type I/II^{195,201}). Peak integration suggested that product yields were ~5% and ~4% for the *cis*P3H Type I and *cis*P4H reactions, respectively. During the course of my work, *cis*P4H enzymes have been shown to produce *cis*-3-OH-L-Aze,³⁸¹ suggesting that our hypothesis is correct.



Scheme 3.29: Proposed reaction of (a) *cis*P3H Type I and (b) *cis*P4H using a L-azetidine-2-carboxylic acid (L-Aze): both enzymes catalyse the *cis*-3-hydroxylation of L-Aze to produce *cis*-3-hydroxy-L-azetidine-2-carboxylic acid (*cis*-3-OH-L-Aze).

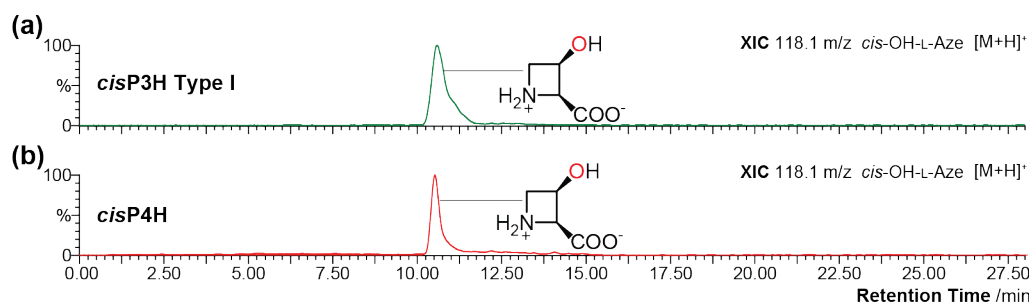


Figure 3.54: LC/MS-based analysis of proline hydroxylase [(a) *cis*P3H Type I and (b) *cis*P4H] reactions with L-azetidine-2-carboxylic acid (L-Aze).

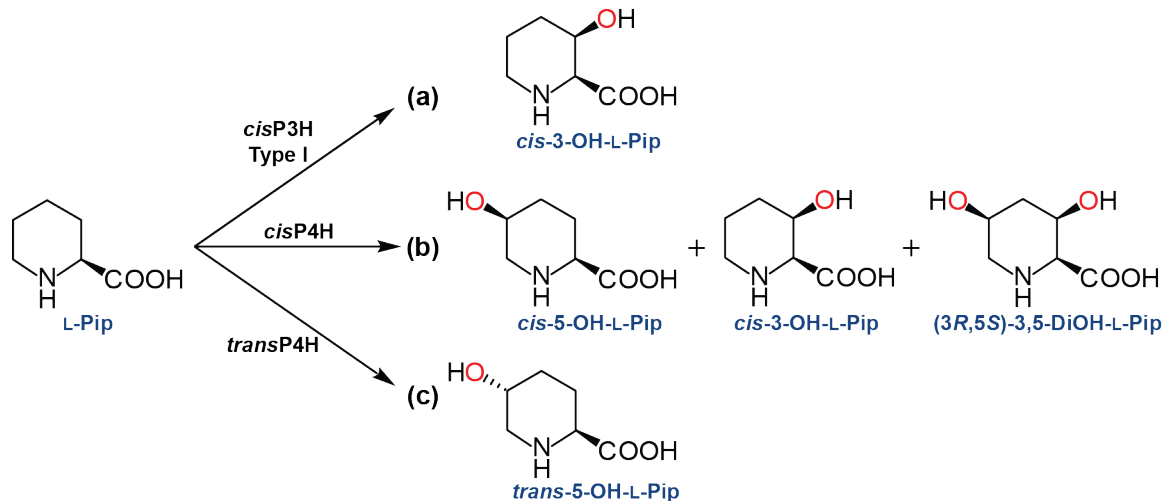
Extracted-ion chromatogram showing 118.1 m/z $[M+H]^+$ ions contained a single peak corresponding to *cis*-3-hydroxy-L-azetidine-2-carboxylic acid (*cis*-3-OH-L-Aze).

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M $(NH_4)_2Fe(SO_4)_2$, 10 mM 2-OG, 1 mM L-Aze. 100 μ L reactions were carried out in at room temperature (25 $^{\circ}$ C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 $\text{\AA}$, 5 μ m) [4.6 x 250mm].

L-Pipecolic acid (L-Pip), a 6-membered proline analogue, was also tested as a substrate of the proline hydroxylases (*cis*P3H Type I, *cis*P4H, and *trans*P4H); L-Pip had been previously reported to be a very good proline hydroxylase substrate,^{191,195,198,201} and so it was thought that new, uncharacterised products may be produced from the PH enzymes from L-Pip. The results showed that, with a

L-Pip substrate, *cis*P3H Type I produces *cis*-3-hydroxy-L-pipecolic acid (*cis*-3-OH-L-Pip) (100% yield), *cis*P4H produces *cis*-3-hydroxy-L-pipecolic acid (*cis*-3-OH-L-Pip) (~45% yield) and *cis*-5-hydroxy-L-pipecolic acid (*cis*-5-OH-L-Pip) (~38% yield), and *trans*P4H produces *trans*-5-hydroxy-L-pipecolic acid (*trans*-5-OH-L-Pip) (100% yield) (Schemes 3.30 a-c), respectively; extracted-ion chromatograms (XICs) showing 146.1 m/z [M+H]⁺ ions contained 1 peak, 2 peaks, and 1 peak for reactions using *cis*P3H Type I, *cis*P4H, and *trans*P4H (Figures 3.55 a-c), respectively. In addition to these characterised products,^{191,195,198,201} LC/MS-analysis showed that the *cis*P4H reactions with L-Pip also contained a product with a 162.1 m/z [M+H]⁺ mass (Figure 3.55 b); this LC/MS peak corresponded with the dihydroxylation product, (3*R*,5*S*)-dihydroxy-L-pipecolic acid [(3*R*,5*S*)-3,5-DiOH-L-Pip]. The (3*R*,5*S*)-3,5-DiOH-L-Pip product has apparently gone undetected in previous studies.^{191,195,198,201}



Scheme 3.30: Proposed reactions catalysed by the proline hydroxylases using an L-pipecolic acid (L-Pip) substrate [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H]:

- (a) *cis*P3H Type I produces *cis*-3-hydroxy-L-pipecolic acid (*cis*-3-OH-L-Pip);
- (b) *cis*P4H produces *cis*-3-hydroxy-L-pipecolic acid (*cis*-3-OH-L-Pip), *cis*-5-hydroxy-L-pipecolic acid (*cis*-5-OH-L-Pip), and (3*R*,5*S*)-dihydroxy-L-pipecolic acid [(3*R*,5*S*)-3,5-DiOH-L-Pip];
- (c) *trans*P4H produces *trans*-5-hydroxy-L-pipecolic acid (*trans*-5-OH-L-Pip).

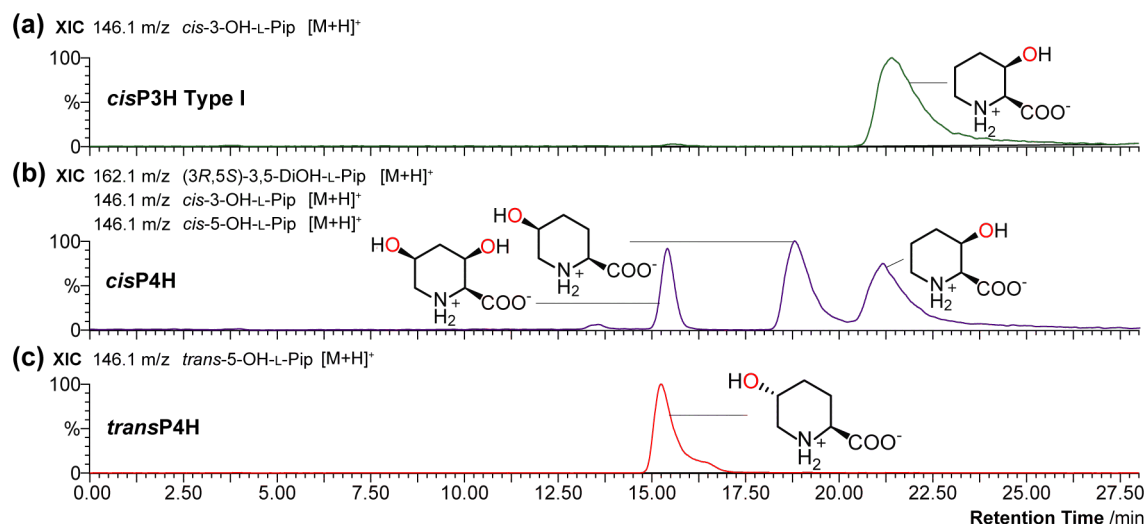


Figure 3.55: LC/MS-based analysis of proline hydroxylase [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H] reactions with L-pipecolic acid (L-Pip).

Extracted-ion chromatogram showing 146.1 m/z [M+H]⁺ ions contained peaks corresponding to:

(a) *cis*-3-hydroxy-L-pipecolic acid (*cis*-3-OH-L-Pip);

(b) *cis*-3-hydroxy-L-pipecolic acid (*cis*-3-OH-L-Pip) and *cis*-5-hydroxy-L-pipecolic acid (*cis*-5-OH-L-Pip);

(c) *trans*-5-hydroxy-L-pipecolic acid (*trans*-5-OH-L-Pip).

In addition, *cis*P4H reactions contained 162.1 m/z [M+H]⁺ ions corresponding to:

(b) (3*R*,5*S*)-dihydroxy-L-pipecolic acid [(3*R*,5*S*)-3,5-DiOH-L-Pip].

PH reactions used 15 μM *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μM (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM L-Pip. 100 μL reactions were carried out in at room temperature (25 °C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μm) [4.6 x 250mm].

The L-Pip PH reactions were scaled up for purification of the products for NMR characterisation; below is a spectrum used to assign the structures of the novel *cis*P4H reaction product, (3*R*,5*S*)-dihydroxy-L-pipecolic acid [(3*R*,5*S*)-3,5-DiOH-L-Pip] (Figure 3.56).

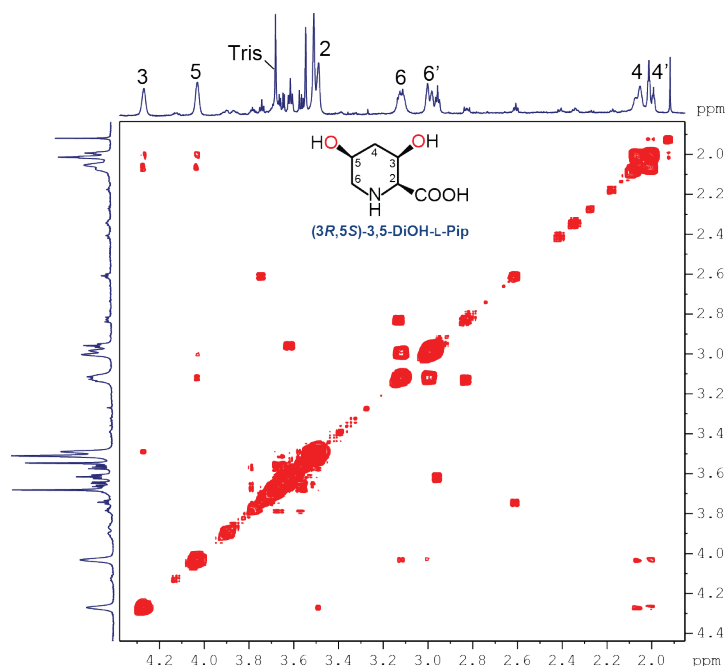
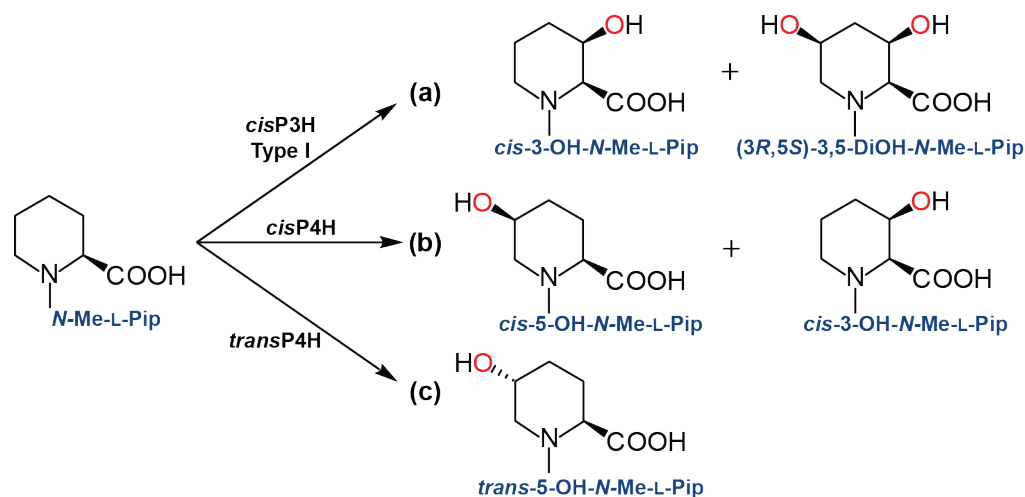


Figure 3.56: COSY-NMR analysis of the product from the *cis*P4H reactions with L-pipecolic acid (L-Pip): NMR assignments confirmed that the product is (3*R*,5*S*)-dihydroxy-L-pipecolic acid [(3*R*,5*S*)-3,5-DiOH-L-Pip].

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D_2O containing 0.05% (w/v) TSP- d_4 , NMR chemical shift reference/internal standard.

Another 6-membered proline analogue, *N*-methyl-L-pipecolic acid (*N*-Me-L-Pip), was also tested as a proline hydroxylase (*cis*P3H Type I, *cis*P4H, and *trans*P4H) substrates; like L-Pip, *N*-Me-L-Pip is hydroxylated by the proline hydroxylases in either of positions C3 or C5, with *cis*P3H Type I producing *cis*-3-hydroxy-*N*-methyl-L-pipecolic acid (*cis*-3-OH-*N*-Me-L-Pip) (~24% yield), *cis*P4H producing *cis*-3-hydroxy-*N*-methyl-L-pipecolic acid (*cis*-3-OH-*N*-Me-L-Pip) (~5% yield) and *cis*-5-hydroxy-*N*-methyl-L-pipecolic acid (*cis*-5-OH-*N*-Me-L-Pip) (~11% yield), and *trans*P4H producing *trans*-5-hydroxy-*N*-methyl-L-pipecolic acid (*trans*-5-OH-*N*-Me-L-Pip) (~25% yield) from L-Pip (Schemes 3.31 a-c), respectively. LC/MS analyses of the proline hydroxylase reaction using a *N*-Me-L-Pip substrate showed single-hydroxylation products (160.1 m/z $[\text{M}+\text{H}]^+$) peaks (Figures 3.57 a-c): the *cis*P3H Type I reaction

extracted-ion chromatogram (XIC) (showing 146.1 m/z $[M+H]^+$ ions) contains a single peak corresponding to *cis*-3-OH-*N*-Me-L-Pip (Figure 3.57 a); the *cis*P4H reaction extracted-ion chromatogram (XIC) (showing 146.1 m/z $[M+H]^+$ ions) contains two peaks corresponding to *cis*-3-OH-*N*-Me-L-Pip and *cis*-5-OH-*N*-Me-L-Pip (Figure 3.57 b); and the *trans*P4H reaction extracted-ion chromatogram (XIC) (showing 146.1 m/z $[M+H]^+$ ions) contains two peaks corresponding to *trans*-4-OH-*N*-Me-L-Pip (Figure 3.57 c). In addition, as with the L-Pip substrate, dihydroxylation of *N*-Me-L-Pip occur to form (3*R*,5*S*)-3,5-dihydroxy-*N*-methyl-L-pipecolic acid [(3*R*,5*S*)-3,5-DiOH-*N*-Me-L-Pip], a reaction which is catalysed by *cis*P3H Type I in this case (as opposed to the L-Pip dihydroxylation reaction, which is catalysed by *cis*P4H) (Scheme 3.31 a); LC/MS-analyses of *cis*P3H Type I reactions using an *N*-Me-L-Pip substrate show a peak corresponding to (3*R*,5*S*)-3,5-DiOH-*N*-Me-L-Pip (176.1 m/z $[M+H]^+$ ions) (Figure 3.57 a).



Scheme 3.31: Proposed reactions catalysed by the proline hydroxylases using an *N*-methyl-L-pipecolic acid (*N*-Me-L-Pip) substrate [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H]:

(a) *cis*P3H Type I produces *cis*-3-hydroxy-L-pipecolic acid (*cis*-3-OH-L-Pip) and (3*R*,5*S*)-dihydroxy-*N*-methyl-L-pipecolic acid [(3*R*,5*S*)-3,5-DiOH-*N*-Me-L-Pip];

(b) *cis*P4H produces *cis*-3-hydroxy-*N*-methyl-L-pipecolic acid (*cis*-3-OH-*N*-Me-L-Pip), *cis*-5-hydroxy-*N*-methyl-L-pipecolic acid (*cis*-5-OH-*N*-Me-L-Pip);

(c) *trans*P4H produces *trans*-5-hydroxy-L-pipecolic acid (*trans*-5-OH-L-Pip).

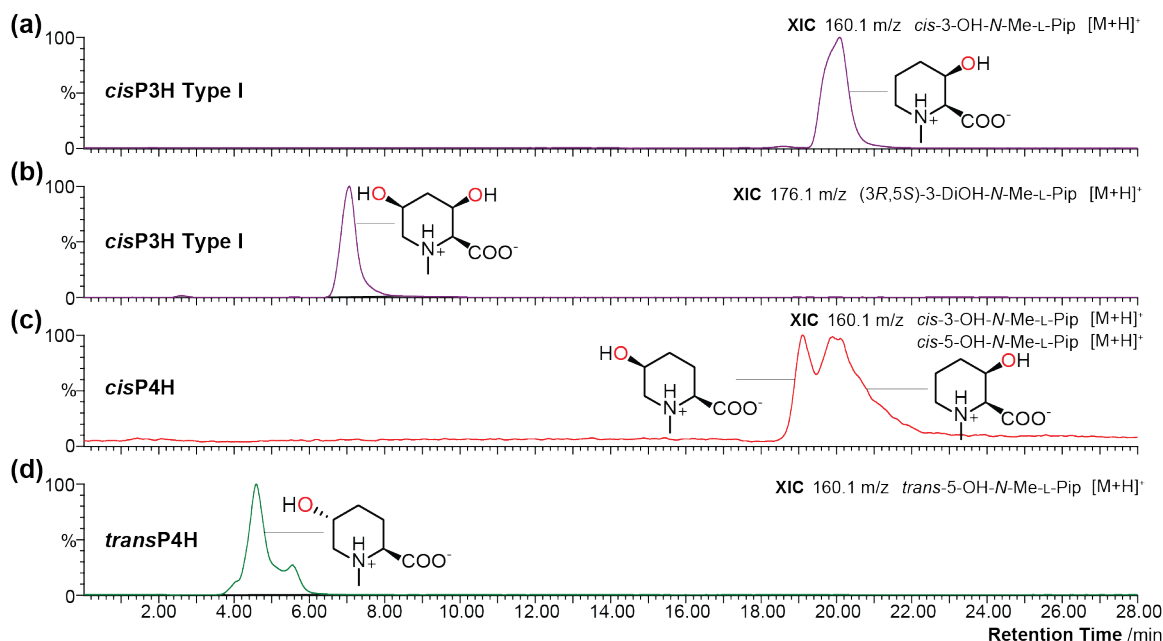


Figure 3.57: LC/MS-based analysis of proline hydroxylase reactions with *N*-methyl-L-pipecolic acid (L-Pip) [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H].

Extracted-ion chromatogram showing 160.1 m/z $[M+H]^+$ ions contained peaks corresponding to:

- (a) *cis*-3-hydroxy-*N*-methyl-L-pipecolic acid (*cis*-3-OH-*N*-Me-L-Pip);
- (b) *cis*-3-hydroxy-*N*-methyl-L-pipecolic acid (*cis*-3-OH-*N*-Me-L-Pip) and *cis*-5-hydroxy-*N*-methyl-L-pipecolic acid (*cis*-5-OH-*N*-Me-L-Pip);
- (c) *trans*-5-hydroxy-*N*-methyl-L-pipecolic acid (*trans*-5-OH-*N*-Me-L-Pip).

In addition, *cis*P3H Type I reactions contained 176.1 m/z $[M+H]^+$ ions corresponding to:

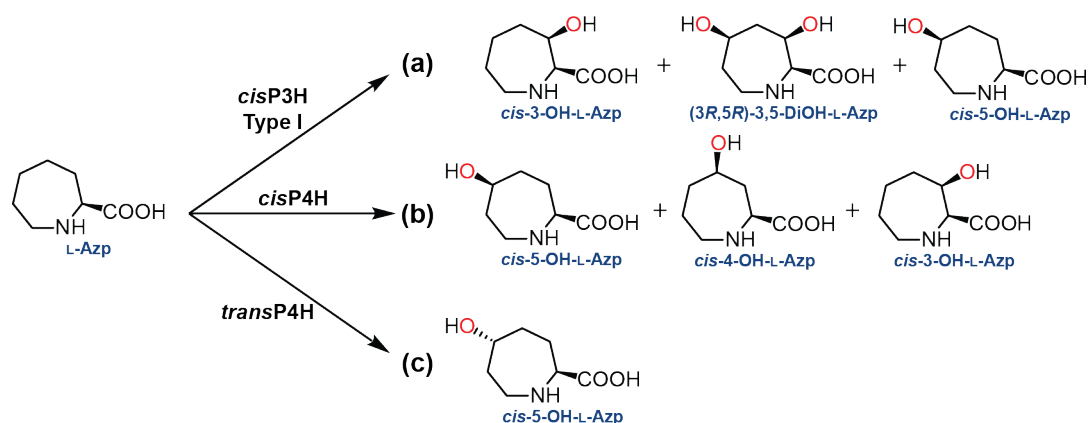
- (a) (3*R*,5*S*)-dihydroxy-*N*-methyl-L-pipecolic acid [(3*R*,5*S*)-3,5-DiOH-*N*-Me-L-Pip].

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 10 mM 2-OG, 1 mM *N*-Me-L-Pip. 100 μ L reactions were carried out in at room temperature (25 $^\circ\text{C}$) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 $\text{\AA}$, 5 μ m) [4.6 x 250mm].

A 7-membered proline analogue, L-azepane-2-carboxylic acid (L-Azp), was tested as a proline hydroxylase (*cis*P3H Type I, *cis*P4H, and *trans*P4H) substrate; LC/MS-analysis indicated that a number of products were produced, and there products are thought to be produced by C3- or C5-hydroxylation (or a combination there of), in a fashion analogous that seen with an L-Pip substrate (Schemes 3.32 a-c). For example, LC/MS-chromatograms from analysed *cis*P3H Type I reactions contained two 160.1 m/z $[M+H]^+$ peaks thought to be *cis*-5-hydroxy-L-azepane-carboxylic

acid (*cis*-5-OH-L-Azp) (~5% yield) and *cis*-3-hydroxy-L-azepane-carboxylic acid (*cis*-3-OH-L-Azp) (~31% yield) (Figure 3.58 a); additionally, a 176.1 m/z $[M+H]^+$ peak was observed (~3% yield), which is thought to correspond to the *cis*-3,5-dihydroxylation product, (3*R*,5*R*)-dihydroxy-L-azepane-2-carboxylic acid [(3*R*,5*R*)-DiOH-L-Azp] (Figure 3.58 a). LC/MS-chromatograms from analysed *cis*P4H reactions contained three 160.1 m/z $[M+H]^+$ peaks, thought to be *cis*-5-hydroxy-L-azepane-carboxylic acid (*cis*-5-OH-L-Azp) (~8% yield), *cis*-4-hydroxy-L-azepane-carboxylic acid (*cis*-4-OH-L-Azp) (~3% yield), and *cis*-3-hydroxy-L-azepane-carboxylic acid (*cis*-3-OH-L-Azp) (~2.5% yield) (Figure 3.58 b). LC/MS-chromatograms from analysed *trans*P4H reactions contained a single 160.1 m/z $[M+H]^+$ peak, thought to be *trans*-5-hydroxy-L-azepane-carboxylic acid (*trans*-5-OH-L-Azp) (~8% yield) (Figure 3.58 c).



Scheme 3.32: Proposed reactions catalysed by the proline hydroxylases using an L-azepane-2-carboxylic acid (L-Apn) substrate [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H]:

(a) *cis*P3H Type I produces *cis*-3-hydroxy-L-azepane-2-carboxylic acid (*cis*-3-OH-L-Apn), *cis*-5-hydroxy-L-azepane-2-carboxylic acid (*cis*-5-OH-L-Apn), and (3*R*,5*R*)-dihydroxy-L-azepane-2-carboxylic acid [(3*R*,5*R*)-DiOH-L-Azp];

(b) *cis*P4H produces *cis*-5-hydroxy-L-azepane-carboxylic acid (*cis*-5-OH-L-Azp), *cis*-4-hydroxy-L-azepane-carboxylic acid (*cis*-4-OH-L-Azp), and *cis*-3-hydroxy-L-azepane-carboxylic acid (*cis*-3-OH-L-Azp);

(c) *trans*P4H produces *trans*-5-hydroxy-L-azepane-2-carboxylic acid (*trans*-5-OH-L-Apn).

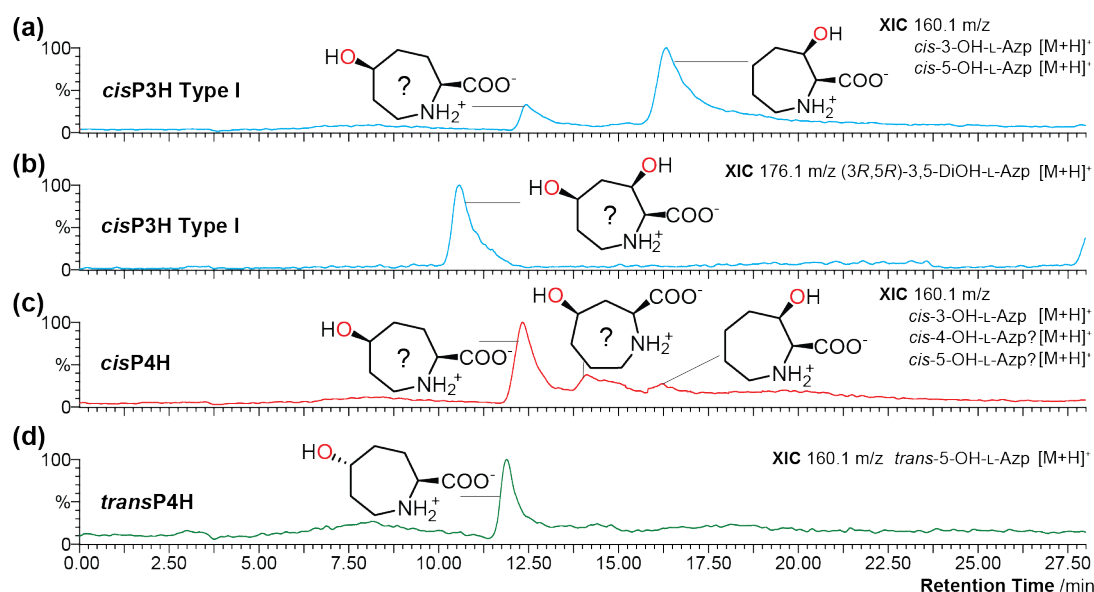


Figure 3.58: LC/MS-based analysis of proline hydroxylase reactions with L-azepane-2-carboxylic acid (L-Azp) [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H].

Extracted-ion chromatogram showing 160.1 m/z $[M+H]^+$ ions contained peaks corresponding to: (a) *cis*-3-hydroxy-L-azepane-2-carboxylic acid (*cis*-3-OH-L-Apn) and *cis*-5-hydroxy-L-azepane-2-carboxylic acid (*cis*-5-OH-L-Apn);

(b) *cis*-5-hydroxy-L-azepane-carboxylic acid (*cis*-5-OH-L-Azp), *cis*-4-hydroxy-L-azepane-carboxylic acid (*cis*-4-OH-L-Azp), and *cis*-3-hydroxy-L-azepane-carboxylic acid (*cis*-3-OH-L-Azp);

(c) *trans*P4H produces *trans*-5-hydroxy-L-azepane-2-carboxylic acid (*trans*-5-OH-L-Apn).

In addition, *cis*P3H reactions contained 176.1 m/z $[M+H]^+$ ions corresponding to:

(a) (3*R*,5*R*)-dihydroxy-L-azepane-2-carboxylic acid [(3*R*,5*R*)-DiOH-L-Azp].

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 10 mM 2-OG, 1 mM L-Azp. 100 μ L reactions were carried out in at room temperature (25 $^\circ\text{C}$) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 $\text{\AA}$, 5 μ m) [4.6 x 250mm].

The proline hydroxylase (*cis*P3H Type I, *cis*P4H, and *trans*P4H) reactions with L-Azp were scaled up for HPLC-purification for the products to be fully characterised by NMR; to date evidence has been acquired to confirm that *cis*P3H Type I and *trans*P4H produce *cis*-3-OH-L-Azp and *trans*-5-OH-L-Azp, respectively (the other products are partially characterised). For example, Figure 3.59 shows the COSY spectrum of the *cis*-3-hydroxy-L-azepane-2-carboxylic acid (*cis*-3-OH-L-Azp) product from the *cis*P3H Type I reaction using an L-Azp substrate.

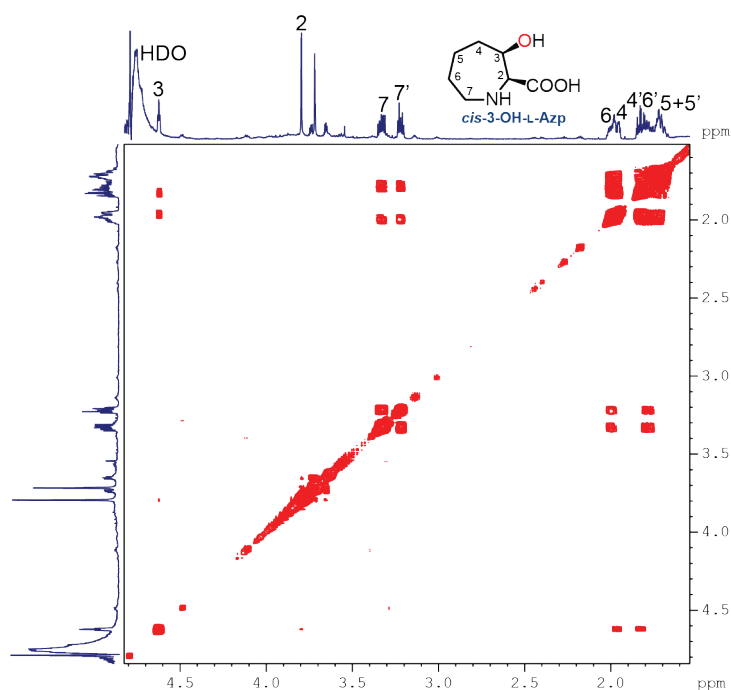


Figure 3.59: COSY-NMR analysis of the product from the *cis*P3H reaction with L-azepane-2-carboxylic acid (L-Azp): NMR assignments confirmed that the product is *cis*-3-hydroxy-L-azepane-2-carboxylic acid (*cis*-3-OH-L-Azp).

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D_2O containing 0.05% (w/v) TSP- d_4 , NMR chemical shift reference/internal standard.

3.4.4.4.1 Substrate Turnover Summary: Substrate Ring Size

The following table (Table 3.12) summarises the proline hydroxylase activities with proline analogues of various carbon-ring sizes (which were screened for proline hydroxylase substrate activity in this section). See below for details of the reaction yields and products.

Substrate	Length/h	Product(s) [+ Yield(s)]			
		<i>cis</i> P3H Type I	<i>cis</i> P4H	<i>trans</i> P4H	
	14	 "cis-3-OH-Me-tCMP" 100%	 "cis-3-OH-Me-tCMP" 4%	-	0%
	14	 <i>cis</i> -3-OH-L-Pip 100%	 <i>cis</i> -5-OH-L-Pip 38% <i>cis</i> -3-OH-L-Pip 45% (3 <i>R</i> ,5 <i>S</i>)-3,5-DiOH-L-Pip 17%	 <i>trans</i> -5-OH-L-Pip 100%	
	14	 <i>cis</i> -3-OH-N-Me-L-Pip 24% (3 <i>R</i> ,5 <i>S</i>)-3,5-DiOH-N-Me-L-Pip 5%	 <i>trans</i> -5-OH-N-Me-L-Pip 5% <i>cis</i> -3-OH-N-Me-L-Pip 11%	 <i>trans</i> -5-OH-L-Pip 100%	
	14	 <i>cis</i> -3-OH-L-Azp 31% <i>cis</i> -5-OH-L-Azp 5% (3 <i>R</i> ,5 <i>R</i>)-3,5-DiOH-L-Azp 3%	 <i>cis</i> -3-OH-L-Azp 2.5% <i>cis</i> -4-OH-L-Azp 3% <i>cis</i> -5-OH-L-Azp 8%	 <i>trans</i> -5-OH-L-Azp 8%	

Table 3.12: Proline hydroxylase substrate analogue screening: the effects of substrate ring size on proline hydroxylase activity.

L-azetidine-2-carboxylic acid (L-Aze), L-pipecolic acid (L-Pip), *N*-methyl-L-pipecolic acid (*N*-Me-L-Pip), and L-azepane-2-carboxylic acid (L-Azp) were screened for proline hydroxylase activity.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM Substrate (and 250 μ M (6*S*)-6-(isopenten-2-yl)-Me-tCMP).

3.4.4.5 Proline analogues with Fused & Bridged Ring Systems

Proline analogues containing fused or bridged ring systems (Figure 3.60) were then tested as substrates of the proline hydroxylases (*cis*P3H Type I, *cis*P4H, and *trans*P4H).

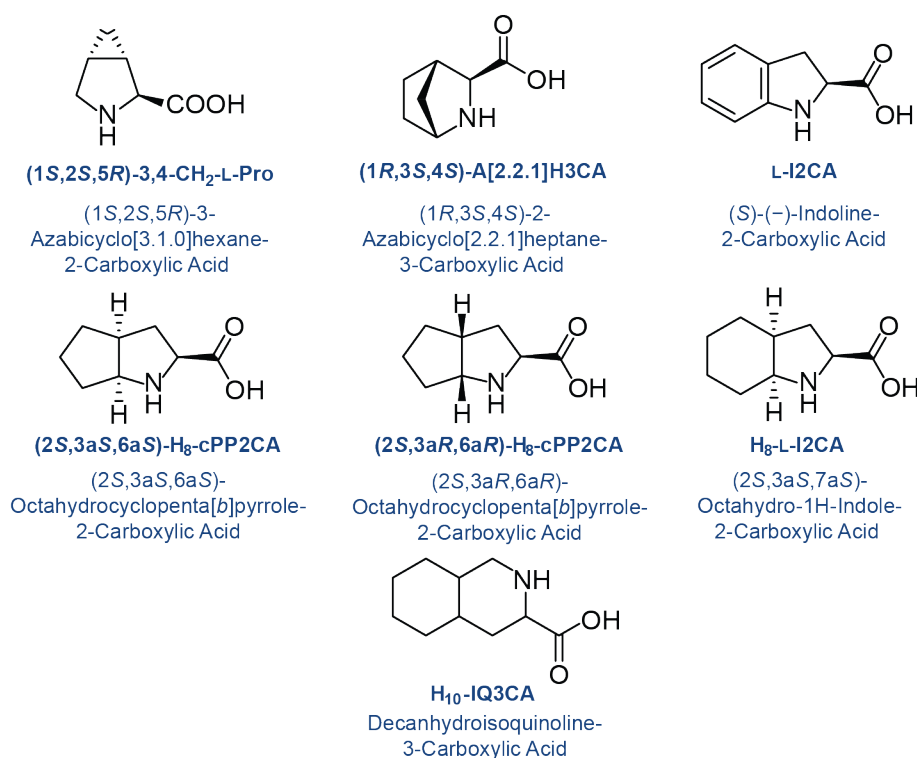


Figure 3.60: Proline analogues with fused and bridged ring systems were tested as proline hydroxylase (*cis*P3H Type I, *cis*P4H, *trans*P4H).

Compounds included: (1S,2S,5R)-3-azabicyclo[3.1.0]hexane-2-carboxylic acid [(1S,2S,5R)-3,4-CH₂-L-Pro], (1R,3S,4S)-2-azabicyclo[2.2.1]heptane-3-carboxylic acid [(1R,3S,4S)-A[2.2.1]H3CA], (S)-(-)-indoline-2-carboxylic acid (L-I2CA), (2S,3aS,6aS)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2S,3aS,6aS)-H₈-cPP2CA], (2S,3aR,6aR)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2S,3aR,6aR)-H₈-cPP2CA], (2S,3aS,7aS)-octahydro-1H-indole-2-carboxylic acid (H₈-L-I2CA), and decanhydroisoquinoline-3-carboxylic acid (H₁₀-IQ3CA).

LC/MS-analyses of *cis*P3H Type I and *cis*P4H reactions involving (1*S*,2*S*,5*R*)-3-azabicyclo[3.1.0]hexane-2-carboxylic acid [(1*S*,2*S*,5*R*)-3,4-CH₂-L-Pro] showed the presence of a 144.1 m/z [M+H]⁺ substance missing from negative control reactions; extracted-ion chromatograms (XICs) contained a broad, shouldered peak (Figures 3.61 a and b), suggesting that *cis*P3H Type I and *cis*P4H could, potentially, hydroxylate the (1*S*,2*S*,5*R*)-3,4-CH₂-L-Pro substrate. Klein *et al.* previously reported that both *cis*P3Hs and *cis*P4H hydroxylate *trans*-3-methyl-L-proline (*trans*-3-Me-L-Pro) to form a tertiary alcohol, (3*R*,4*S*)-4-hydroxy-3-methyl-L-proline [(3*R*,4*S*)-4-OH-3-Me-L-Pro],²⁹⁵ and so, perhaps, the same enzymes could hydroxylate the C3- and/or C4-positions to form a tertiary alcohol product; a shoulder in the peak may suggest a ‘major’ and a ‘minor’ product being formed. However, turnover was, at best, very low (<1%), peak integration suggesting yields of 0.35% and 0.25% for *cis*P3H Type I and *cis*P4H, respectively, and therefore it is not fully clear whether the peaks really correspond to substrate hydroxylation products, or whether these peaks are just artifacts of the method; the fact the 144.1 m/z [M+H]⁺ peaks are only missing from the *trans*P4H samples (a reaction which is blocked by the *trans*-position of the 3-membered ring) may suggest that the enzyme-catalysed reaction is genuine (but very slow).

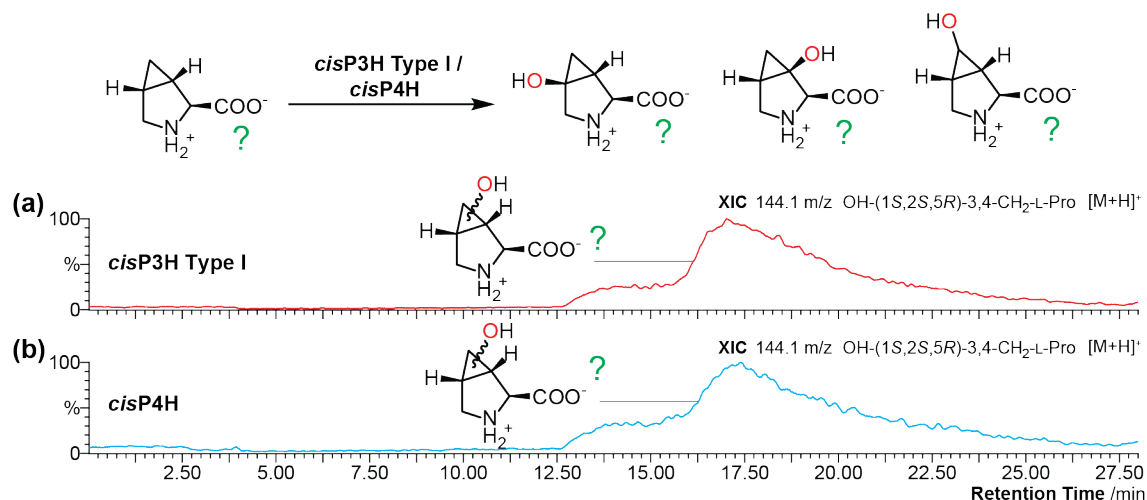


Figure 3.61: LC/MS-based analysis of proline hydroxylase reactions with (1S,2S,5R)-3-azabicyclo[3.1.0]hexane-2-carboxylic acid [(1S,2S,5R)-3,4-CH₂-L-Pro] [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H].

Extracted-ion chromatogram showing 144.1 m/z [M+H]⁺ ions contained peaks potentially corresponding to hydroxylated (1S,2S,5R)-3-azabicyclo[3.1.0]hexane-2-carboxylic acid [OH-(1S,2S,5R)-3,4-CH₂-L-Pro].

PH reactions used 15 μM *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μM (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM (1S,2S,5R)-3,4-CH₂-L-Pro. 100 μL reactions were carried out in at room temperature (25 °C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μm) [4.6 x 250mm]. These measurements were collected alongside Dr Refaat Hamed.

When (1R,3S,4S)-2-azabicyclo[2.2.1]heptane-3-carboxylic acid [(1R,3S,4S)-A[2.2.1]H3CA] and (S)-(-)-indoline-2-carboxylic acid (L-I2CA) were tested as proline hydroxylase substrates, they were found to be non-substrates; the methylene bridge in (1R,3S,4S)-A[2.2.1]H3CA must sterically hinder the PH-catalysed hydroxylation, and the delocalisation of L-I2CA's ring amine into the aromatic ring makes amine lone pair less available for hydrogen bonding/protonation, rendering the substrate unhydroxylatable (substrates were found to require amine functionality for PH activity).

Two related proline analogues containing two fused, 5-membered rings (Figure 3.62) were also tested as substrates of the proline hydroxylases (*cis*P3H Type I, *cis*P4H, and *trans*P4H) and interesting hydroxylation selectivity trends were ob-

served; in these compounds the ‘second’ 5-membered ring is fused to carbons 4 and 5 of L-proline (Figure 3.62). In (2*S*,3*aS*,6*aS*)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aS*,6*aS*)-H₈-cPP2CA], the ‘second’ 5-membered ring is fused to L-proline so that the ring backbone is ‘*cis*’ with respect to the carboxylic acid group [(3*aS*,6*aS*)-stereochemistry] (*i.e.* both are on the same side/face of the proline ring) (Figure 3.62 a); in (2*S*,3*aR*,6*aR*)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aR*,6*aR*)-H₈-cPP2CA], the second 5-membered ring is fused to L-proline so that the ring backbone is ‘*trans*’ with respect to the carboxylic acid group [(3*aR*,6*aR*)-stereochemistry] (*i.e.* both are on the same side/face of the proline ring) (Figure 3.62 b).

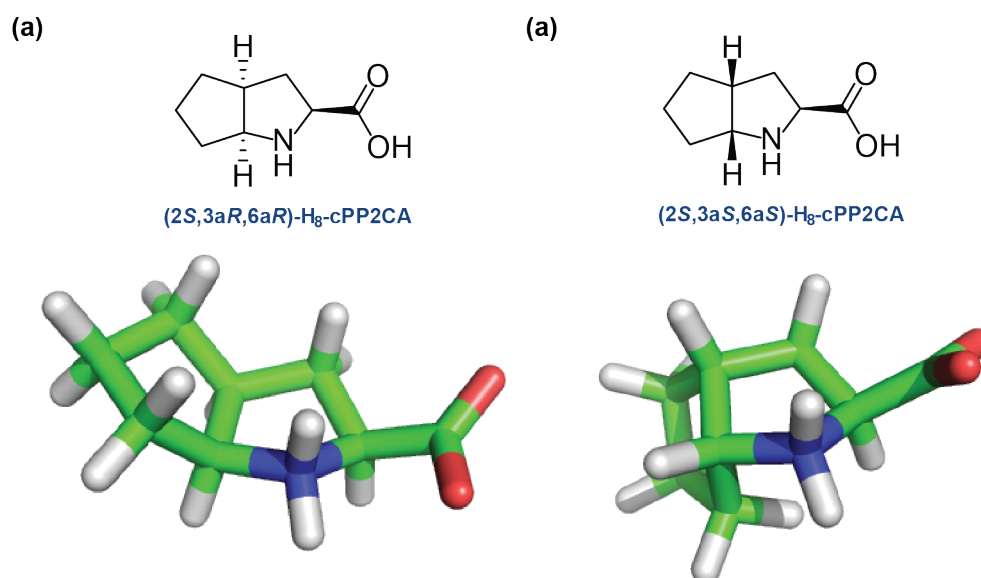


Figure 3.62: Ring conformations of proline analogues with a C4,C5-fused 5-membered ring:

- (a) in (2*S*,3*aS*,6*aS*)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aS*,6*aS*)-H₈-cPP2CA], the second 5-membered ring is fused to L-proline so that the ring backbone is ‘*cis*’ with respect to the carboxylic acid group [(3*aS*,6*aS*)-stereochemistry];
- (b) in (2*S*,3*aR*,6*aR*)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aR*,6*aR*)-H₈-cPP2CA], the second 5-membered ring is fused to L-proline so that the ring backbone is ‘*trans*’ with respect to the carboxylic acid group [(3*aR*,6*aR*)-stereochemistry].

LC/MS-analysis of proline hydroxylase (*cis*P3H Type I, *cis*P4H, and *trans*P4H) reactions showed that (2*S*,3*aS*,6*aS*)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aS*,6*aS*)-H₈-cPP2CA] was a substrate of all three proline hydroxylase enzymes (Figures 3.63 a-c); extracted-ion chromatograms (XICs) showing 172.1 m/z [M+H]⁺ ions contained broad peaks, suggesting hydroxylation products were present. For the *cis*P3H Type I reaction, 2 peaks were seen (retention time = ~15 min and ~21 min) in the 172.1 m/z XIC (Figure 3.63 a), suggesting two products are produced by *cis*P3H Type I (thought to be the *cis*-4- and *cis*-3-hydroxylation products, respectively); the products were produced with yields of ~17% and ~5%, respectively. The *cis*P4H 172.1 m/z XIC contained one peak with a retention time of ~15 min (Figure 3.63 b), a retention time which is roughly the same as that for the first product from the *cis*P3H Type I reaction; this *cis*P4H product is also thought to give the same, *cis*-4-hydroxylation product; the yields was ~3.5%. Analysis of the *trans*P4H reaction showed one 172.1 m/z [M+H]⁺ product, which may be the *trans*-4-hydroxylation product; the yields was ~1%.

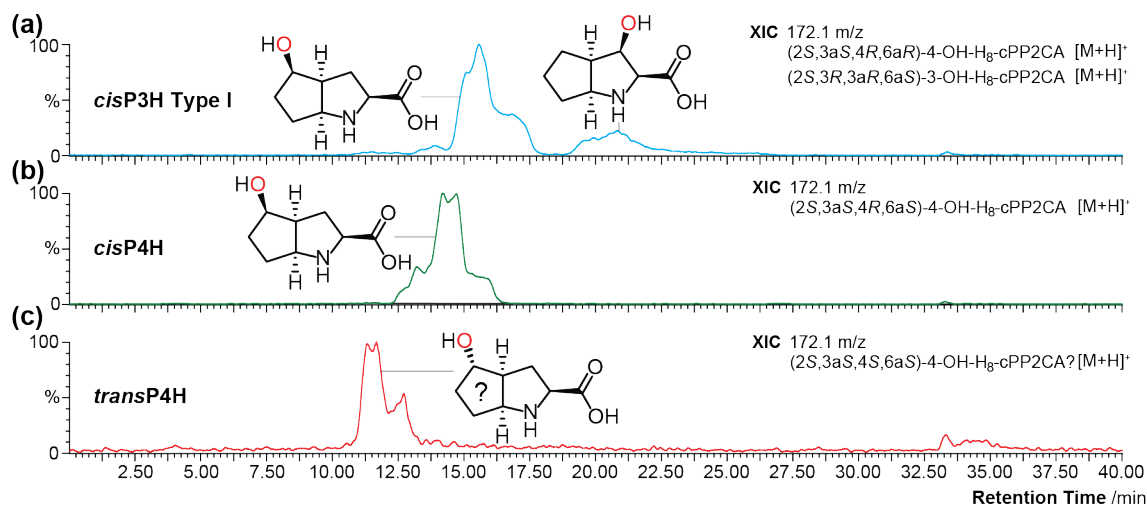


Figure 3.63: LC/MS-based analysis of proline hydroxylase reactions with (2*S*,3*aS*,6*aS*)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aS*,6*aS*)-H₈-cPP2CA] [(a) *cis*P3H Type I, (b) *cis*P4H, and (c) *trans*P4H].

Extracted-ion chromatogram showing 172.1 m/z [M+H]⁺ ions contained peaks thought to correspond to:

- (a) (2*S*,3*aS*,4*R*,6*aR*)-4-hydroxy-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aS*,4*R*,6*aR*)-4-OH-H₈-cPP2CA] and (2*S*,3*R*,3*aS*,6*aR*)-3-hydroxy-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*R*,3*aS*,6*aR*)-3-OH-H₈-cPP2CA];
- (b) (2*S*,3*aS*,4*R*,6*aS*)-4-hydroxy-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aS*,4*R*,6*aS*)-4-OH-H₈-cPP2CA];
- (c) (2*S*,3*aS*,4*S*,6*aS*)-4-hydroxy-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aS*,4*S*,6*aS*)-4-OH-H₈-cPP2CA].

PH reactions used 15 μM *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μM (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM (2*S*,3*aS*,6*aS*)-H₈-cPP2CA. 100 μL reactions were carried out in at room temperature (25 °C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μm) [4.6 x 250mm]. These measurements were collected with Dr Refaat Hamed.

The fused ring of (2*S*,3*aS*,6*aS*)-H₈-cPP2CA is pointing in a direction ('up' in my figures) towards the anticipated hydroxylation sites of *cis*P3H Type I and *cis*P4H (making it accessible for PH hydroxylation), and the C4-bridge proton is pointing away from ferryl species (making it an unfavourable hydroxylation site); therefore, it was thought that these enzymes may favour hydroxylation of the second ring over the 'usual' C3 or C4 sites. The major product from the *cis*P3H Type I and *cis*P4H reactions was purified from large scale PH-reactions (20-fold larger), and characterised by NMR (Figure 3.63). Although the sample was not entirely pure (~20%), the product NMR peaks were resolvable from impurity peaks (Figure 3.64 a); gradient-enhanced 1D-selective TOCSY spectra were also used to confirm which peaks belonged to the (2*S*,3*aS*,6*aS*)-H₈-cPP2CA product (Figure 3.64 a). COSY (correlation spectroscopy) experiments confirmed that the hydroxylation site of the major product was, indeed, on the second, fused, 5-membered ring, with spectra being consistent with *cis*-4-hydroxylation (Figure 3.64 b). The (2*S*,3*aS*,6*aS*)-H₈-cPP2CA substrate is therefore the first PH substrate found to have PH hydroxylation activity at a site which is not on the proline ring system.

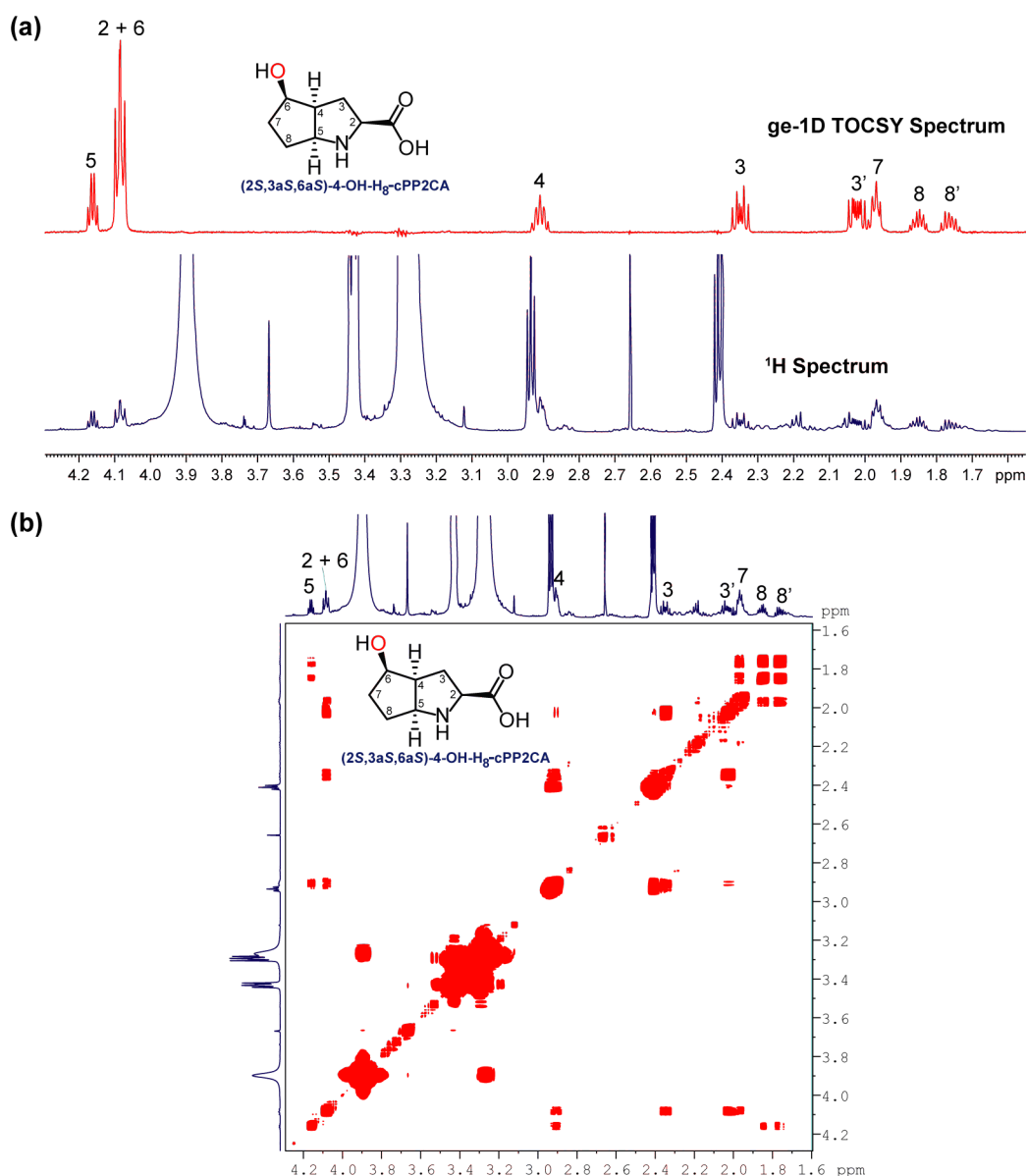


Figure 3.64: Proton-, TOCSY-, COSY-NMR analysis of the ‘major’ product from *cis*P3H Type I and *cis*P4H reactions using a $(2S,3aS,6aS)$ - H_8 -cPP2CA substrate: NMR assignments confirmed that the product is $(2S,3aS,4R,6aR)$ -4-hydroxy-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [$(2S,3aS,4R,6aR)$ -4-OH- H_8 -cPP2CA].

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1H/^{13}C/^{15}N$ TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μ L) D_2O containing 0.05% (w/v) TSP- d_4 , NMR chemical shift reference/internal standard. This compound was purified and analysed with Dr Refaat Hamed.

LC/MS-analyses of proline hydroxylase (*cis*P3H Type I, *cis*P4H, and *trans*P4H) reactions using a (2*S*,3*aR*,6*aR*)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aR*,6*aR*)-H₈-cPP2CA] substrate (a substrate with a C4-/C5-fused, ‘*trans*’ 5-membered ring) showed that (2*S*,3*aR*,6*aR*)-H₈-cPP2CA is a substrate of both *cis*P3H Type I and *trans*P4H (Figures 3.65 a and b); extracted-ion chromatograms showing 172.1 m/z [M+H]⁺ ions showed two products were produced by *cis*P3H Type I (~25% and ~16% yields, respectively) (Figure 3.65 a) and *trans*P4H (~1% yield) (Figure 3.65 b).

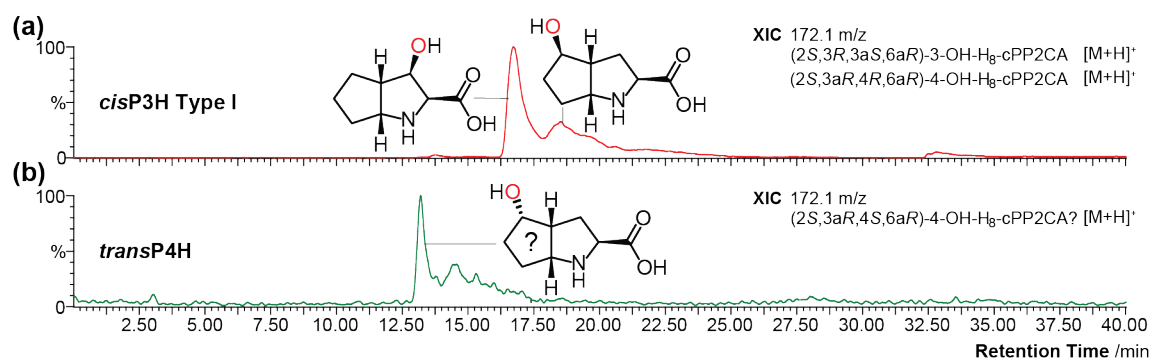


Figure 3.65: LC/MS-based analysis of proline hydroxylase reactions with (2*S*,3*aR*,6*aR*)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aR*,6*aR*)-H₈-cPP2CA] [(a) *cis*P3H Type I, and (b) *trans*P4H].

Extracted-ion chromatogram showing 172.1 m/z [M+H]⁺ ions contained peaks thought to correspond to:

(a) (2*S*,3*R*,3*aS*,6*aR*)-3-hydroxy-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*R*,3*aS*,6*aR*)-3-OH-H₈-cPP2CA] and (2*S*,3*aR*,4*S*,6*aR*)-4-hydroxy-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aR*,4*S*,6*aR*)-4-OH-H₈-cPP2CA];

(b) (2*S*,3*aR*,4*S*,6*aR*)-4-hydroxy-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aR*,4*S*,6*aR*)-4-OH-H₈-cPP2CA].

PH reactions used 15 μM *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μM (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM (2*S*,3*aR*,6*aR*)-H₈-cPP2CA. 100 μL reactions were carried out in at room temperature (25 °C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μm) [4.6 x 250mm]. These measurements were acquired with Dr Refaat Hamed.

Given that the second, fused ring system in this compound is '*trans*' to the carboxylic acid group (*i.e.* pointing 'down' away from the plane of the L-Pro ring), it is not unsurprising that (2*S*,3*aR*,6*aR*)-H₈-cPP2CA is not a *cis*P4H substrate. The only site hydroxylated by the PH enzymes in the other fused-ring substrate, (2*S*,3*aS*,6*aS*)-H₈-cPP2CA, is the part of the secondary ring system which points towards the preferred hydroxylation site of *cis*P4H; since the fused ring system of (2*S*,3*aR*,6*aR*)-H₈-cPP2CA points away from the preferred hydroxylation sites of *cis*P3H Type I and *cis*P4H, the fused ring positions are less favourable hydroxylation sites (or not at all in the case of *cis*P4H). On the basis of this reasoning, in contrast with PH reactions using a (2*S*,3*aS*,6*aS*)-H₈-cPP2CA substrate, it was predicted that *cis*P3H Type I would favour *cis*-3-hydroxylation over hydroxylation of the less-accessible C6-site (*i.e.* a more 'conventional' *cis*P3H-type selectivity), and the major product would be (2*S*,3*R*,3*aS*,6*aR*)-3-hydroxy-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*R*,3*aS*,6*aR*)-3-OH-H₈-cPP2CA].

To verify the identity of the 'major' product of the *cis*P3H Type I reaction with (2*S*,3*aR*,6*aR*)-H₈-cPP2CA (and confirm the hydroxylation selectivity preference), the product was purified from a 20-fold scaled-up reaction by HPLC, and characterised using NMR techniques (Figures 3.66 and 3.67); the NMR evidence measured for the 'major' product was consistent with the hypothesis that it was the *cis*-3-hydroxylated derivative of (2*S*,3*aR*,6*aR*)-H₈-cPP2CA.

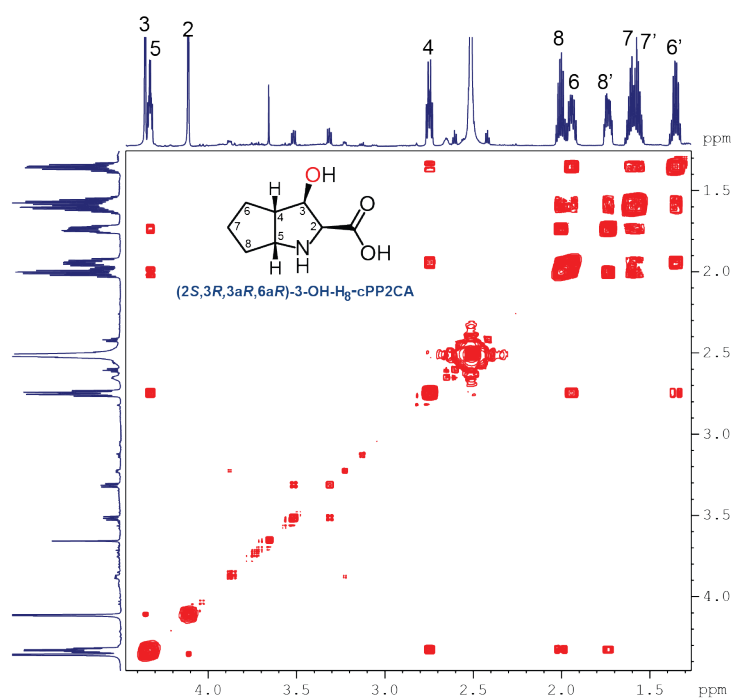


Figure 3.66: COSY-NMR analysis of the ‘major’ product from the *cis*P3H Type I reaction with (2*S*,3*aR*,6*aR*)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aR*,6*aR*)-H₈-cPP2CA]: NMR assignments confirmed that the product is (2*S*,3*R*,3*aS*,6*aR*)-3-hydroxy-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*R*,3*aS*,6*aR*)-3-OH-H₈-cPP2CA].

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a ¹H/¹³C/¹⁵N TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D₂O containing 0.05% (w/v) TSP-*d*₄, NMR chemical shift reference/internal standard. This compound was purified and analysed with Dr Refaat Hamed.

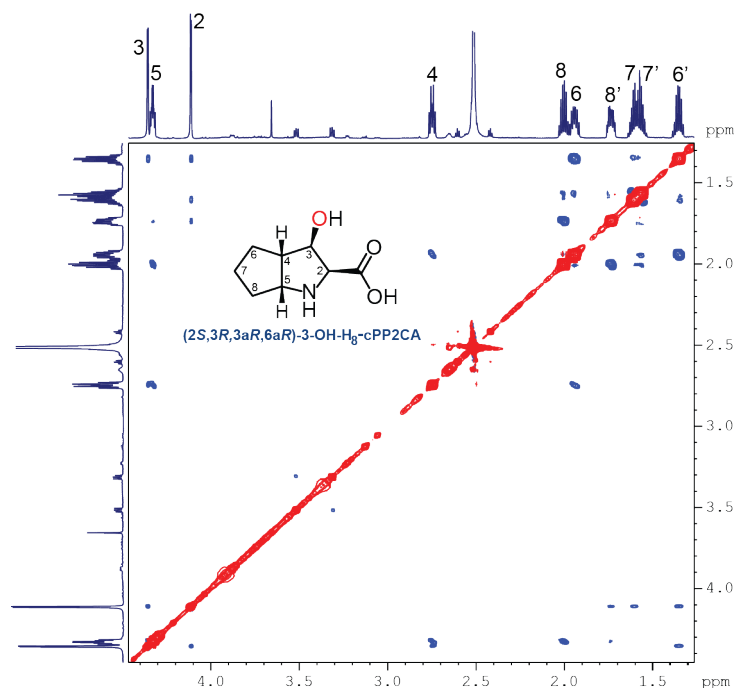


Figure 3.67: NOESY-NMR analyses of the ‘major’ product from the *cis*P3H Type I reaction with (2*S*,3*aR*,6*aR*)-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*aR*,6*aR*)- H_8 -cPP2CA]: NMR assignments confirmed that the product is (2*S*,3*R*,3*aS*,6*aR*)-3-hydroxy-octahydrocyclopenta[*b*]pyrrole-2-carboxylic acid [(2*S*,3*R*,3*aS*,6*aR*)-3-OH- H_8 -cPP2CA].

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D_2O containing 0.05% (w/v) TSP- d_4 , NMR chemical shift reference/internal standard. This compound was purified and analysed with Dr Refaat Hamed.

A proline analogue with a C4/C5-fused cyclohexane ring (6-membered), (2*S*,3*aS*,7*aS*)-octahydro-1*H*-indole-2-carboxylic acid (H_8 -L-I2CA), was then tested as a substrate of the proline hydroxylases (*cis*P3H Type I, *cis*P4H, and *trans*P4H); the ring junction of the compound is 'cis' to the carboxylic acid and points towards the favoured hydroxylation sites of *cis*P3H Type I and *cis*P4H, and, therefore, it was thought that may H_8 -L-I2CA also be hydroxylated on its secondary ring system [much like the (2*S*,3*aR*,4*S*,6*aR*)-4-OH- H_8 -cPP2CA substrate]. LC/MS-analysis showed that H_8 -L-I2CA is a *cis*P3H Type I substrate with a major product thought to be (2*S*,3*aS*,4*R*,7*aS*)-4-hydroxy-octahydro-1*H*-indole-2-carboxylic acid [(2*S*,3*aS*,4*R*,7*aS*)-4-OH- H_8 -L-I2CA]; extracted-ion chromatograms contained a 186.1 *m/z* [*M*+*H*]⁺ ion peak, and peak integration indicated the yield was 6% conversion (Figure 3.68).

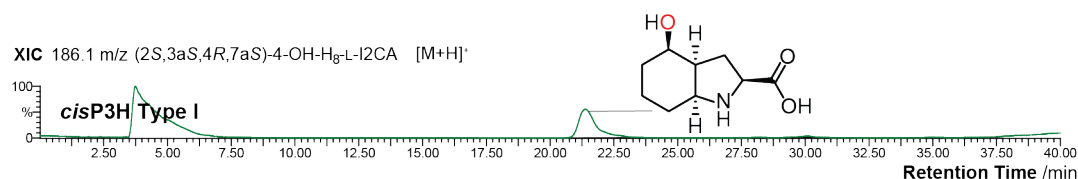


Figure 3.68: LC/MS-based analysis of proline hydroxylase (*cis*P3H Type I) reactions with (2

textit{S},3*aS*,7*aS*)-octahydro-1*H*-indole-2-carboxylic acid (H_8 -L-I2CA).

Extracted-ion-chromatograms (XICs) showing 186.1 *m/z* [*M*+*H*]⁺ ions contained a major peak, thought to correspond to (2*S*,3*aS*,4*R*,7*aS*)-4-hydroxy-octahydro-1*H*-indole-2-carboxylic acid [(2*S*,3*aS*,4*R*,7*aS*)-4-OH- H_8 -L-I2CA].

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M (NH₄)₂Fe(SO₄)₂, 10 mM 2-OG, 1 mM (2*S*,3*aR*,6*aR*)- H_8 -cPP2CA. 100 μ L reactions were carried out in at room temperature (25 °C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μ m) [4.6 x 250mm]. These results were measured alongside Dr Refaat Hamed.

The *cis*P3H Type I reaction using the H_8 -L-I2CA substrate was scaled up, and the product thought to be (2*S*,3*aS*,4*R*,7*aS*)-4-OH- H_8 -L-I2CA was HPLC purified for NMR-characterisation experiments; for example, Figures 3.69 and 3.70 show the COSY and NOESY spectra measured for the reaction product, which unambiguously confirms that *cis*P3H Type I converts H_8 -L-I2CA into (2*S*,3*aS*,4*R*,7*aS*)-4-OH- H_8 -L-

I2CA.

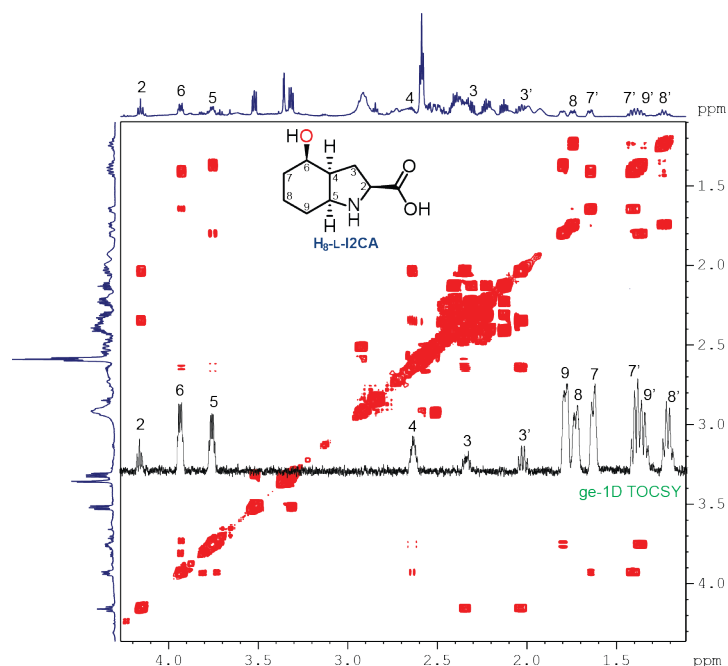


Figure 3.69: COSY-NMR analysis of the ‘major’ product from the *cis*P3H Type I reaction with (2*S*,3*aS*,7*aS*)-octahydro-1*H*-indole-2-carboxylic acid (H_8 -L-I2CA): NMR assignments confirmed that the product is (2*S*,3*aS*,4*R*,7*aS*)-4-hydroxy-octahydro-1*H*-indole-2-carboxylic acid [(2*S*,3*aS*,4*R*,7*aS*)-4-OH- H_8 -L-I2CA].

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D_2O containing 0.05% (w/v) TSP- d_4 , NMR chemical shift reference/internal standard. This compound was purified and analysed with Dr Refaat Hamed.

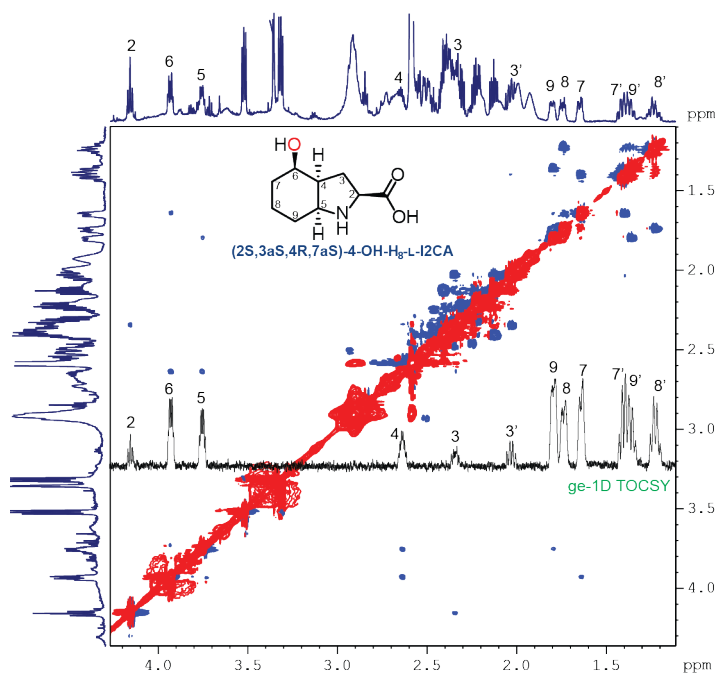


Figure 3.70: NOESY-NMR analyses of the ‘major’ product from the *cis*P3H Type I reaction with (2*S*,3*aS*,7*aS*)-octahydro-1*H*-indole-2-carboxylic acid (*H*₈-L-I₂CA): NMR assignments confirmed that the product is (2*S*,3*aS*,4*R*,7*aS*)-4-hydroxy-octahydro-1*H*-indole-2-carboxylic acid [(2*S*,3*aS*,4*R*,7*aS*)-4-OH-*H*₈-L-I₂CA].

Spectra were measured at 298 K using on a 700 MHz Bruker AVIII 700 spectrometer (with a ¹H/¹³C/¹⁵N TCI CryoProbe); presaturation (with O1P = 4.701 ppm) was used to suppress the broad, HDO-solvent peak. Samples were prepared in 1 mm tubes containing (16 μL) D₂O containing 0.05% (w/v) TSP-*d*₄, NMR chemical shift reference/internal standard. This compound was purified and analysed with Dr Refaat Hamed.

A substrate containing two, fused 6-membered rings, decanhydroisoquinolone-3-carboxylic acid (H_{10} -IQ3CA) was tested as a substrate of the proline hydroxylase enzymes (*cis*P3H Type I, *cis*P4H, and *trans*P4H); LC/MS-analyses showed that *cis*P3H Type I and *trans*P4H reactions contained single-hydroxylation (200.1 m/z $[M+H]^+$ ions) (Figures 3.71 a and b), which suggest that H_{10} -IQ3CA is a substrate of *cis*P3H Type I and *trans*P4H. These reactions were also scaled up for purification of the ‘major’ products and COSY spectra were measured; however, the assignments have yet to be confirmed with any NOE data (such as a NOESY spectrum), and it is not certain with respect to the reaction stereochemistry or the hydroxylation site.

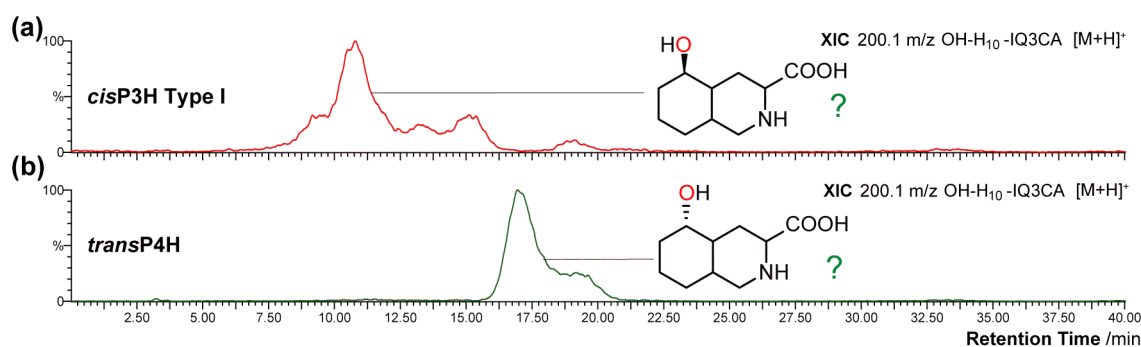


Figure 3.71: LC/MS-based analysis of proline hydroxylase [(a) *cis*P3H Type I and (b) *trans*P4H] reactions with decanhydroisoquinolone-3-carboxylic acid (H_{10} -IQ3CA).

Extracted-ion-chromatograms (XICs) showing 200.1 m/z $[M+H]^+$ ions contained a major peak, thought to correspond to single hydroxylation products of H_8 -L-I2CA.

PH reactions used 15 μ M *cis*P3H Type I/*cis*P4H/*trans*P4H, 500 μ M $(NH_4)_2Fe(SO_4)_2$, 10 mM 2-OG, 1 mM (2*S*,3*aR*,6*aR*)- H_8 -cPP2CA. 100 μ L reactions were carried out in at room temperature (25 °C) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 Å, 5 μ m) [4.6 x 250mm]. These results were measured alongside Dr Refaat Hamed.

3.4.4.5.1 Substrate Turnover Summary: Bicyclic Proline Analogues

The following table (Table 3.13) summarises the proline hydroxylase activities with the bicyclic-substrates screened for proline hydroxylase substrate activity in this section. See below for details of the reaction yields and products.

CHAPTER 3. THE L-PROLINE HYDROXYLASES (PHS)

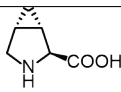
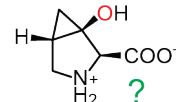
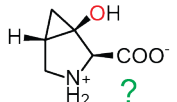
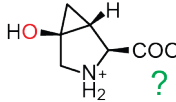
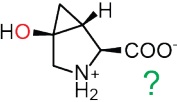
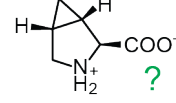
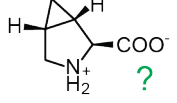
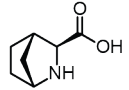
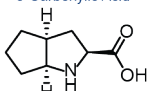
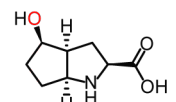
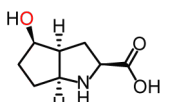
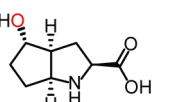
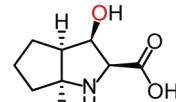
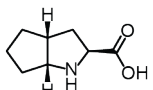
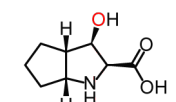
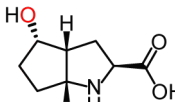
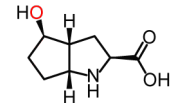
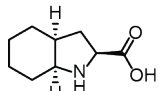
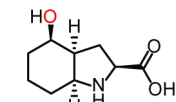
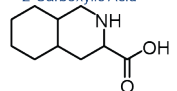
Substrate	Product(s) [+ Yield(s)]			
	<i>cis</i> P3H Type I	<i>cis</i> P4H	<i>trans</i> P4H	
 (1S,2S,5R)-3,4-CH ₂ -L-Pro (1S,2S,5R)-3-Azabicyclo[3.1.0]hexane-2-Carboxylic Acid	 ? 0.25%	 ? 0.35%	-	0%
	 ? 0%	 ? 0%		
	 ? 0%	 ? 0%		
 (1R,3S,4S)-A[2.2.1]H ₃ CA (1R,3S,4S)-2-Azabicyclo[2.2.1]heptane-3-Carboxylic Acid	-	0%	-	0%
 (2S,3aS,6aS)-H ₈ -cPP2CA (2S,3aS,6aS)-Octahydrocyclopenta[b]pyrrole-2-Carboxylic Acid	 (2S,3aS,4R,6aS)-4-OH-H ₈ -cPP2CA 17%	 (2S,3aS,4R,6aS)-4-OH-H ₈ -cPP2CA 3.5%	 (2S,3aS,4S,6aS)-4-OH-H ₈ -cPP2CA 1%	
	 (2S,3R,3aR,6aS)-3-OH-H ₈ -cPP2CA 5%			
 (2S,3aR,6aR)-H ₈ -cPP2CA (2S,3aR,6aR)-Octahydrocyclopenta[b]pyrrole-2-Carboxylic Acid	 (2S,3R,3aS,6aR)-3-OH-H ₈ -cPP2CA 25%	-	0%	 (2S,3aR,4S,6aR)-4-OH-H ₈ -cPP2CA 1%
	 (2S,3aR,4R,6aR)-4-OH-H ₈ -cPP2CA 16%			
 H ₈ -L-12CA (2S,3aS,7aS)-Octahydro-1H-Indole-2-Carboxylic Acid	 (2S,3aS,4R,7aS)-4-OH-H ₈ -L-12CA 6%	-	0%	0%
 H ₁₀ -LQ3CA Decanhydroisoquinoline-3-Carboxylic Acid	?	4%	0%	?
		1%		

Table 3.13: Proline hydroxylase turnover of bicyclic proline derivatives (for 14 h reactions at 25 °C).

3.5 Proline Hydroxylase (PH) Crystallography

At the onset of the work detailed in this thesis, the only crystal structures of a proline hydroxylase was solved by Clifton *et al.* (see Section 1.2.2.4);²⁰⁵ this structure was for a proline *cis*-3-hydroxylase enzyme, *cis*P3H Type II (from *Streptomyces* *sp.* strain TH1). Structures for *cis*P3H Type II were reported in its apo-form (2.3 Å resolution) and its iron(II)-bound form (2.4 Å resolution). Hence, although not a major priority of my work, one aim was to attempt to solve crystal structures for other proline hydroxylases, which had been produced in recombinant form. If possible, it was desirable to obtain structures with substrates and cosubstrates bound. It was hoped that PH crystal structures would likely inform on the structural basis for the stereoselective hydroxylation catalysed by each PH enzyme, and may therefore aid protein engineering studies.

3.5.1 PH Crystallisation Screening

Samples of purified recombinant proline hydroxylases (*cis*P3H Type I, *cis*P4H, and *trans*P4H) were prepared with purities of >90% (as judged by SDS-PAGE analysis) (at least 95% for *cis*P4H and *trans*P4H), and were used to screen a number of protein crystallisation conditions using a sitting-drop, vapor-diffusion crystallisation method.^{293, 382, 383} Initially, commercial screens (from Hampton Research and Molecular Dimensions) were used in attempts to generate protein crystallisation ‘hits’. Where ‘hits’ were obtained, crystallisation conditions were then optimised for wells containing crystals and microcrystals, in order to improve crystal quality (and therefore X-ray diffraction data quality - e.g. increased diffraction spot intensity and increased X-ray diffraction data resolution).

As of writing this thesis, no crystallisation ‘hits’ were found for the *cis*P3H Type I or *trans*P4H proteins. It is possible that the *cis*P3H Type I protein purity (and sam-

ple homogeneity) was insufficient to produce X-ray-diffracting *cis*P3H Type I crystals. Although, highly pure *trans*P4H (>95%) protein was produced (using a 3 column method: nickel-affinity, DEAE Sepharose FF, and Superdex 75 size exclusion), crystals were not obtained. In the case of *cis*P4H, however, protein crystallisation was observed in a number of wells with similar conditions. A number of these conditions were then optimised, in attempts to improve crystal size and quality (for X-ray diffraction). The remaining parts of this chapter focus on *cis*P4H protein crystallography.

3.5.1.1 *cis*P4H Crystallography

The initial screening conditions used for crystallisation of *Sinorhizobium meliloti* *cis*P4H were based, in part, on those used previously for *cis*P3H Type II²⁰⁵ and other 2-OG oxygenases.^{13,384,385} Prior to screening, 15 mg/mL *cis*P4H solutions were prepared by diluting concentrated stocks with 50 mM MES-NaOH, pH 6.5 buffer containing NiSO₄ (in place of redox-unstable ferrous iron), *N*-oxalylglycine (*N*-OG - a 2-OG structural analogue), and proline; the final concentrations of the substrates and cofactors were 1 mM Ni²⁺, 2 mM *N*-OG, and 2 mM L-Pro. To maximise the probability of finding a crystallisation ‘hit’, a number of different 96-well commercial crystallisation screens (Hampton Research: PEG/Ion HT, Crystal Screen HT, Index; Molecular Dimensions: JCSG+, PACT Premier, MIDAS) were used in crystallisation trials; plates were incubated at both 4 °C and 20 °C. 96-Well crystallisation plates (Figure 3.72) were prepared using a crystallisation robot (drops were typically 300 nL total); for each condition well, crystallisation drops were also mixed in 3 different volume-ratios of protein solution and precipitant solution (2:1, 1:1, and 1:2) to increase the breadth of conditions tested (and improve the likelihood of protein crystallisation).

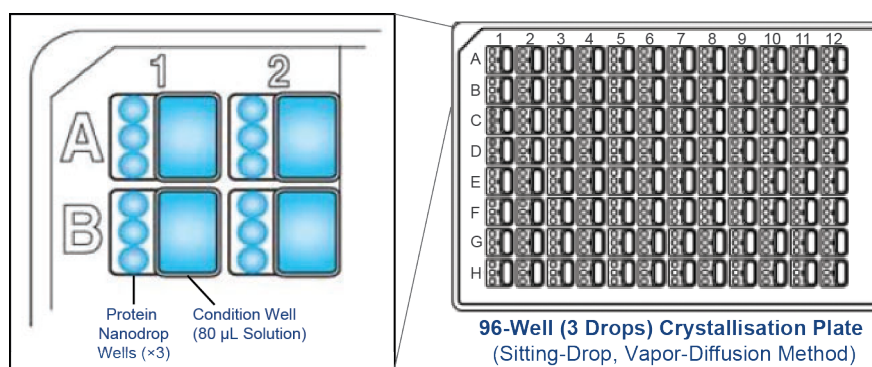


Figure 3.72: 3-Drop, 96-well crystallisation plate: Intelli-Plate 96-3 (Art Robbins Instruments).

The plates facilitate sitting-drop, vapor-diffusion crystallisation experiments. The condition well is typically filled with $\sim 80 \mu\text{L}$ precipitant solution, and the three drop wells are filled with 300 nL protein crystallisation drops (*i.e.* protein solution + precipitant) with various drop ratios (typically 2:1, 1:1, and 1:2). Plates are sealed with optically clear plate seals, which initiates equilibration between the drops and the condition well; the plates are also optically clear, allowing the plates to be inspected for crystals (using microscopy) and imaged. .

After a day, the *cisP4H* plates were inspected for crystals and wells A1 and A2 of plates using a Hampton Research PEG/Ion HT screen incubated at 20°C [conditions A1 and A2 contain 20% (w/v) polyethylene glycol 3350, and 0.2 M sodium fluoride (pH 7.3) and 0.2 M potassium fluoride (pH 7.3), respectively]. Interestingly, in both cases, the crystallisation occurred in protein solution drops, which did not mix with precipitant solution (owing to an equipment malfunction). Unfortunately, the crystals produced without protein-precipitant drop mixing tended to stick to the plate, and were therefore difficult to harvest for X-ray diffraction experiments; these crystals diffracted too poorly to be useful. Methods were then optimised in an attempt to obtain suitable crystals. A Hampton Research PEG/Ion HT screening plate (incubated at 4°C) gave a 'star-shaped' formation of crystal needles in well F2 [20% (w/v) Polyethylene glycol 3350, 8% (v/v) Tacsimate, pH 6.0], drop 2 (1:1 drop ratio). A crystal with the approximate dimensions of $100 \mu\text{M} \times 30 \mu\text{M} \times 30 \mu\text{M}$ was gently separated crystals with a crystal loop and flash frozen with liquid nitrogen. X-ray carried out at the Diamond Light Source Synchrotron (Didcot, Oxfordshire,

UK) showed that the *cis*P4H crystal diffracted to 2.65 Å; indexing showed that the space group of the *cis*P4H crystals was $P2_12_12_1$. Using *cis*P3H Type II structures [apo: PDB 1E5R; Fe(II)-bound: PDB 1E5S] to generate X-ray diffraction phases, molecular replacement was used to generate a model *cis*P4H structure solution; this structure solution indicated that *cis*P4H exists as a tetramer in this $P2_12_12_1$ crystal form (Figure 3.73). The crystal harvesting was carried out by Dr Rasheduzzaman Chowdhury, who also helped collect the X-ray diffraction data set and generate the initial crystal structure solution.

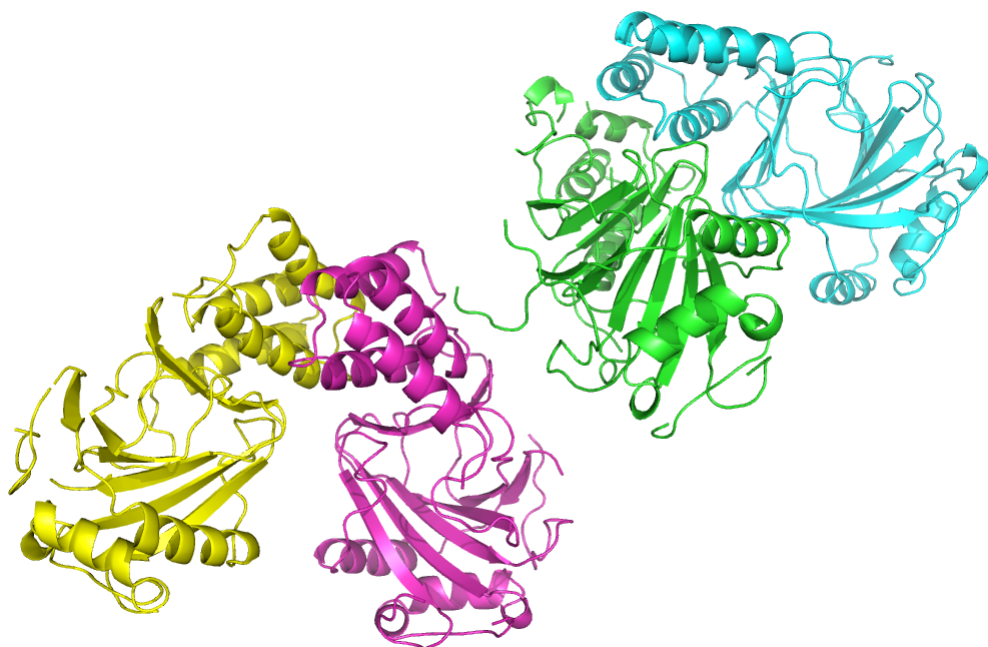


Figure 3.73: *cis*P4H is tetramer in the unit cell of $P2_12_12_1$ protein crystals.

A partially refined *cis*P4H structure indicated that the secondary structure of *cis*P4H is mainly composed of at least 8 α -helices, 10 β -strands, and 18 loops/turns (Figure 3.74), as for other 2-OG oxygenases. Eight of the β -strands fold to form the conserved, distorted double-stranded β -helix (DSBH) fold (‘jelly roll’ motif), which is characteristic of the 2-OG oxygenases.¹³ The active site is located at the more open end of the DSBH, which supports the iron(II)- and 2-OG-binding residues, in a manner very similar to those predicted for other 2-OG oxygenases.

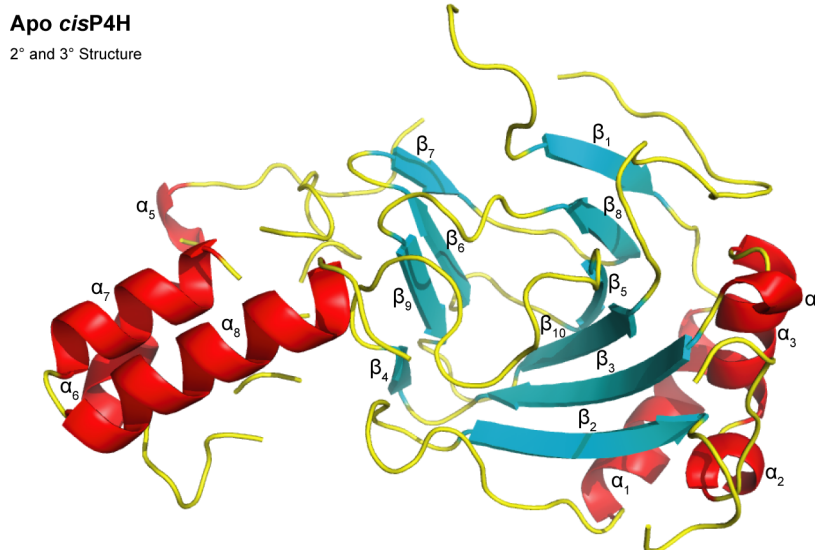


Figure 3.74: ‘Cartoon’ depiction of the secondary and tertiary structure of apo-*cis*P4H monomer units.

*cis*P4H is mainly composed of at least 8 α -helices, 10 β -strands, and 18 loops/turns, and 8 β -strands fold to form the conserved, distorted double-stranded β -helix (DSBH) fold (‘jelly roll’ motif).

Calculations (using PyMOL modelling software) showed that the root-mean-square deviation (rmsd) of *cis*P4H-backbone-chain atoms compared with *cis*P3H Type II atoms is 1.431 Å (Figure 3.75); this low rmsd value suggests the two proline hydroxylase enzymes are similar in structure. They are consistent with the successful molecular-replacement solution using the *cis*P3H Type II structure (PDB accession number 1E5S), despite there only being ~32% sequence identity between the two PH proteins. Inspection of the aligned *cis*P3H Type II and *cis*P4H structures (‘by eye’) shows that the two proteins share the same ‘major’ α -helices and β -strands (in similar positions), and differ mainly in the number and positioning of loops/turns; in the partially refined *cis*P4H structure, it is likely that amino acid residues N219-F233 form an α -helix analogous to that seen in the *cis*P3H Type II structure (residues N226-K240), since these amino acids occupy very similar positions as the corresponding residues in *cis*P3H Type II, and form a (currently) ‘broken up’ helical chain.

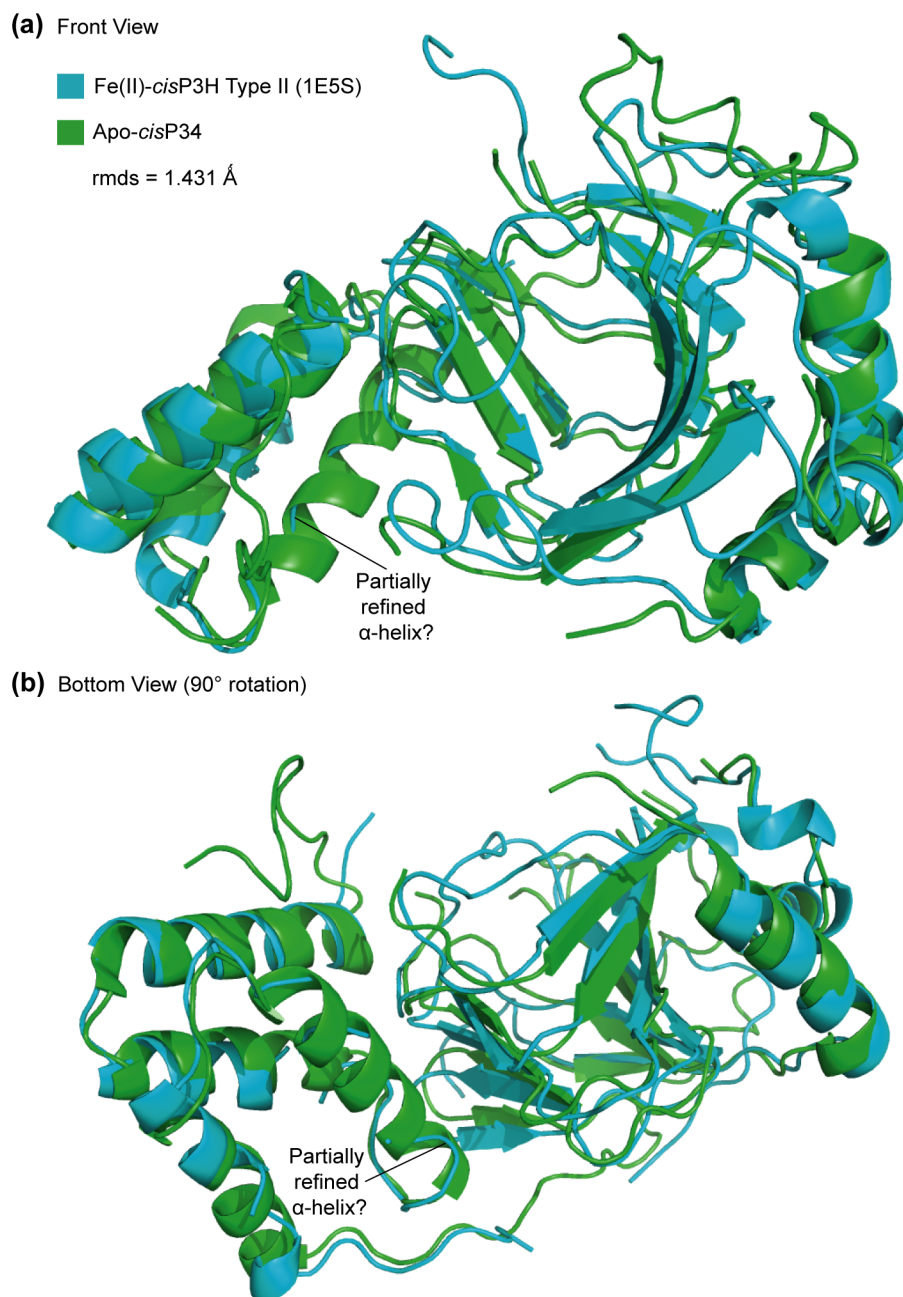
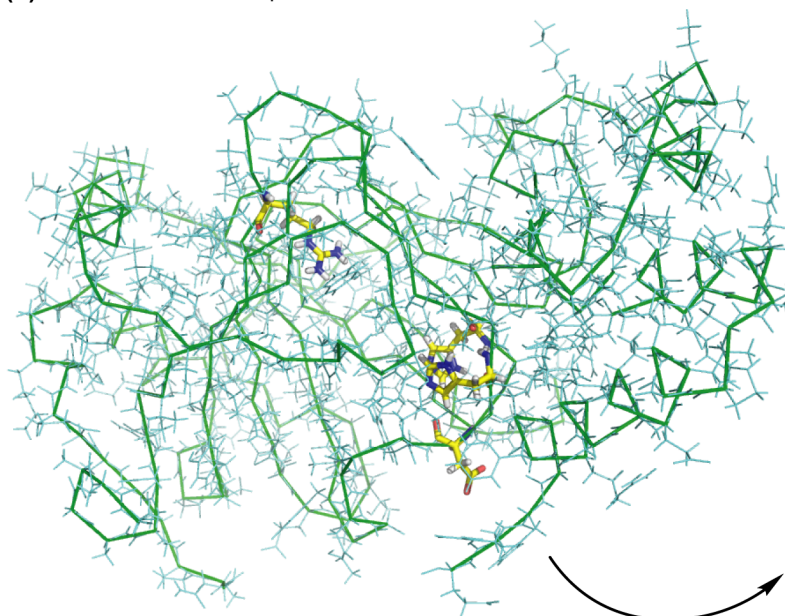


Figure 3.75: Molecular alignment of *cis*P3H Type II (Fe(II)-bound form - PDB accession number 1E5S) and *cis*P4H (apo) crystal structures.

The two proline hydroxylase structures show clear similarities, sharing the same ‘major’ α -helices and β -strands (in similar positions), and differ mainly in the number and positioning of loops/turns; it is likely that *cis*P4H amino acid residues N219-F233 form an α -helix analogous to that seen in the *cis*P3H Type II structure (residues N226-K240).

Although the *Sinorhizobium meliloti* *cis*P4H crystallisation solutions contained NiSO_4 , *N*-oxoglutarate (*N*-OG), and L-proline (1 mM, 1 mM, and 2 mM, respectively), before being mixed with ‘precipitant’ solution, the electron-density maps for the *cis*P4H structure showed that the crystallised protein was in the apo-form of *cis*P4H (Figure 3.76 a). Presumably, the binding of the cofactors and substrates to the *cis*P4H protein is too weak under the crystallisation conditions used for them to remain present in the protein crystals. Interestingly, in this *cis*P4H crystal structure, the carboxymethyl sidechain of the iron(II)-binding aspartate residue (D108) of the active site is observed to ‘point away’ from the usual position in which it chelates Fe(II) in many 2-OG oxygenases (*i.e.* bound to $2 \times \text{His}$ & $1 \times \text{Asp/Glu}$ residues) (Figure 3.76 b).

(a) *Sinorhizobium meliloti* Apo-*cis*P4H



(b) Active Site Residues

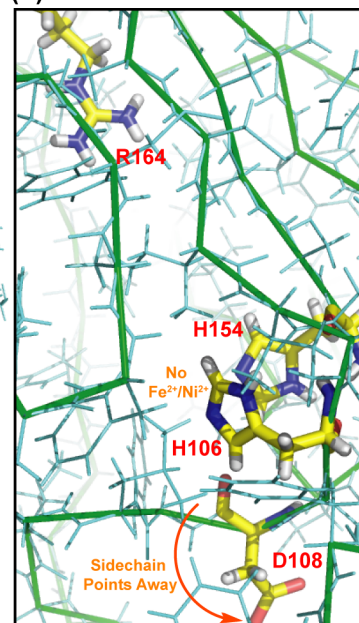


Figure 3.76: Active site residues of the apo-*cis*P4H crystal structure.

(a) The ‘lines’ and ‘ribbon’ model of the apo-*cis*P4H structure.

(b) The iron(II)-binding, active-site residues of *cis*P4H (a $2 \times \text{His}$ - $1 \times \text{Asp}$ motif), where the D108 residue is pointing away from the iron(II)-binding site; electron-density maps show ‘zero-occupancy’ for the enzyme cofactors and substrates, showing that they were absent from these *cis*P4H crystals.

3.5.1.2 *cis*P4H D108 Variants

Given that *cis*P4H crystallised with the D108 sidechain pointing away from its site of Fe(II) ligation, the role of D108 in the mechanism of *cis*P4H-mediated proline hydroxylation was of interest. To investigate the effect of the D108 residue on *cis*P4H activity, D108-variant proteins were subsequently produced (Figure 3.77); D108 was placed with Gly, Ala, Leu, Thr, Glu, and Pro.

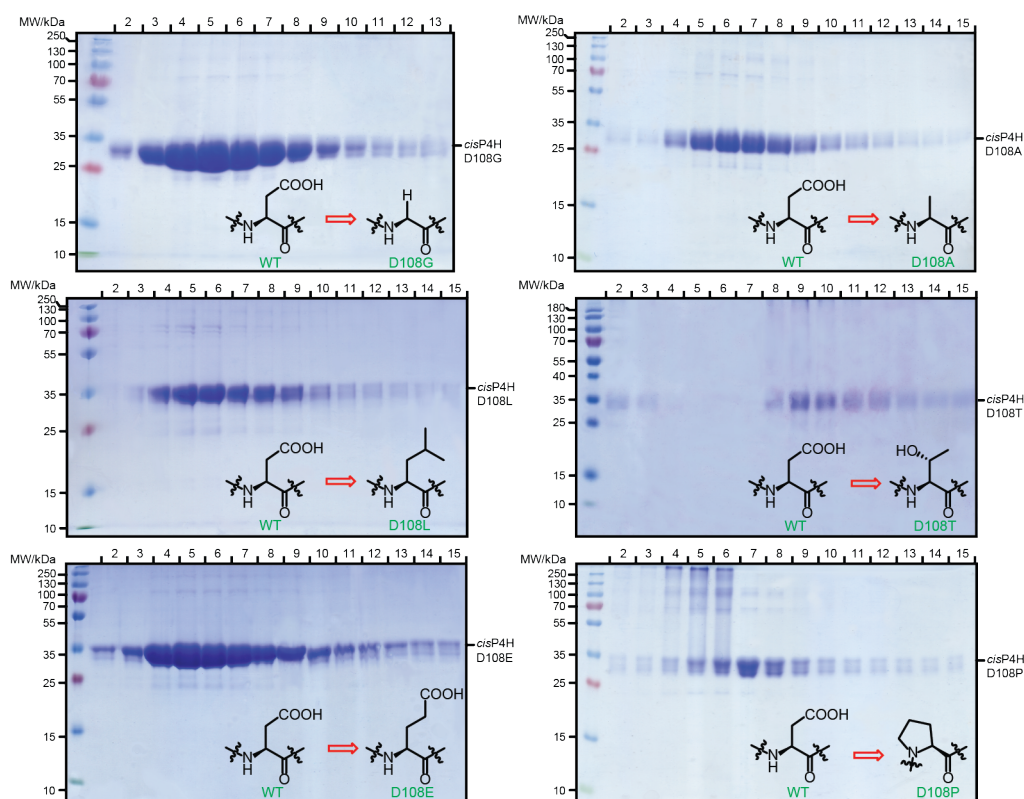


Figure 3.77: SDS-PAGE analysis of nickel-affinity purifications of D108 *cis*P4H variants: six mutations in the *Sinorhizobium meliloti cis*P4H gene to produce D108G/A/L/T/E/P variants.

The Gly, Ala, Leu, and Thr D108 variants were all inactive in LC/MS-assays (*i.e.* <5% activity). In contrast, the D108E and D108P variants were found to be active by LC/MS assay (Figure 3.78). Preliminary activity assays using end-point time points indicated the D108E variant has a level of activity to the wildtype (*i.e.* complete conversion of proline was observed) (Figure 3.78 a), and timecourse

assays were then used to confirm these observations (Figure 3.79 a); this is not entirely surprising, given the structural similarities of the Asp and Glu residues (both have terminal-carboxylic-acid sidechains). The D108P variant was also active (Figure 3.78 b), but at a much reduced level (~5% of the relative activity of the wild type enzyme) (Figure 3.79 b). This is surprising given that proline and aspartate are very different residues. In the case of the other 2-OG oxygenases, variable results for structure of the Asp of the 'HX(D/E)' motif have been obtained. For example, in the case of an asparaginyl hydroxylase, factor inhibiting hypoxia-inducible factor (FIH), the D201A (Asp to Ala) and D201E (Asp to Glu) variants were found to be inactive,³⁸⁶ but the D201G variant was active.³⁸⁶ To my knowledge, there are no previous reports of an analogous Glu-to-Pro ODD variant being active.

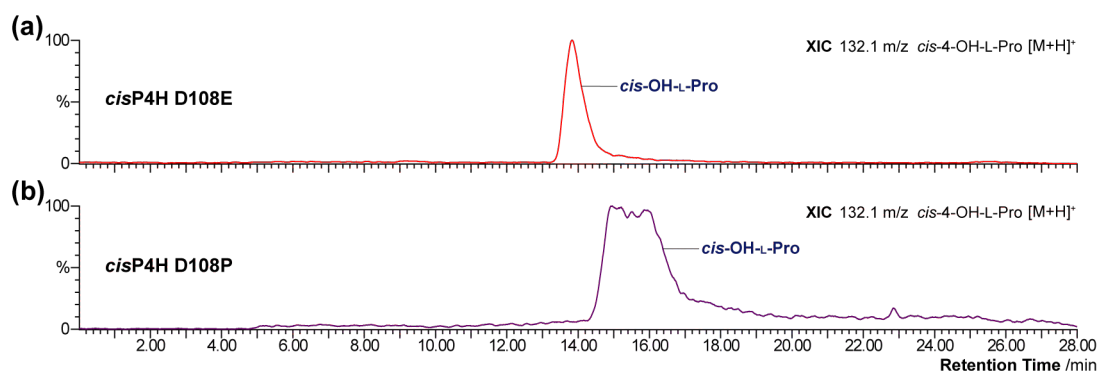
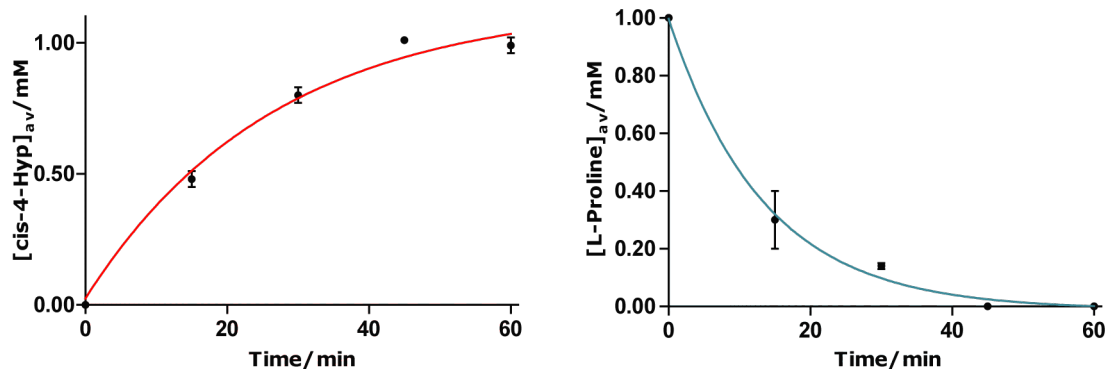
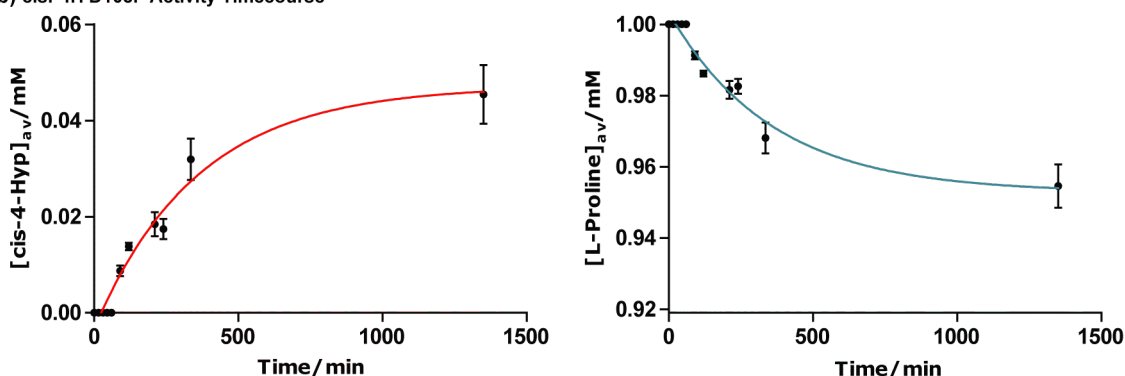


Figure 3.78: LC/MS-based analysis of *cisP4H* variant reactions [(a) *cisP3H* Type I and (b) *transP4H*].

Extracted-ion-chromatograms (XICs) showing 132.1 m/z $[M+H]^+$ ions contained a major peak, corresponding to single hydroxylation products.

PH reactions used 15 μ M *cisP4H*-D108E/D108P, 500 μ M $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 10 mM 2-OG, 1 mM L-Pro. 100 μ L reactions were carried out in at room temperature (25 $^\circ\text{C}$) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 $\text{\AA}$, 5 μ m) [4.6 x 250mm].

(a) *cis*P4H D108E Activity Timecourse(b) *cis*P4H D108P Activity TimecourseFigure 3.79: *cis*P4H variant timecourse reactions [(a) D108E (b) D108P].

PH reactions used 15 μ M *cis*P4H-D108E/D108P, 500 μ M $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 10 mM 2-OG, 1 mM L-Pro. 100 μ L reactions were carried out in at room temperature (25 $^\circ\text{C}$) for 14 h in 50 mM MES-NaOH, pH 6.5 buffer.

Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 $\text{\AA}$, 5 μm) [4.6 x 250mm].

3.5.1.3 Summary of PH Crystallographic Work

Although the crystallographic analyses on *Sinorhizobium meliloti cis*P4H should be treated as preliminary, methods for the reliable crystallisation of the protein have been developed. Future work can now focus on obtaining a structure of complexes with substrates and substrate analogues, with a view to enabling protein engineering.

The fact that the D residue of the iron(II)-binding, 'HXD' motif can be replaced with E (Glu) or P (Pro) and still retain *cis*P4H activity is interesting, particularly

in the latter case. One possible line of investigation is to test if altered iron(II) binding alters the substrate selectivity. In this respect, it's notable that translocation of bound iron(II) is possible for most of the 2-OG oxygenases (being used previously as a method for inhibition),³⁸⁷ and, thus, use of proline hydroxylases with only 2 iron ligands may enable expanded substrate selectivity.

3.6 Conclusions

Throughout this project, the proline hydroxylase enzymes have been shown to accept a wide range of substrates; the PH enzymes can hydroxylate proline derivatives with smaller substituents, such as methyl, fluoro, and hydroxyl groups, as well as substrates with relatively large sidechains, such 2-propynyl groups and carboxymethyl substituents. In addition, the proline hydroxylases were found to be good catalysts for hydroxylation of substrates of varying ring sizes (4-7 carbons) (*i.e.* L-Aze, L-Pip, and L-Azp), as well as bicyclic proline analogues with fused rings, which can be hydroxylated on the 'non-proline' ring. As well as single hydroxylation products, a number of dihydroxylation products have been indicated by my work, and interesting product mixtures can, in some cases, be observed. The most catalytically diverse enzyme was *cis*P3H Type I, which accepted the broadest range of substrates, but still maintained high stereoselectivity, whereas the least 'promiscuous' enzyme was *trans*P4H, which accepted the fewest range of substrates, but was also very stereoselective; *cis*P4H also accepted many of the same substrates as *cis*P3H Type I, albeit often with a lower activity, and often producing a number of products (by hydroxylating both C3- and C4-positions of the proline ring, or the analogous positions in other, similar substrates).

A number of products have been characterised by mass spectrometry and NMR, and it is likely that many others can also be similarly characterised with further

optimisation of HPLC purification methods (for highly polar products) and improved turnover yields; future work would pursue full product characterisation experiments to verify our hypotheses and increase our understanding of the proline hydroxylase substrate selectivity.

It would be interesting to see if many of these substrate analogue proline hydroxylase reactions work on a larger scale in both *in vivo* and *in vitro*, as these enzyme could be tools to produce compounds useful, for example, in pharmaceutical/agricultural production. In particular, given its high activity with a broad spectrum of interesting proline analogues, *cis*P3H Type I (or its isozyme “Type II”) may be useful in larger scale reactions (such as fermenter cell cultures) for producing chemicals of biological and pharmaceutical interest.

Preliminary crystallographic investigations on *cis*P4H was carried out, and was used as the basis for mutagenesis studies. One interesting aspect of the results is that *cis*P4H can catalyse hydroxylation with only two iron(II)-binding residues (*i.e.* H106 and H154 in the D108P variant); whether this may enable expandable substrate selectivity is a possible line of investigation.

Chapter 4

Ethylene Biosynthesis

4.1 Industrial Ethylene Production

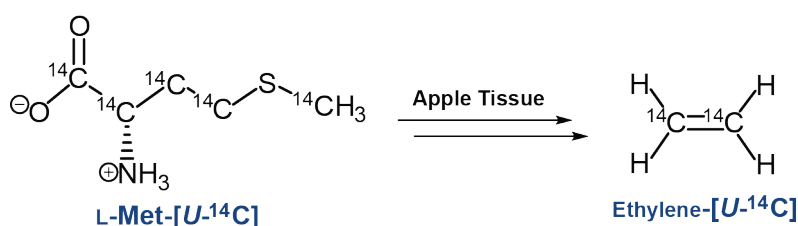
As detailed in Chapter 1, ethylene is the most produced organic compound,²¹³ and is a common industrial feedstock.^{214–216} The work in this chapter was aimed at contributing to the development of a biological role for ethylene production.

4.2 Pathways of Ethylene Biosynthesis

Ethylene is produced biologically by higher plants, acting as a plant hormone that modulates growth and development; ethylene stimulates a defensive response to abiotic factors or stress (such as pathogenic attack or wounding).^{217–219} A variety of microbes, including bacteria and fungi, also produce ethylene enzymatically. The discovery that microbes make ethylene is linked to the effect of ethylene on plant growth and development (though their ethylene production).^{235–241} This following section provides an outline of the biochemical pathways used for ethylene production.

4.2.1 Plant Metabolism of L-Methionine to Ethylene

Plants produce ethylene gas from L-methionine in the presence of air.²²⁰ A notable, early example studied was from apple tissue,^{220,227,388–390} and other examples in other fruits and flowers have been investigated (e.g. tomato,^{391–393} avocado,^{394,395} melon,^{396,397} and carnation flower³⁹⁸). In order to analyse the course of this ethylene-forming pathway, apple tissue samples were fed different L-methionine samples labelled with radioactive isotopes in a number of positions. Apple tissue fed with L-[U-¹⁴C]-methionine efficiently metabolised the [¹⁴C]-labelled L-methionine to produce [¹⁴C]-labelled ethylene in the presence of air (Scheme 4.1).²²⁰



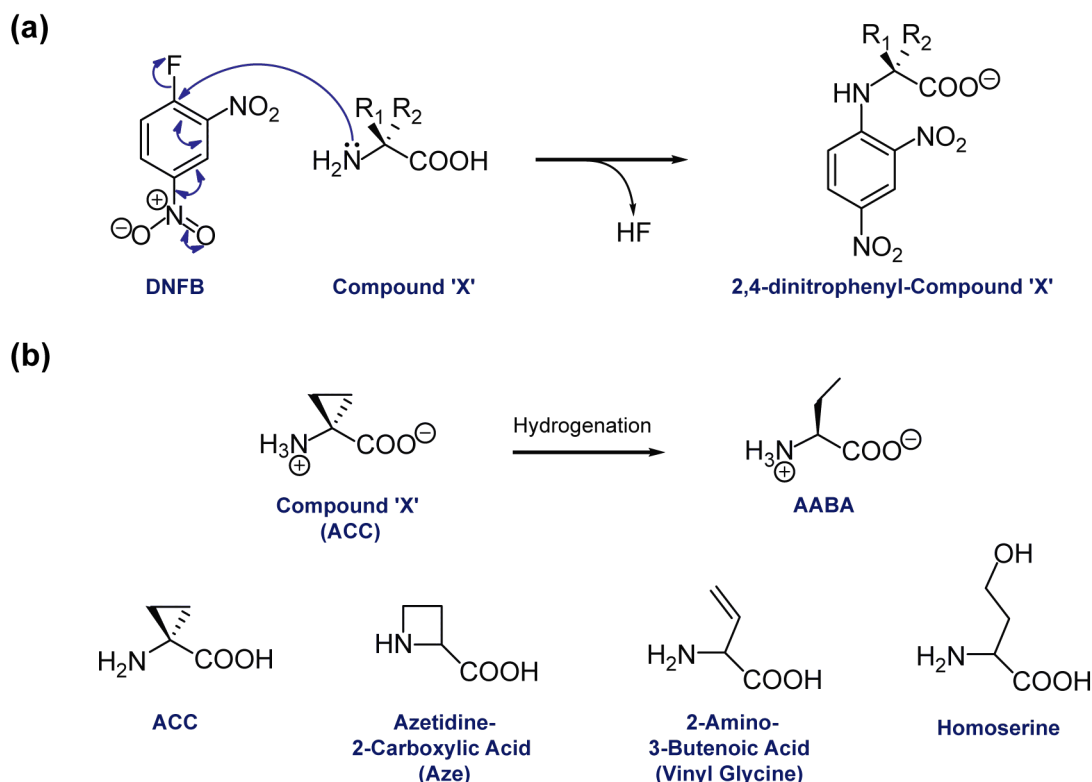
Scheme 4.1: Apple tissue metabolises L-methionine to ethylene: L-methionine-[U-¹⁴C] was shown to be converted to ¹⁴C-labelled ethylene.²²⁰

Under a nitrogen atmosphere, it was found that apple tissue converts L-methionine to an intermediate (compound ‘X’), which was not fully converted to ethylene.²²⁰ The intermediate (‘X’) was found to be cationic on the basis of electrophoresis experiments at pH 2.2, neutral at pH 6.8, and anionic at pH 11.0; the intermediate (‘X’) reacts with 2,4-dinitrofluorobenzene (DNFB - ‘Sanger’s Reagent’) to produce a derivative that migrated as an anion at pH 6.8 (Scheme 4.2 a).²²⁰ These observations supported the hypothesis that intermediate ‘X’ was an amino acid. The hydrogenation product of intermediate ‘X’ was found to be α -aminobutyric acid (AABA) using data gathered from paper chromatography, paper electrophoresis, and co-crystallisation (with authentic AABA to constant specific radioactivity) experiments (Scheme 4.2 b). The mobility of compound ‘X’ in paper chromatography

and electrophoresis experiments was then compared with that of authentic samples of desaturated forms of AABA, namely 1-aminocyclopropane-1-carboxylic acid (ACC), azetidine-2-carboxylic acid (Aze), 2-amino-3-butenic acid (vinyl glycine), and homoserine (Scheme 4.2 b); only ACC was found to comigrate with compound 'X' (at pH 2.2 and 10.8).²²⁰ Hydrogenation of known samples of ACC did, in fact, produce AABA.²²⁰ Compound 'X' also reacted very slowly with ninhydrin, consistent with it containing no α -hydrogen atoms (as is the case for ACC). Further confirmation that compound 'X' was ACC was verified when the radioactive compound 'X' material was cocrystallised with known ACC to constant specific radioactivity.²²⁰

Apple tissues fed with L-methionine-[*methyl*-¹⁴C] and L-methionine-[³²S] under a nitrogen atmosphere metabolised the isotopically labelled L-methionine to produce the corresponding, radioactively labelled methylthioadenosine (MeSAdo) molecules (Scheme 4.3 a);²²⁰ this result suggested that methionine is initially converted to *S*-adenosyl-L-methionine (*S*-Ado-L-Met, SAM), which is then, in turn, fragmented to ACC and methylthioadenosine (MeSAdo),²²⁰ an α,γ -elimination step was subsequently shown to be catalysed by ACC synthase (Scheme 4.3 b).^{221–225} MeSAdo is then readily hydrolysed in apple tissue to produce methylthioribose (MeSRib).

Radioactive [*U*-¹⁴C]-ACC was shown to be converted to [*U*-¹⁴C]-ethylene,²²⁰ a reaction which was later shown to be catalysed by ACC oxidase;^{226,227} ACC oxidase catalyses the oxidative, step-wise²³² fragmentation of ACC to ethylene and cyanide (CN⁻) (Scheme 4.4).^{227,393} Support for the ACC oxidase reaction products was initially provided by experiments studying the metabolism of [¹⁴C]-radiolabelled ACC (*e.g.* ACC-[*carboxyl*-¹⁴C]) in several plant tissues which produce ethylene;³⁹⁹ these radiolabelling experiments showed that the carboxyl carbon atom of ACC was liberated in CO₂, whereas the C-1 of ACC was found in β -cyanoalanine in an amount equivalent to that of the ethylene produced.³⁹⁹ Since HCN was shown

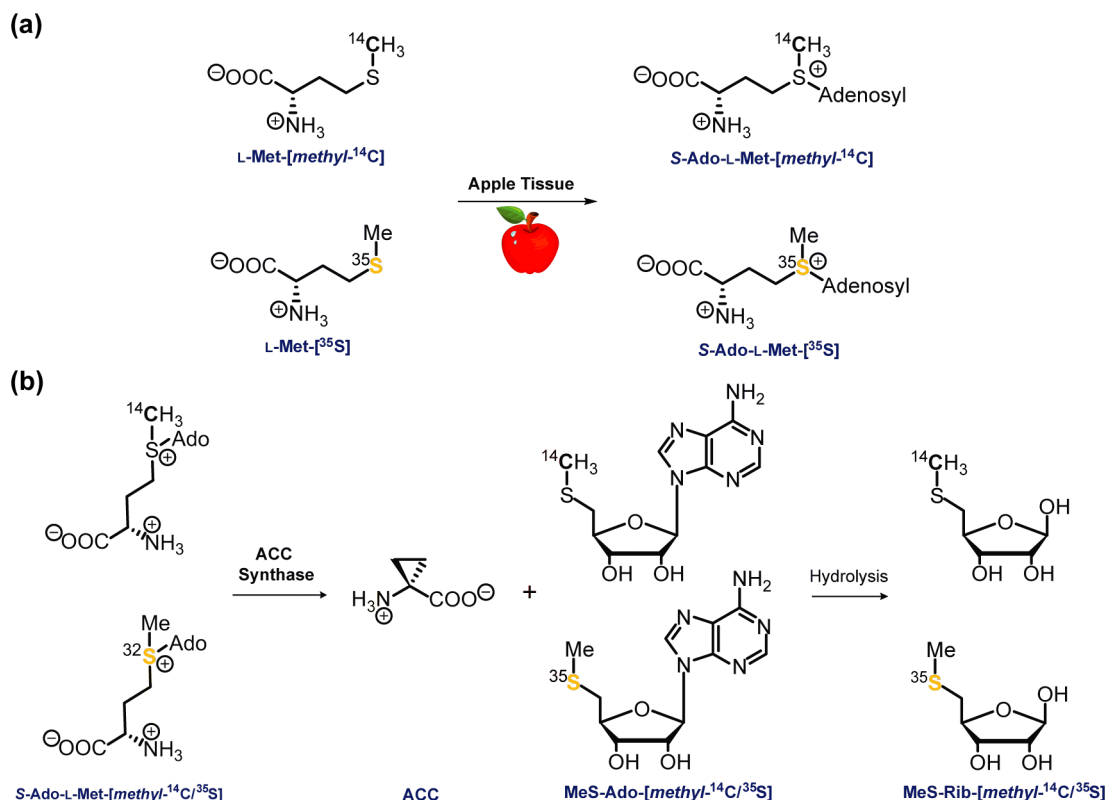


Scheme 4.2: Reactions used to determine the identity of compound 'X', an intermediate produced from L-methionine in the plant biosynthesis of ethylene.

(a) Compound 'X' reacted with dinitrofluorobenzene (DNFB - 'Sanger's Reagent') to produce an anionic substance (at pH 6.8), suggesting compound 'X' was an amino acid.

(b) The hydrogenation of compound 'X' produced α -aminobutyric acid (AABA), consistent with compound 'X' being one of the following: aminocyclopropane-1-carboxylic acid (ACC), azetidine-2-carboxylic acid (Aze), 2-amino-3-butenic acid (vinyl glycine), or homoserine. Compound 'X' was confirmed to be ACC using a number of chemical identification experiments.

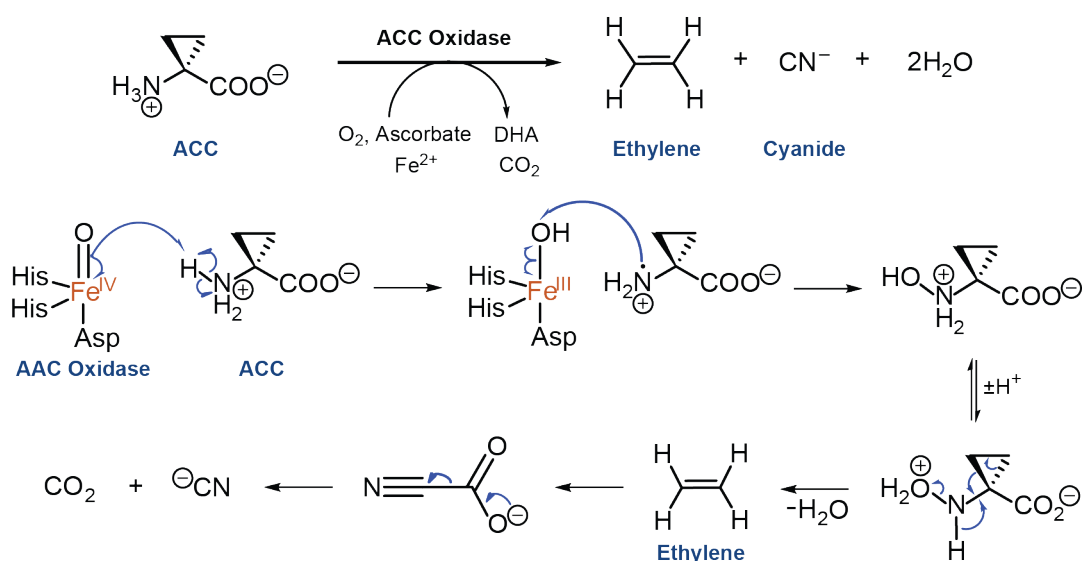
to be metabolised into β -cyanoalanine conjugates in plant tissues,²³³ the presence of β -cyanoalanine in ethylene-producing plant tissues implied that CN^- is the initial product derived from C-1 of ACC during ACC oxidase-mediated ethylene production.³⁹⁹ The ACC oxidase reaction requires molecular oxygen and ferrous iron for catalytic activity (Scheme 4.4);^{227, 388, 389, 393, 395, 397} there is evidence that it uses ascorbate as a co-substrate, as opposed to the 2-OG-dependent oxygenases, where ascorbate is not directly involved in the catalytic cycle. Studies of *in vitro* reactions, using purified ACC oxidase, showed that 1 mol equivalent of O_2 is consumed



Scheme 4.3: Use of isotopic-labelling experiments with radioactive L-methionine to follow the pathway of ethylene biosynthesis in apple tissue; [^{14}C]- and [^{35}S]-labels were used.²²⁰

(a) L-Methionine-(*methyl*- ^{14}C) and L-methionine- ^{35}S are converted by apple tissues to S-adenosyl-L-methionine- ^{14}C and S-adenosyl-L-methionine- ^{35}S , respectively;
 (b) these isotopically labelled S-Ado-L-Met samples were shown to be converted to 1-aminocyclopropane-1-carboxylic acid (ACC) and methylthioadenosine (MeSAdo), in a reaction catalysed by ACC synthase in the apple tissue. The MeSAdo is hydrolysed to give methylthioribose (MeSRib).

for each 1 mol equivalent of ACC turned over by ACC oxidase,²²⁷ and equimolar quantities of dehydroascorbate, ethylene, CN^- , and CO_2 are formed by the reaction (Scheme 4.4);²²⁷ headspace concentrations of ethylene and CO_2 in reaction tubes were determined with a gas chromatograph and an infrared analyzer, respectively.²²⁷ Radioactive $^{14}\text{CO}_2$, produced from [*carboxyl*- ^{14}C]ACC was trapped and quantified using a scintillation counter, whereas O_2 consumption was monitored with an oxygen electrode.²²⁷

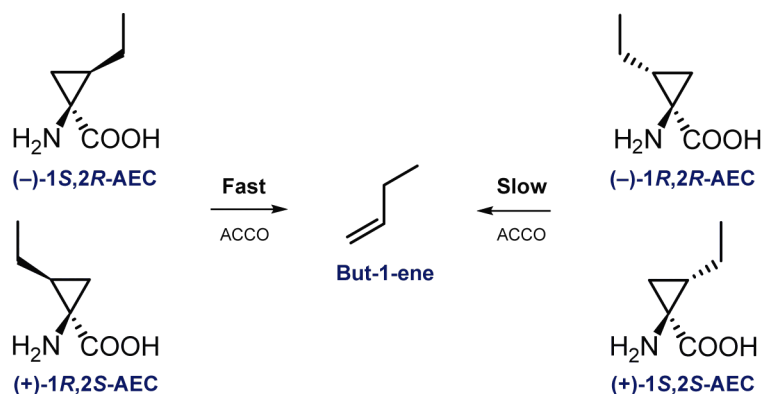


Scheme 4.4: The overall reaction catalysed by ACC oxidase and the outline mechanism of ACC turnover by ACC oxidase.

Hydroxylation of the amino group of ACC produces an unstable, *N*-hydroxyl-ACC intermediate (a nitrenium ion equivalent), which fragments in a step-wise manner step-wise²³² to release ethylene gas and cyanofornate; cyanofornate then spontaneously decarboxylates to release CN⁻ and CO₂.^{227,393}

Interestingly, CO₂ was shown to directly activate catalysis by ACC Oxidase,^{227–230} increasing the activity of apple fruit ACC oxidase by >10-times its ambient air (0.03% CO₂) activity under conditions of 4% CO₂;²²⁷ substitution of added, gas phase CO₂ for HCO₃⁻ in the reaction mixture similarly stimulates an increase in ACC oxidase activity (maximal >10 mM HCO₃⁻²²⁷), probably by increasing the concentration of gaseous CO₂.^{227–230}

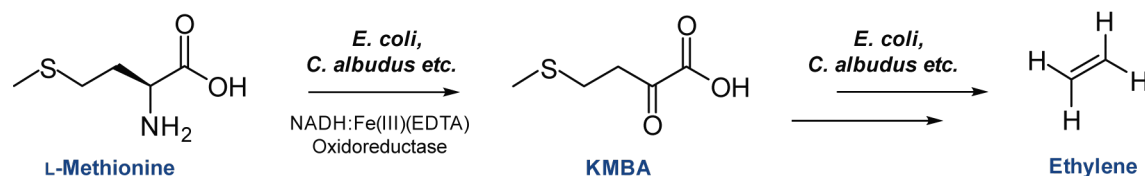
Purified ACC oxidases have been shown to catalyse the turnover various stereoisomers of the ACC-analogue, 1-amino-2-ethylcyclopropane (AEC - ‘2-ethyl-ACC’), to produce but-1-ene by a mechanism equivalent to that used in ethylene production (Scheme 4.5);^{388,389,392,393,395,396} mixtures of (1*R*,2*S*)- and (1*S*,2*R*)-AEC [(±)-allocoronamic acid] were found to be efficiently converted into but-1-ene, whereas mixtures of (1*R*,2*R*)- and (1*S*,2*S*)-AEC [(±)-coronamic acid] were converted much less efficiently to but-1-ene (approximately 21 times slower) (Scheme 4.5).^{388,389,393,395,396}



Scheme 4.5: ACC oxidase preferentially converts mixtures of (1R,2S)- and (1S,2R)-AEC to but-1-ene over converting mixtures (1R,2R)- and (1S,2S)-AEC to but-1-ene.^{227–230}

4.2.2 Microbial Ethylene Biosynthesis

As well as methionine being the precursor of ethylene in all higher plants, some species of fungi and a number of bacteria examined also make use of methionine for ethylene production;⁴⁰⁰ microbially produced ethylene, especially in waterlogged soils, has deleterious effects on plants.⁴⁰¹ In some cases, such as in *Escherichia coli*, *Chromobacterium violaceum*, and *Cryptococcus albidus*, bacterial ethylene biosynthesis can occur via 2-oxo-4-methylthiobutyric acid (KMBA),^{235, 236, 402, 403} a transaminated derivative of methionine (Scheme 4.6); the transamination of methionine to produce KMBA is proposed to be catalysed by an NADH:Fe(III)(EDTA) oxidoreductase, in a reaction which is enhanced by conditions of limiting ammonia (C:N atom ratio = 20:1) (Scheme 4.6).^{235, 236} However, there is now evidence that the transformation of KMBA to ethylene can be non-enzymatically directed using hydrogen peroxide. Some think that KMBA formation may be a method of recovering amino nitrogen from methionine, resulting in the spontaneous generation of KMBA-derived ethylene.^{235, 236}



Scheme 4.6: The transformation of 2-keto-4-methylthiobutyric acid (KMBA) to ethylene by *Escherichia coli*, *Cryptococcus albidus*, and *Chromobacterium violaceum* etc.; KMBA is produced from L-methionine in NADH:Fe(III)(EDTA) oxidoreductase-mediated, transamination reaction.

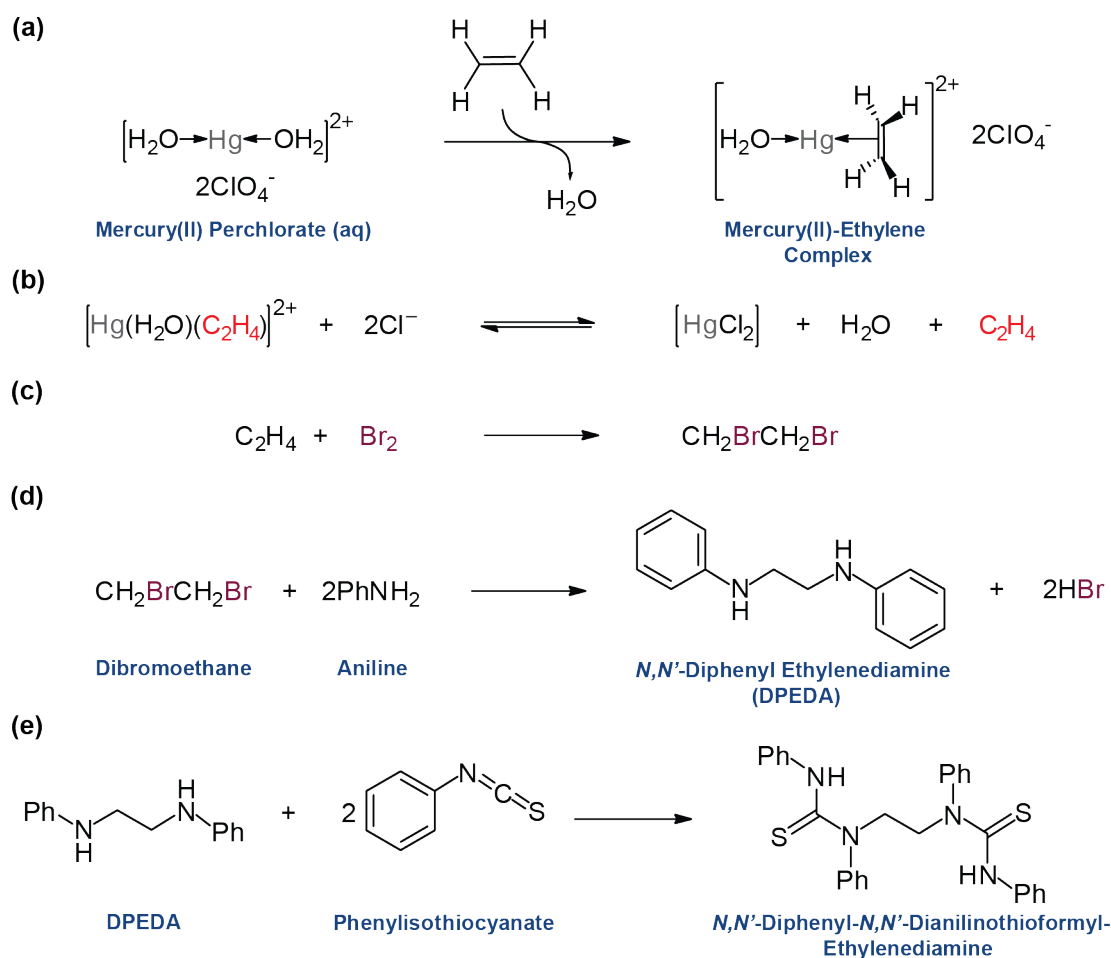
4.2.2.1 Ethylene Biosynthesis from 2-OG

An early study found that emanations from fruits (e.g. lemons) inoculated with *Penicillium digitatum* ('green mould') showed accelerated CO₂ release and colour development over control plants, suggesting a plant growth hormone (ethylene) was being produced by the *P. digitatum*.²³⁷ The increase in CO₂ evolution by fruit subjected to 'green mould' vapours (over the control fruit) varied from 50% to 100%, a significant change.²³⁷ The CO₂ production was quantified by absorbing the evolved CO₂ in a solution of barium hydroxide (forming a barium carbonate precipitate).²³⁷ In another early study, Florida-grown oranges, tangerines, limes, lemons, and grapefruit produced gases, which induced epinasty in potted tomato/potato/sunflower plants and potato cuttings.²³⁸ Epinasty is characterised by an increase in growth of the upper surface of an organ over of the lower side (especially leaves), causing it to bend downward.^{404,405} Oranges and grapefruit inoculated with *P. digitatum* also showed signs of epinasty, but more quickly than without *P. digitatum* inoculation, implying that *P. digitatum* produces ethylene gas;²³⁸ positive epinasty results were also obtained when using pure cultures of *P. digitatum*, further validating the hypothesis for *P. digitatum*-based ethylene biosynthesis.²³⁸ Later, ethylene was chemically identified as the plant growth hormone produced by *P. digitatum*;²³⁹ a solution of mercuric perchlorate [Hg(ClO₄)₂] was used to remove ethylene from the stream of air being passed over a series of cultures (by forming a

$[\text{Hg}(\text{C}_2\text{H}_4)(\text{H}_2\text{O})]^{2+}$ complex-ion).²³⁹ The accumulated ethylene was subsequently released quantitatively into a reaction flask by adding of hydrochloric acid to the ethylene-mercury complex solution (a ligand replacement reaction, where Cl^- ions replace $\text{H}_2\text{O}/\text{C}_2\text{H}_4$ ligands) (Scheme 4.7 b); the ethylene was brominated in a reaction flask, forming 1,2-dibromoethane (Scheme 4.7 b).²³⁹ The 1,2-dibromoethane was reacted with aniline to form *N,N'*-diphenyl ethylenediamine (DPEDA) (Scheme 4.7 c), which was then reacted with phenyl isothiocyanate (PITC) to produce *N,N'*-diphenyl-*N,N'*-dianilinothioformyl-ethylenediamine (Scheme 4.7 d);²³⁹ the melting points of both ethylene-derived DPEDA and *N,N'*-diphenyl-*N,N'*-dianilinothioformyl-ethylenediamine matched those measured for known samples of these compounds, confirming that the gas isolated from *P. digitatum*.²³⁹

Jacobsen *et al.* proposed that the origin of the *P. digitatum*-produced ethylene carbon skeleton is intimately associated with Krebs cycle [or citric acid/tricarboxylic acid (TCA) cycle⁴⁰⁶] acids.²⁴⁰ [^{14}C]-Labelled Krebs cycle intermediates (e.g. fumarate-[1,4- ^{14}C]/[2,3- ^{14}C], succinate-[1,4- ^{14}C]/[2,3- ^{14}C], DL-malate-[3- ^{14}C], sodium acetate-[1- ^{14}C]/[2- ^{14}C]) were introduced to *P. digitatum* growth media; the [^{14}C]-labels were found in the ethylene gas produced,²⁴⁰ supporting the involvement of a Krebs cycle intermediate in the *P. digitatum* ethylene-forming pathway. Certain carbon atoms of other, similar compounds studied (e.g. L-methionine-[U - ^{14}C]/[methyl- ^{14}C], DL-methionine-[1- ^{14}C]/[2- ^{14}C], β -alanine-[2- ^{14}C], sodium propionate-[1- ^{14}C]/[2- ^{14}C]/[3- ^{14}C]) were also incorporated into the carbon skeleton of ethylene (probably *via* the Krebs cycle);²⁴⁰ carbon atoms of acrylic acid (potassium acrylate-[1- ^{14}C]/[2- ^{14}C]/[3- ^{14}C] were fed to *P. digitatum*), particularly C-2, were also found to be incorporated in the ethylene skeleton.²⁴⁰ A novel ethylene-forming enzyme (EFE) from *Pseudomonas syringae* pv. *phaseolicola* PK2 was then shown to catalyse the formation of ethylene from 2-oxoglutarate (2-OG) (Scheme 4.8),²⁴¹ a known Krebs cycle intermediate; this confirmed the hypothesis Jacobsen *et al.* put forward, which

suggested that the carbon backbone of ethylene from EFEs originates from a Krebs cycle intermediate.²⁴⁰

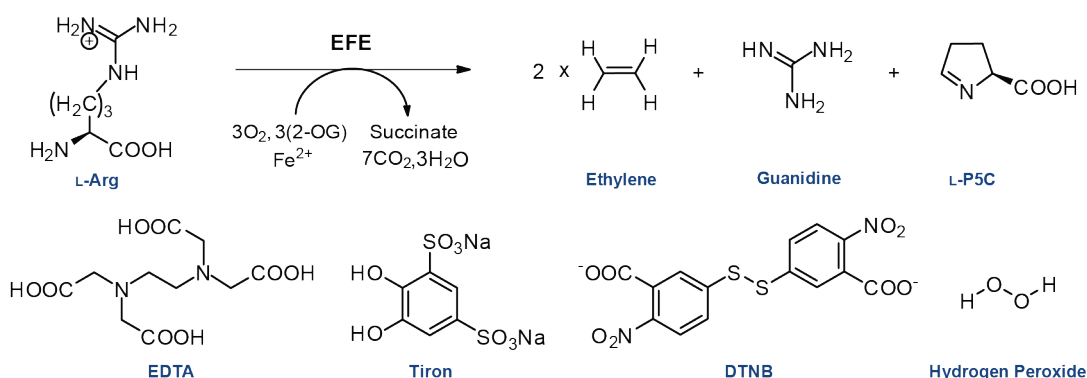


Scheme 4.7: Chemical identification of ethylene produced by *P. digitatum*.

- (a) Ethylene forms a complex with aqueous Hg^{2+} .
 (b) Ethylene ligands are displaced from Hg^{2+} by Cl^- ions from HCl.
 (c) Ethylene reacts with bromine in an electrophilic addition reaction to produce 1,2-dibromoethane.
 (d) Dibromoethane reacts aniline in a nucleophilic substitution reaction to produce DPEDA.
 (e) DPEDA reacts with phenylisothiocyanate in a nucleophilic addition reaction to produce N,N'-dianilinothioformyl-ethylenediamine.

Gas-chromatograph analyses of cell-free extracts from *P. digitatum* and *Pseudomonas syringae* pv. *phaseolicola* PK2 indicated that EFEs require 2-OG, L-arginine, ferrous iron (Fe^{2+}), and molecular oxygen for catalysis (*i.e.* EFEs are ODDs) of ethylene production,^{241, 254, 255, 407} the EFE enzymes are reported to be specific for

2-OG and L-arginine as substrate and cofactor,^{241,254,407} respectively, and are effectively inhibited by EDTA, Tiron (the trade name for disodium 4,5-dihydroxy-1,3-benzenedisulfonate), DTNB [5,5'-dithio-bis(2-nitrobenzoate)], and hydrogen peroxide (Scheme 4.8).²⁴¹ It was reported that inspection of the EFE reactions later revealed that the complete set of products were: succinate, carbon dioxide, L- Δ^1 -pyrroline-5-carboxylate (L-P5C), water, and guanidine (Scheme 4.8);²⁵⁵ the concentrations of EFE reaction components were followed using a thiokinase kit (succinate), glutamate dehydrogenase (2-OG), by amino-acid derivatisation (L-arginine and L-histidine), and using a Clark-type oxygen electrode (molecular oxygen).²⁵⁵ However, the detailed evidence for L- Δ^1 -pyrroline-5-carboxylate (L-P5C) (and guanidine) was not reported.



Scheme 4.8: The reactions catalysed by EFE: O₂, 2-OG, and L-arginine are consumed to yield L- Δ^1 -pyrroline-5-carboxylate (L-P5C), guanidine, succinate, and CO₂.

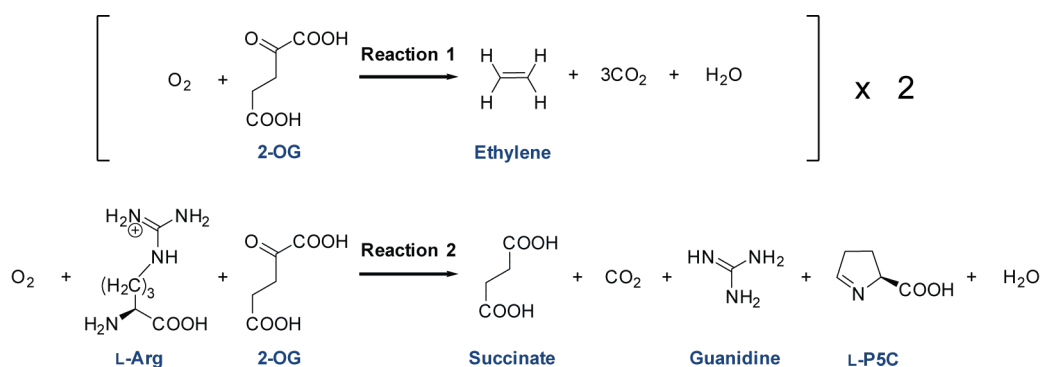
Interestingly, the presence of histidine is reported to increase the efficiency of ethylene formation by the EFE;²⁴¹ under optimal conditions for ethylene formation, without L-histidine, the activity of *P. syringae* pv. *phaseolicola* PK2 EFE is 50% of the activity of this EFE in the presence of L-histidine.²⁴¹ The activity of *P. syringae* pv. *phaseolicola* PK2 EFE in the presence of D-histidine was reported to be 90% of the activity of this EFE in the presence of L-histidine; EFE activity is somewhat stimulated by the presence of imidazole.²⁴¹ It is proposed that L-histidine acts a

scavenger of reactive oxygen species,⁴⁰⁸ which damage the EFE. In addition, EFE from *Pseudomonas syringae* pv. *phaseolicola* was shown to have no requirement for the presence of DTT (dithiothreitol - a reducing agent) for activity.²⁴¹

EFE reactions carried out with [U - ^{14}C]-radiolabelled 2-OG and L-arginine also showed that [^{14}C]-labelled ethylene was formed only from [U - ^{14}C]-2-OG and not from [U - ^{14}C]-L-arginine, implying that ethylene is formed from 2-OG alone and not L-arginine;²⁵⁴ the ethylene generated was trapped with 0.1 M mercuric acetate in methanol or 0.25 M mercuric perchlorate, and the [^{14}C]-radioactivity of the EFE reaction products was determined with a scintillation counter.²⁵⁴ The [^{14}C]-label, originating from a [^{14}C]-labelled 2-OG substrate, was also found in a $^{14}\text{CO}_2$ product, suggesting that the CO_2 was produced from 2-OG;²⁵⁵ the $^{14}\text{CO}_2$ was trapped with a KOH solution and analysed using a scintillation counter.²⁵⁵

An unusual dual-circuit mechanism for EFE catalysis was proposed (Scheme 4.9), based on both the apparent stoichiometry of the consumed reactants (2-OG:L-arginine = 3:1) and the products of the EFE reactions (ethylene:succinate:L-P5C = 2:1:1);²⁵⁵ in one pathway (reaction 1), L-arginine remains bound as a cofactor while 2-OG molecules are consumed by two cycles to produce 6 (3×2) CO_2 molecules, 2 water molecule, and 2 ethylene molecules, whereas in the other EFE pathway (reaction 2), both 2-OG L-arginine are consumed to produce L-P5C, succinate, water, and CO_2 (Scheme 4.9).²⁵⁵ In the ethylene-forming pathway (reaction 1), it was proposed that the bridged Fe(IV)-peroxy intermediate (formed when O_2 binds EFE), fragments in a process which may involve the decarboxylative elimination of the C-4 carboxylate group (attached to C-3) of the 2-OG ligand (releasing CO_2);²⁵⁵ this decarboxylation drives the elimination of ethylene from 2-OG and the formation of an Fe(IV)-oxo ['ferryl-Fe(IV)'] species (Scheme 4.10).²⁵⁵ Ethylene is therefore derived from the $-\text{CH}_2-$ groups at positions C-2 and C-3 of 2-OG. In the succinate-forming pathway (reaction 2), however, Fe(IV)-peroxy intermediate frag-

mentation may be facilitated by the decarboxylation of the C-1-containing carboxylate of EFE-bound 2-OG (*i.e.* the iron-chelating COOH group), driving the direct formation of an Fe(IV)-ferryl intermediate (Scheme 4.10);²⁵⁵ this mechanistic step mirrors the proposed method of 2-OG turnover by more ‘conventional’ ODDs,^{13–15} and allows the newly formed Fe(IV)-ferryl intermediate to hydroxylate the C-5 site of L-arginine (*i.e.* beside the guanidino group) (Scheme 4.10 b).²⁵⁵ The hydroxyguanidino intermediate can spontaneously collapse to form an aldehyde and release a guanidinium group. The L-glutamate-5-semialdehyde can cyclise to produce L- Δ^1 -pyrroline-5-carboxylate (L-P5C) (Scheme 4.10 b).²⁵⁵

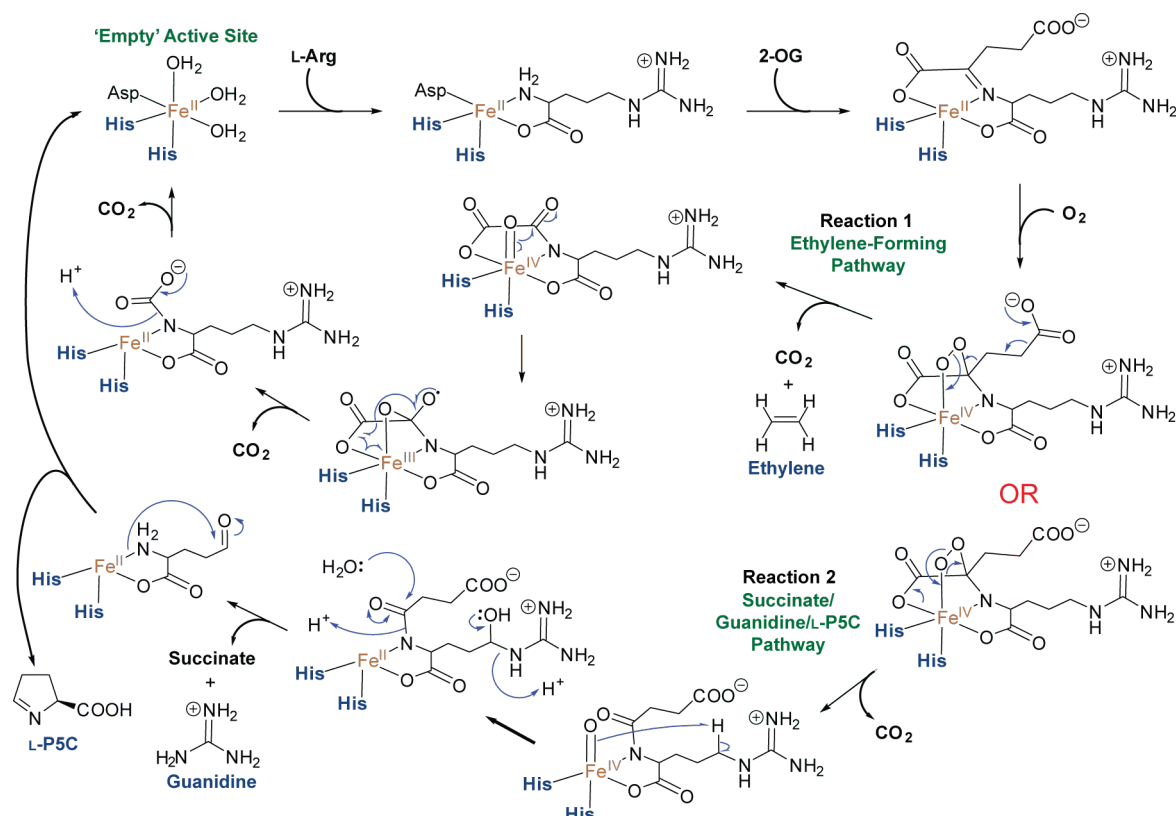


Scheme 4.9: Proposed reactions catalysed by EFE.²⁵⁵

There are two pathways thought to be catalysed by EFE:²⁵⁵

- (1) in reaction 1 (major), two reaction cycles produce 6CO₂, 2H₂O, and 2 ethylenes from two 2-OG molecules;²⁵⁵
- (2) in reaction 2 (minor), one cycle consumes both 2-OG and L-arginine molecules to generate L-P5C, guanidine, succinate, water, and CO₂.²⁵⁵

However, the proposed EFE mechanism²⁵⁵ only partly correlates with mechanisms proposed for 2-OG oxygenases.¹² With catalysis by most 2-OG oxygenases, oxidative decomposition of 2-OG into CO₂ and succinate occurs with coupling to the hydroxylation of a co-substrate (*e.g.* L-arginine to L-hydroxyarginine).¹² My work aimed to investigate the products, stoichiometry, mechanism, and structure of the EFE from *Pseudomonas syringae* *pv.* *phaseolicola* PK2.



Scheme 4.10: Proposed reaction pathways catalysed by EFE (modified from Fukuda *et al.*²⁵⁵).

In reaction 1 (ethylene-forming pathway), the bridged Fe(IV)-peroxy intermediate (formed when O₂ binds) fragments by decarboxylation elimination of the C-4-carboxylic acid group (attached to C-3) of EFE-bound 2-OG; this drives the elimination of ethylene from 2-OG and the formation of an Fe(IV)-oxo ('ferryl-Fe(IV)') species.²⁵⁵ Ethylene is produced from the -CH₂- groups at positions C-2 and C-3 of 2-OG.

In reaction 2 (succinate-forming pathway), Fe(IV)-peroxy intermediate fragmentation is facilitated by the decarboxylation of the C-1-containing carboxylate of EFE-bound 2-OG (*i.e.* the iron-chelating COOH group); this drives the direct formation of an Fe(IV)-ferryl intermediate.²⁵⁵ The newly formed Fe(IV)-ferryl intermediate then hydroxylates the C-5 position of L-arginine (*i.e.* beside the guanidino group), and the 5-hydroxy-L-arginine collapses to produce L-glutamate-5-semialdehyde, a compound which cyclises to produce L-Δ¹-pyrroline-5-carboxylate (L-P5C).²⁵⁵

4.3 Studies on the Ethylene-Forming Enzyme (EFE) from *Pseudomonas syringae* pv. *phaseolicola* PK2

I began by investigating assays for the *Pseudomonas syringae* pv. *phaseolicola* PK2 EFE. Methods were developed for the direct, experimental detection of EFE reaction components for the identification and quantification of substrates and products.

4.3.1 Solution-State NMR Detection Assays

A ^1H -NMR method was developed for the identification of products formed from both 2-OG and L-arginine. A substance's concentration in solution is (often) directly proportional to the integral of its corresponding peak in its NMR spectrum,^{335,336} and therefore peak integration can be used as a method for chemical quantification. NMR spectroscopy has been previously used to follow the reaction progress of ODD enzymes (for both small-molecule and peptide substrate systems),^{17,337–339} and so it was hoped such a method would be successful for following EFE reactions. The EFE was prepared as described in Section 2.3.2.1.

Initially, it was not clear whether substrate turnover by EFE in an NMR tube would be sufficient for detection, and so an EFE reaction was carried out in an Eppendorf tube to allow efficient diffusion of gases into (O_2) and out (CO_2) of the reaction mixture (for efficient EFE reaction catalysis). NMR tubes are relatively long and thin, causing diffusion of molecular oxygen into the bulk reaction mixture to be insufficient for optimal catalysis in some ODD reactions (see Section 3.3.1.2),³⁴⁸ particularly if they are very active (as is the case for a number of small-molecule hydroxylases).^{17,207,338} EFE reaction conditions were based, in part, on conditions previously reported for efficient catalytic activity,^{242,255} and also chosen to aid effective NMR detection. EFE reactions were carried out in deuterated-tris buffer (*i.e.*

containing $^2\text{H}/\text{D}$ - tris- d_{11}) in D_2O to maximise the possibility of detection of reactant and products signals; deuterated positions are not usually detected in proton-NMR experiments, because D nuclei do not have a magnetic dipole moment, but a quadrupole moment (for D, $S=1$, where S is the spin quantum number of an elementary particle), which leads to very broad signals with very poor detection sensitivity.³³⁵ The EFE protein samples used for the NMR assay reactions were also buffer-exchanged into the same 50 mM Tris-DCl, pH 7.5 reaction buffer, to remove complex HEPES buffer signals from the ^1H -NMR spectrum, which may overlap with important reactant or product peaks, and interfere with detection. The EFE reaction mixtures were left for several hours (~ 14 h overnight) at room temperature ($\sim 25^\circ\text{C}$) to maximise product yields and improve the likelihood of product detection; the completed reaction solutions were then transferred to 3 mm NMR tubes for analysis. Appropriate ‘no-enzyme’ controls were also carried out to demonstrate reactions were EFE-catalysed, and not non-enzymatic; these control samples also provide information on the positions of specific reactant proton-NMR signals. Proton-NMR analyses showed that EFE turned over almost all of the 2-OG in the reaction mixture, leaving only very tiny peaks in the ^1H -spectrum (Figure 4.1 a) of the final reaction mixture (only visible by zooming in using Topspin 3.0 analysis software); a large, sharp singlet was also present in this spectrum, indicating the presence of succinate (containing equivalent C-2 and C-3 protons), presumably produced from 2-OG (Figure 4.1 a). In addition, a noticeable resonance was detected at ~ 7.7 ppm, and was proposed to originate from the proton attached to carbon-5 (unsaturated centre) of a $\text{L-}\Delta^1$ -pyrroline-5-carboxylate (L-P5C) product formed from L-arginine; the resonances produced by L-arginine (Figure 4.1 a) were also slightly reduced in intensity (by $\sim 10\%$) compared with those in the ‘no-enzyme’ control mixtures (Figures 4.1 b and c), consistent with L-Arg being consumed by the EFE reaction produce L-P5C. The other signals for protons in L-P5C are close to those of other

species in the mixture, and so resolution of these peaks is low (Figure 4.1 a); a more sensitive NMR spectrometer (possibly of higher field) would be required to more clearly resolve these peaks. Control reactions containing no EFE showed no detectable signs of turnover (Figure 4.1 b, c, and d), suggesting the changes in the reaction mixture composition in complete EFE reactions resulted from enzymatic catalysis.

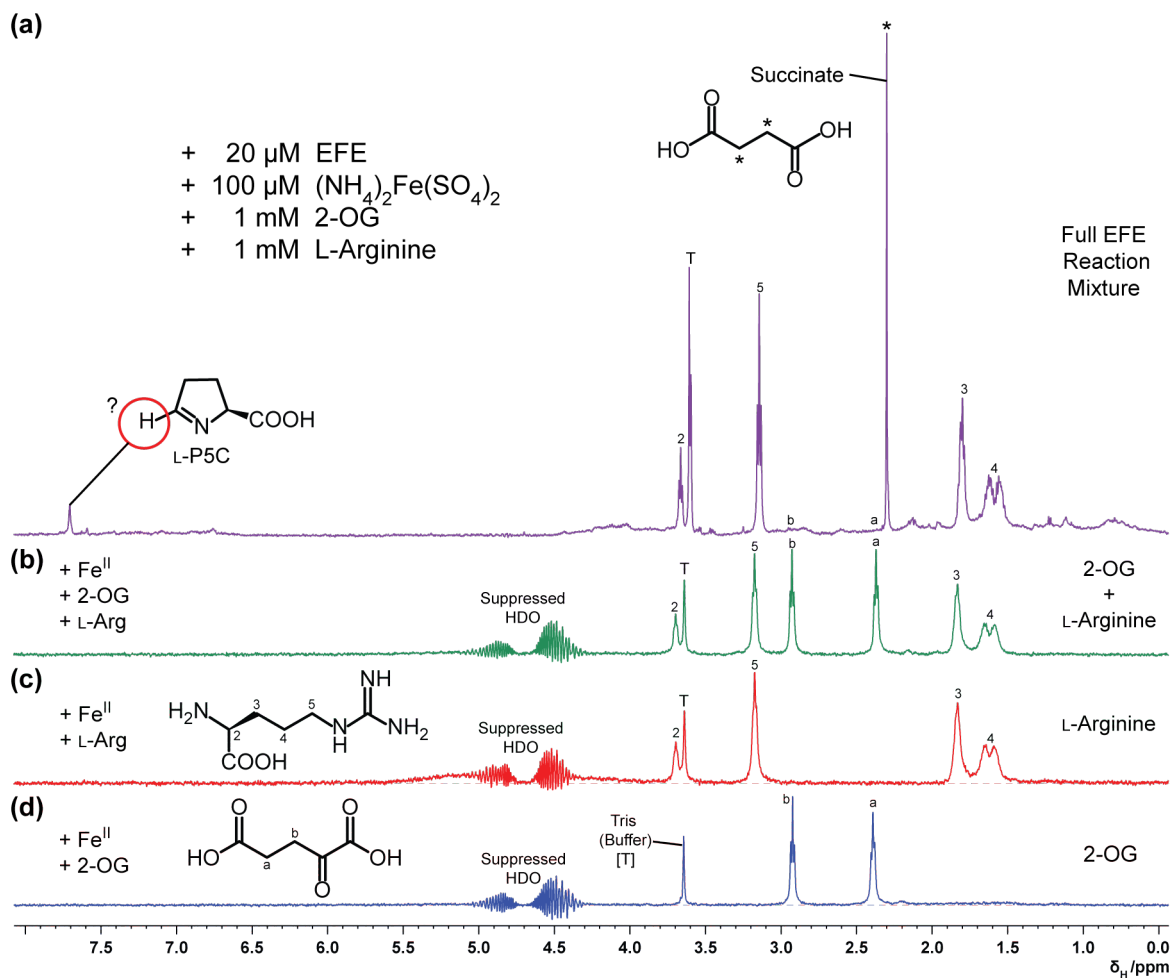


Figure 4.1: NMR analyses of EFE reactions: spectra were measured for:

- (a) an EFE reaction containing L-Arg, disodium 2-OG, $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, and EFE (initially 1 mM, 1 mM, 100 μM , and 20 μM , respectively);
- (b) a ‘no-enzyme’ control reaction (*i.e.* no EFE added) containing all the required EFE reaction components (L-Arg, 2-OG, Fe^{II});
- (c) control reactions containing no 2-OG;
- (d) control reactions containing no L-Arg.

Reactions were carried out at room temperature in 50 mM Tris- d_{11} -DCl, pD 7.5. ^1H -NMR spectra were measured at 298 K for 165 μL reactions in 3 mm NMR tubes using a 600 MHz Bruker AVIII 600 NMR spectrometer [equipped with a 5 mm broadband (BBO)- $^{19}\text{F}/^1\text{H}$ Prodigy N_2 CryoProbe]; presaturation (with O1P = 4.701 ppm) was used to suppress broad, HDO-solvent peaks.

Closer analyses of the EFE reactions on a higher-field NMR spectrometer (700 MHz resonance for ^1H) with higher sensitivity allowed separate signals to be resolved more effectively, particularly those of the species thought to be L-P5C (Figure 4.2 a); in particular, the peak thought to correspond to the L-P5C C-2 proton (α' -proton) was more clearly apparent than in previous spectra, being quite sharp at ~ 4.45 ppm (Figure 4.2 a), whereas in Figure 4.1 a, this α - ^1H peak was too close to the broad, HDO-solvent peak, and was suppressed by the solvent-suppression routine of the experiments pulse sequence. Other peaks produced by the L-P5C product were also partially resolved, despite overlap with peaks from other species, probably corresponding to protons the C-3 (~ 2.15 ppm and ~ 1.70 ppm) and C-4 (~ 2.55 ppm) positions.

Further analysis of the EFE-reaction samples using selective ge-1D TOCSY (gradient-enhanced, 1-dimensional total correlation spectroscopy)³³⁵ experiments more clearly identified the peaks which correspond to those of L-P5C (Figure 4.2 b and c). Selective 1D TOCSY experiments work by irradiating the sample with RF-pulses of a frequency, which corresponds to a particular peak in a proton-NMR spectrum, produced by a compound of interest;³³⁵ magnetization transfer between protons within the same J-coupling network ('spin system') as the irradiated proton, then affords a 1D-proton subspectrum, containing only peaks from that spin system (*i.e.* the spectrum contains only the proton peaks from the J-coupled sections of the same molecule). Selective 1D TOCSY experiments are therefore an effective method of producing a simplified ^1H -spectrum of a particular compound, which is present in a mixture of substances;³³⁵ the substance is effectively 'purified by NMR'. Irradiation of the peak thought to correspond to the C-2 proton (4.451 ppm) of L-P5C produced a 1D TOCSY spectrum that was an accurate representation of the ^1H -spectrum of L-P5C, with all proton signals being highly visible above background noise (Figure 4.2 b); irradiation of the C-5 proton peak produced a 1D TOSCY spectrum contain-

ing the same peaks (Figure 4.2 c), but with lower intensities (but which are still clearly visible using an appropriate scale, particularly if you zoom in using analysis software). A comparison of the ^1H and ge-1D TOCSY EFE reaction spectra with a ^1H -spectrum of authentic DL-P5C (Figure 4.3) suggests that the product of L-Arg turnover is, in fact, L-P5C.

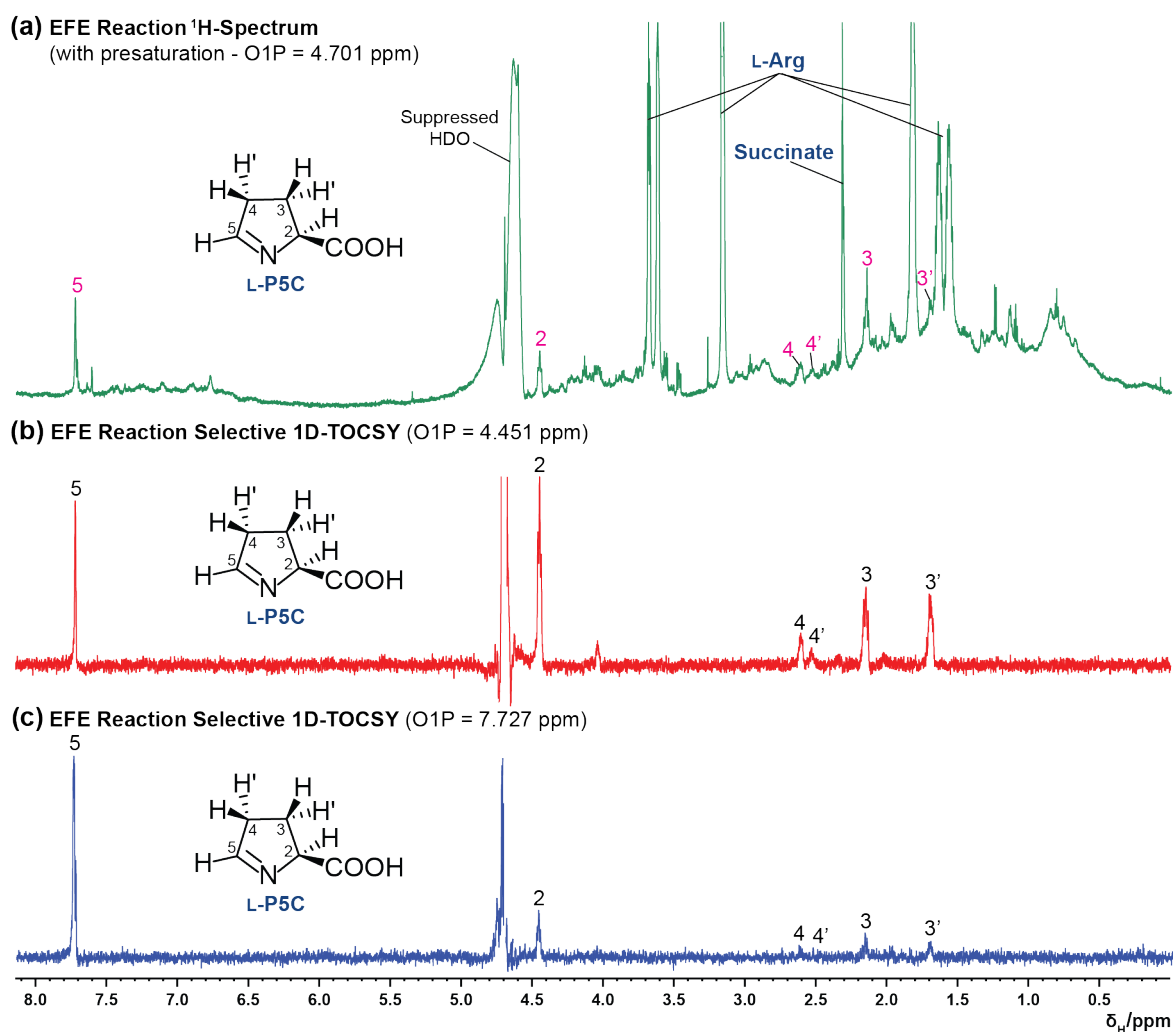


Figure 4.2: Further NMR analyses of EFE reactions: (a) the proton-NMR spectrum of an EFE reaction mixture; (b) and (c) the ge-1D TOCSY spectrum of an EFE reaction (O1P = 4.451 ppm and 7.727 ppm, respectively).

Initial reaction conditions were 1 mM L-Arg, 1 mM 2-OG, 100 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 20 μM EFE. Spectra were measured at 298 K on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe) with samples in 3 mm NMR tubes. Reactions were carried in 50 mM Tris- d_{11} -DCl at room temperature (25 $^\circ\text{C}$).

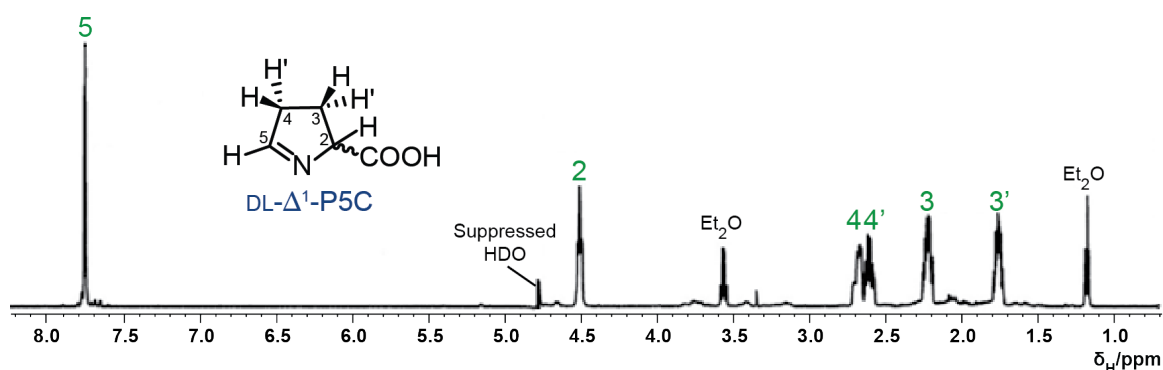


Figure 4.3: ^1H Spectrum of DL-P5C. Part of this figure was taken from a figure reported by Farrant *et al.*⁴⁰⁹

EFE reactions were also carried out without L-arginine to measure EFE's capacity for substrate-uncoupled turnover. In the absence of L-Arg, slight turnover of 2-OG to succinate appeared to occur (Figure 4.4 a), but to a lesser extent than in the presence of L-Arg (Figure 4.4 b). Close inspection of the ^1H -NMR spectrum (Figure 4.4 a) for this substrate-uncoupled reaction reveals a relatively small, sharp singlet (produced by succinate) at ~ 2.3 ppm, whereas a L-Arg-coupled EFE reactions produce a large succinate singlet in their ^1H -spectrum (Figure 4.2 b), with virtually complete turnover of 2-OG (*n.b.* 2-OG peaks are practically absent from this ^1H -spectrum). The estimated yield of conversion of 2-OG succinate was calculated to be 10% using peak integration information (and assuming all products originate from 2-OG). Interestingly, other peaks are also present in the substrate-uncoupled EFE reaction spectrum, which overlap with the 2-OG and succinate peaks (Figure 4.4 a); it is unclear what species these peaks (denoted by *) correspond to, since they are not present in spectra from previous NMR-based studies on ODDs,^{17,337–339} though they may be produced owing to EFE activity. Perhaps the species are deuterated derivatives of 2-OG or succinate (or similar compounds), since these EFE reactions were carried out in buffer dissolved in D_2O ; alternatively, these peaks may previously undetected product of similar structure to 2-OG and succinate. Further analysis of these products is required to elucidate their identities; the products only

appear in the solutions containing EFE and not in the ‘no-enzyme’ (-ve) control mixtures (Figure 4.4 c), suggesting that the products are produced by EFE activity.

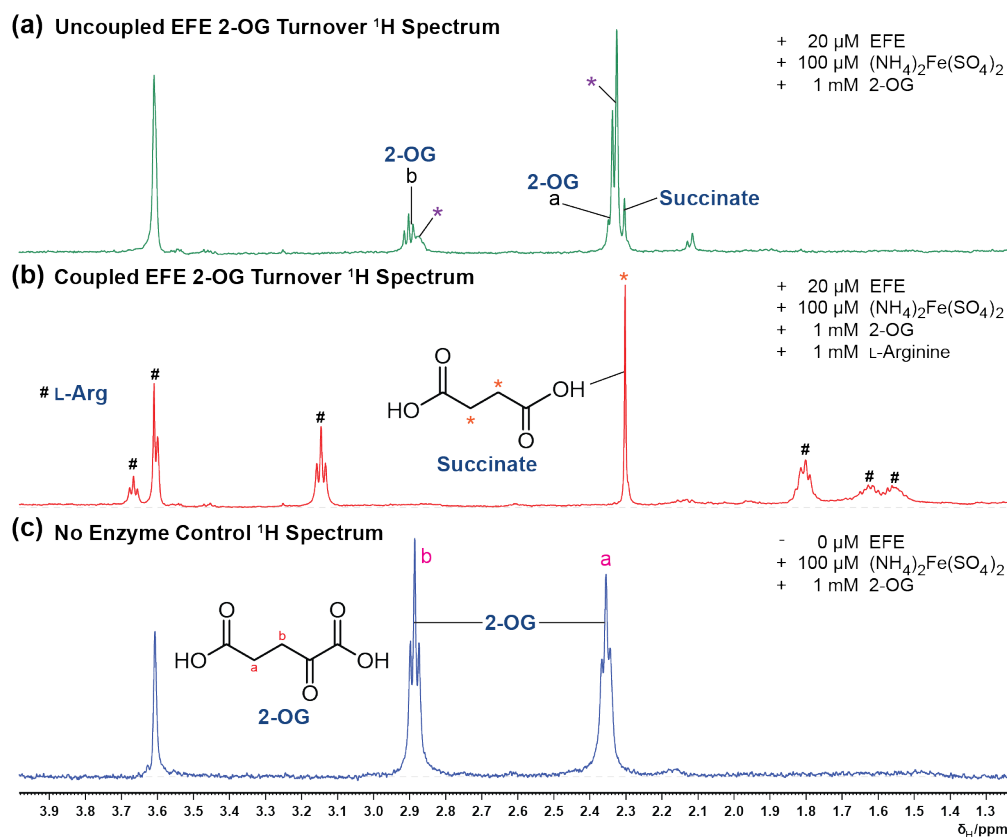


Figure 4.4: ^1H -Spectrum of uncoupled turnover of 2-OG by EFE: a comparison with spectra measured for a 2-OG-coupled reaction (with L-Arg) and a ‘no-enzyme’ (-ve) control sample.

(a) ^1H -Spectrum of the substrate-uncoupled (*i.e.* without L-Arg) turnover of 2-OG by EFE; initial conditions were: 20 μM EFE, 100 μM Fe^{2+} , 1 mM 2-OG. An unassigned product was observed (marked with an * in the spectrum).

(b) ^1H -Spectrum of the substrate-coupled turnover of 2-OG by EFE (*i.e.* with L-Arg - ‘complete’ reaction); initial conditions were: 20 μM EFE, 100 μM Fe^{II} , 1 mM 2-OG, 1 mM L-Arg.

(c) ^1H -Spectrum of a ‘no-enzyme’ control mixture for an uncoupled 2-OG turnover reaction; initial conditions were: 100 μM Fe^{II} , 1 mM 2-OG. No reactions were observed.

All reactions were prepared in 50 mM Tris- d_{11} -DCl, pD 7.5 buffer in D_2O and incubated at 25 $^\circ\text{C}$ for ~14 h overnight. ^1H -NMR spectra were measured at 298 K for 165 μL reactions in 3 mm NMR tubes using a 600 MHz Bruker AVIII 600 NMR spectrometer [equipped with a 5 mm broadband (BBO)- $^{19}\text{F}/^1\text{H}$ Prodigy N_2 CryoProbe]; presaturation (with O1P = 4.701 ppm) was used to suppress broad, HDO-solvent peaks.

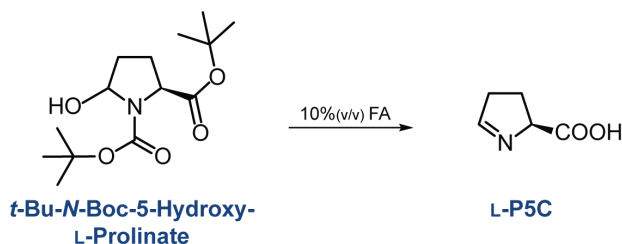
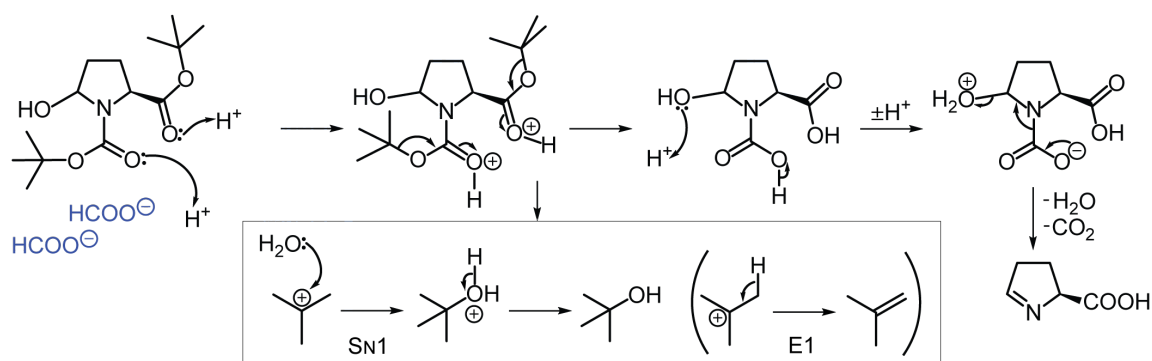
4.3.2 Solution-State LC/MS-Detection Assays

In order to further study the solution-state products of EFE reactions, a liquid-chromatography/mass-spectrometry (LC/MS) detection assay was developed. Given the chemical similarity of L-P5C with L-proline and other amino acids, the method developed was based, in part, on a method used for detection of proline and derivatives of proline (predominantly hydroxylated proline analogues) (see Section 3.3.2); the same analytical, mixed-mode column (250 mm×3.2 mm SIELC Primesep 100) and solvents were used to perform HPLC runs, using similar HPLC gradient methods. The mass spectrometry setting were also identical as those used previously for proline hydroxylase LC/MS assays.

Initial 100 μ L EFE reactions were carried under similar conditions to those used in the EFE NMR assays, with similar buffer conditions (50m mM Tris-HCl, pH 7.5 in water) and at room temperature; however, deuterated buffers and D₂O were not required and were replaced with conventional ¹H-containing materials and solutions. Samples were then analysed using the LC/MS method. Samples containing all the necessary EFE reaction components (as reported by Nagahama *et al.*^{241,407} and Fukuda *et al.*^{242,255}) contained a substance with a retention time of ~33 min and a mass of ~114 Da, corresponding to the protonated molecular-ion peak ([M+H]⁺) of L-P5C (Figure 4.5 a). Reactions containing supplementary L-ascorbate were also carried out to determine whether a reducing agent, such as L-ascorbate, would improve the catalytic activity of EFE, as it does with many other ODD enzymes (perhaps by maintaining Fe^{II} levels by reducing Fe^{III}).^{18,40} The addition of L-ascorbate did not appear to improve or diminish the catalytic efficiency of EFE to produce L-P5C from L-Arg; the integral for the ion-extracted peaks for L-P5C was roughly the same for reactions with and without L-ascorbate (Figures 4.5 a and b). Ascorbate having no apparent effect on EFE turnover is consistent with published re-

ports suggesting that DTT (dithiothreitol), another reducing agent, has no effect on EFE's catalytic efficiency.²⁴¹ Further evidence for the identity of the LC/MS peak with a retention time of ~33 min, proposed to correspond to L-P5C, came from analysis of an authentic standard, run using the same LC/MS method; both the standard and the enzyme reaction product had the same retention time. A precursor of L-P5C (*tert*-butyl-*N*-Boc-5-hydroxy-L-prolinate) was chemically synthesised in another project (studying carbapenem biosynthesis),^{357,358,360} and L-P5C was prepared directly before use by deprotecting the precursor with a 10% (v/v) formic acid solution (Scheme 4.11) (and was then stored at -80 °C to preserve the L-P5C). The 'no-enzyme' controls of the EFE reactions (Figure 4.5 c and d) showed no peaks corresponding to L-P5C in their ion-extracted LC/MS chromatograms, showing EFE does indeed produce L-P5C from L-Arg.

(a)

(b) Deprotection Mechanism (S_N1) (+ some E1)

Scheme 4.11: Preparation of L-P5C from a *t*-Bu-*N*-Boc-5-OH-L-prolinate precursor:

(a) treating the precursor with 10% (v/v) removes the *t*-Bu and Boc protecting groups, and then L-P5C is spontaneously formed;

(b) this conversion employs an acid-catalysed mechanism, (predominantly) involving the S_N1-based removal of *t*-Bu groups, which triggers decarboxylation and dehydration steps.

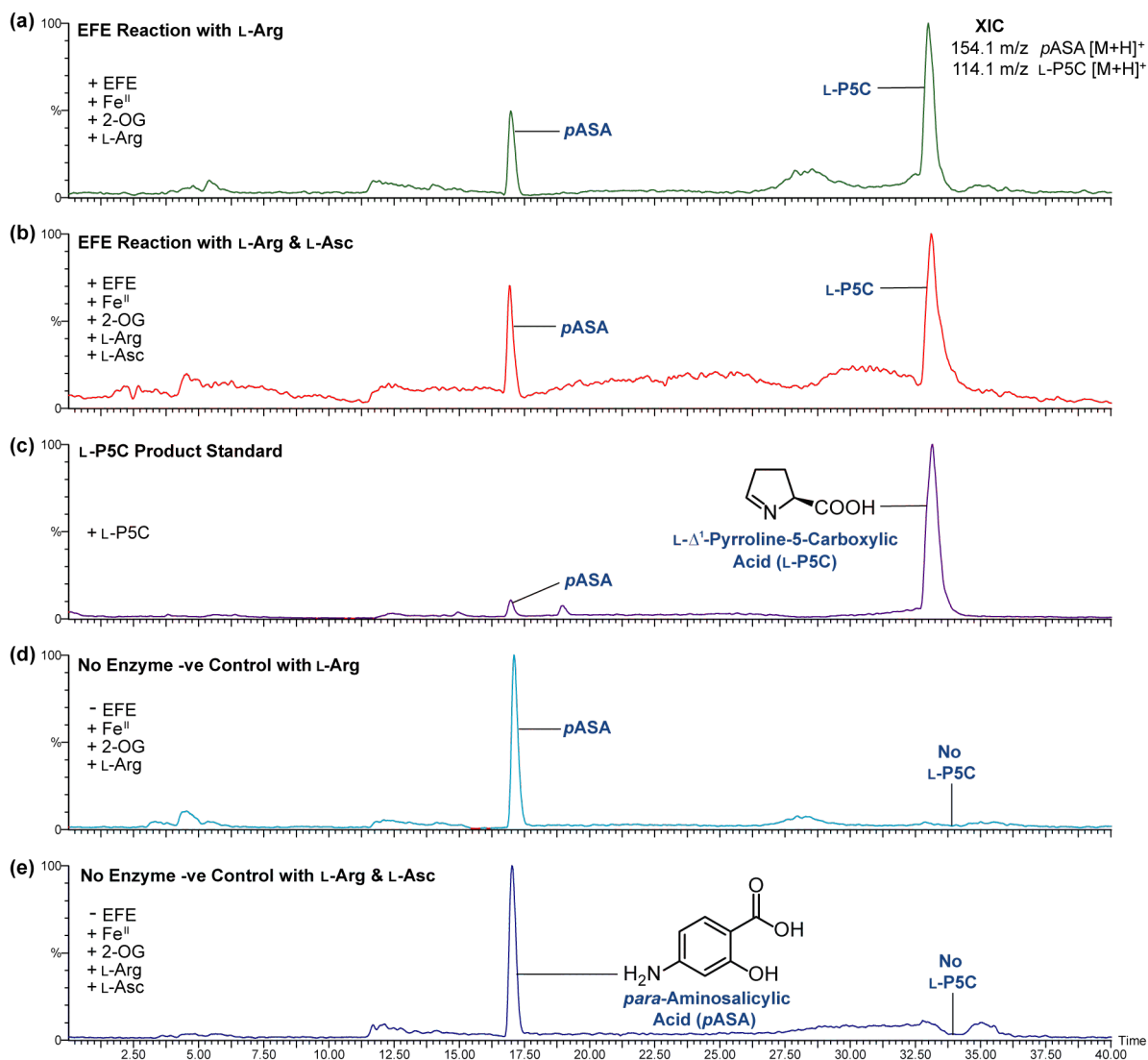


Figure 4.5: LC/MS-based analysis of EFE reactions; 100 μ L reactions were carried out at room temperature (25 $^{\circ}$ C) for 14 h in 50 mM Tris-HCl, pH 7.5 buffer.

- (a) An EFE reaction with 15 μ M EFE, 100 μ M (NH₄)₂Fe(SO₄)₂, 1 mM 2-OG, 1 mM L-Arg.
 (b) An EFE reaction with 15 μ M EFE, 100 μ M (NH₄)₂Fe(SO₄)₂, 1 mM 2-OG, 1 mM L-Arg, 1 mM sodium L-ascorbate.
 (c) A chemical synthesised L-P5C standard.
 (d) A 'no-enzyme' (-ve) control with 100 μ M (NH₄)₂Fe(SO₄)₂, 1 mM 2-OG, 1 mM L-Arg.
 (e) A 'no-enzyme' (-ve) control with 100 μ M (NH₄)₂Fe(SO₄)₂, 1 mM 2-OG, 1 mM L-Arg, 1 mM sodium L-ascorbate.

Complete reactions were quenched with 100 μ L methanol (1:1) containing a *para*-aminosalicylic acid (pASA) standard, and then microcentrifuged to remove precipitate. Product analysis was carried out using a Waters Quattro Micro ESI-MS system (using positive mode) and HPLC methods used a SiELC Primesep 100 HPLC column (100 $\text{\AA}$, 5 μ m) [4.6 $\times$ 250mm].

4.3.2.1 L-Arg Substrate Analogue Studies

To investigate the substrate-selectivity of EFE, compounds which are structural similar to L-arginine were tested for catalytic turnover (Figure 4.6); it was thought that these experiments may suggest which structural motifs (of substrates) are necessary for catalysis, and improve our mechanistic understanding of EFE. Reactions were carried out using the similar conditions as used in Section 4.3.2, and the products were analysed using a similar LC/MS method to determine whether any L-P5C was produced.

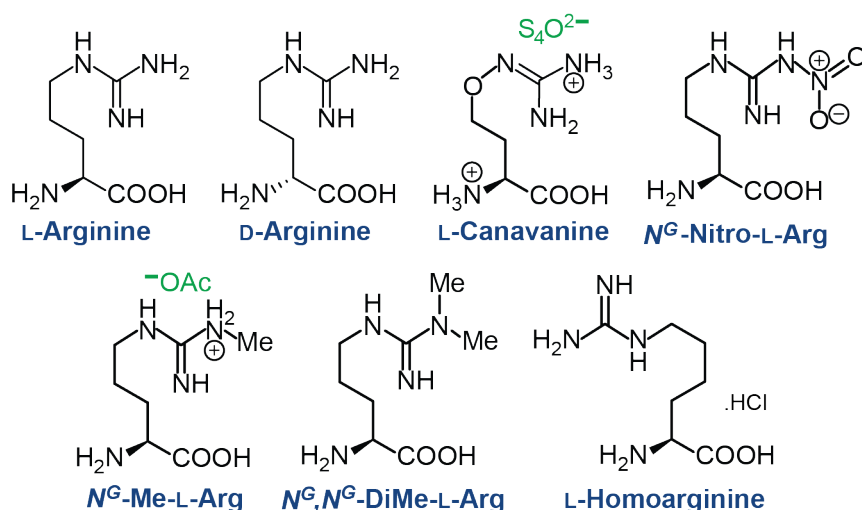


Figure 4.6: L-Arginine substrate analogues were tested for turnover by EFE: an LC/MS assay was used to detect any products that may be formed (*i.e.* L-P5C).

To investigate the effect that arginine's C-2 ('α') stereochemistry has on the efficiency of the EFE reaction, the D-isomer of arginine was tested alongside L-Arg for substrate turnover; previous studies, observing ethylene production, suggested that only a very small amount of ethylene is produced by EFE when using a D-Arg substrate analogue (~3% relative to levels using L-Arg).²⁴¹ Whilst almost complete turnover of L-Arg into L-P5C was seen (~97% - L-P5C retention time = ~16 min - Figure 4.7 c), no detectable quantities of D-P5C appeared to be produced by EFE from D-Arg (Figure 4.7 a); the EFE reaction using D-Arg as a substrate produced a

near identical LC/MS chromatogram to that seen for the D-Arg ‘no-enzyme’ mixture (Figure 4.7 b), suggesting D-Arg is not a substrate. The sensitivity of arginine detection (both L- and D- forms) was relatively poor, possibly because in positive-mode ESI-MS (electrospray-ionisation mass spectrometry), it may exist in 2 charged states (*i.e.* single and double protonation), both with different mass:charge (m/z) ratios; thus, arginine could therefore exist in a population of different ions. The sensitivity was poorer below ~ 100 m/z (high noise), a problem given the mass/charge ratio of the $[(DL\text{-Arg})+2H]^{2+}$ (87.6 m/z) molecular ions. It is possible that arginine would exist as singly negatively charged molecular ions in negative-mode, ESI-MS experiments (*i.e.* $[(DL\text{-Arg})-H]^{-}$), allowing for accurate detection of the arginine levels.

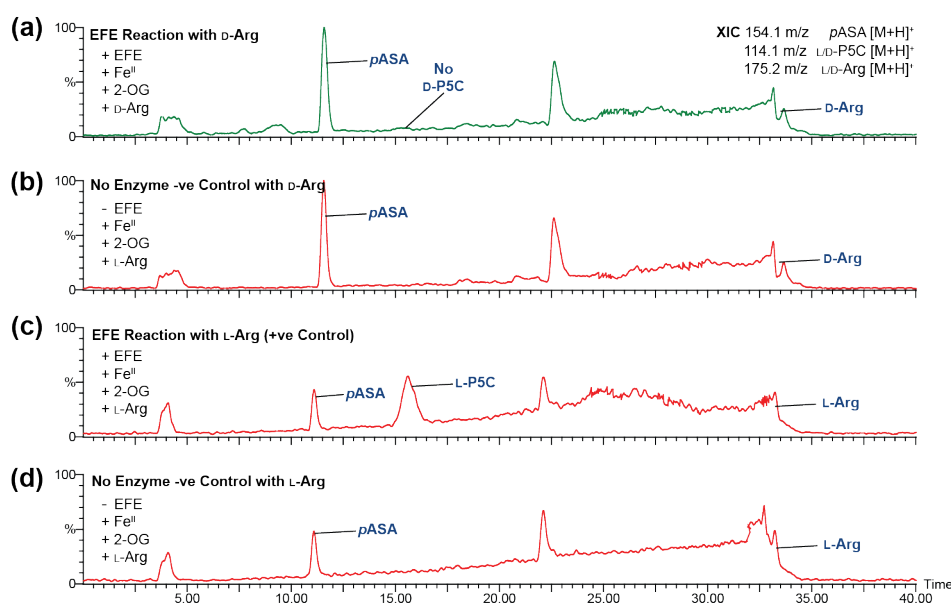


Figure 4.7: D- and L-Arginine EFE Reactions: the LC/MS analysis of the reaction products.

(a) An EFE reaction with 15 μM EFE, 100 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM 2-OG, 1 mM D-Arg.

(b) A ‘no-enzyme’ (-ve) control with 100 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM 2-OG, 1 mM D-Arg.

(c) An EFE reaction with 15 μM EFE, 100 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM 2-OG, 1 mM L-Arg.

(e) A ‘no-enzyme’ (-ve) control with 100 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM 2-OG, 1 mM L-Arg.

100 μL reactions were carried out at room temperature (25 $^\circ\text{C}$) for 14 h in 10 mM HEPES-NaOH, pH 7.5 buffer. Complete reactions were quenched with 100 μL methanol (1:1) containing a *para*-aminosalicylic acid (*pASA*) standard (0.25 mM), and then microcentrifuged to remove precipitate.

L-(+)-(S)-Canavanine (L-Cav) is a non-proteinogenic α -amino acid which is isosterically analogous to L-arginine; the sole difference between L-Arg and canavanine is the C-5 methylene bridge (-CH₂-) in arginine is replaced with an oxa group (-O-, oxygen atom) in canavanine. L-Canavanine was previously tested as a substrate of EFE, being reported to stimulate poor ethylene production (~7% relative to levels using L-Arg).²⁴¹ LC/MS analysis of EFE reactions using an L-canavanine substrate showed no L-P5C was produced from L-canavanine (Figure 4.8); the chromatograms for the complete EFE reaction mixtures (Figure 4.8 a) and the 'no-enzyme' negative control (Figure 4.8 b) are identical. These observations suggest L-canavanine is not a substrate of EFE; this conclusion is not particularly surprising, given that EFE was predicted to hydroxylate carbon-5 of L-Arg (*i.e.* adjacent to the guanidino group),²⁴² the methylene position which is replaced with an oxygen atom in L-canavanine. L-Canavanine (L-Cav) elutes at the same retention time as the *para*-aminosalicylic acid (*p*ASA) standard, being detected in the form of a singly charged molecular ion (single protonation state). The electronwithdrawing effect of the oxygen atom (Pauling electronegativity = 3.44 - very high)¹ in L-Cav likely causes its guanidinoxy moiety to be much less basic ($pK_a \sim 7.05$)^{410,411} than the guanidino group of L-Arg ($pK_a \sim 12.5$);^{410,411} the guanidinoxy group of L-Cav is therefore much less likely to be protonated than L-Arg's guanidino group, predominantly adopting a more-rigid *endo*-tautomer (whereas L-Arg tends to adopt more of the *exo*-tautomer).^{410,411}

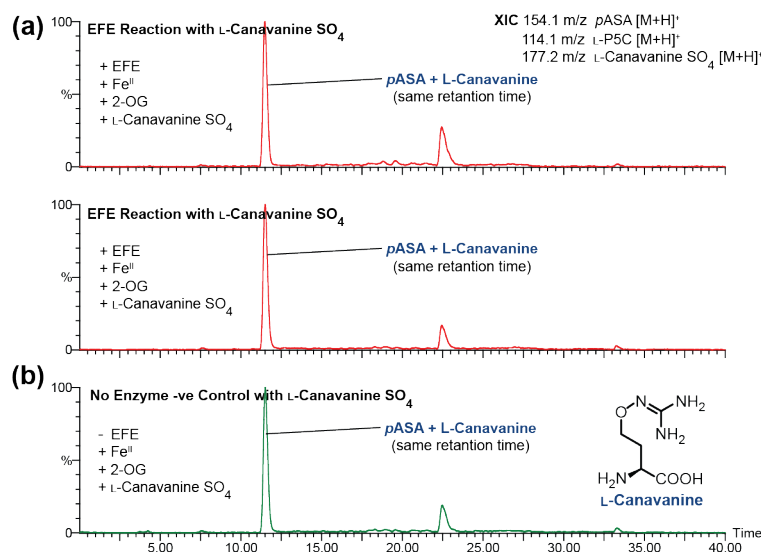


Figure 4.8: EFE reactions with a L-canavanine substrate: the LC/MS analysis of the reaction products.

(a) EFE reactions with 15 μ M EFE, 100 μ M (NH₄)₂Fe(SO₄)₂, 1 mM 2-OG, 1 mM L-Cav.

(b) A 'no-enzyme' (-ve) control with 100 μ M (NH₄)₂Fe(SO₄)₂, 1 mM 2-OG, 1 mM L-Cav.

100 μ L reactions were carried out at room temperature (25 °C) for 14 h in 10 mM HEPES-NaOH, pH 7.5 buffer. Complete reactions were quenched with 100 μ L methanol (1:1) containing a *para*-aminosalicylic acid (pASA) standard (0.25 mM), and then microcentrifuged to remove precipitate.

L-Arginine derivatives with substituents on the guanidino group were tested for EFE turnover. Notably, N^G -Methyl-L-Arg showed almost complete turnover (98%) into L-P5C; LC/MS analysis of complete N^G -Me-L-Arg EFE reactions showed a prominent L-P5C peak ($m/z = 114.13 - [(L-P5C) + H]^+$) in their ion-extracted chromatograms (Figure 4.9 a). The product of the N^G -Me-L-Arg EFE reactions had the same retention time (~ 16 min) as the L-P5C product from L-Arg EFE reactions (and the correct mass), suggesting the product was, in fact, L-P5C. The L-P5C peak is missing from the 'no-enzyme' (-ve) controls (Figure 4.9 b), which indicates the production of L-P5C was catalysed by EFE. Similarly, LC/MS-analysis of products from EFE reactions using a N^G, N^N -dimethyl-L-arginine (Asymmetric DiMe-L-Arg) showed that L-P5C was also efficiently produced from N^G, N^G -DiMe-L-Arg (Figure 4.10), with a similar yield (96%) as the N^G -Me-L-Arg substrate. Analyses of both N^G -Me-L-Arg and N^G, N^G -DiMe-L-Arg EFE reactions gave poor detection of N^G -Me-L-Arg and N^G, N^G -DiMe-L-Arg (as for arginine, under the conditions used for analysis, N^G -methyl- and N^G, N^G -dimethyl-L-arginine will predominantly exist as doubly protonated, doubly charged molecular ions, which are difficult to detect).

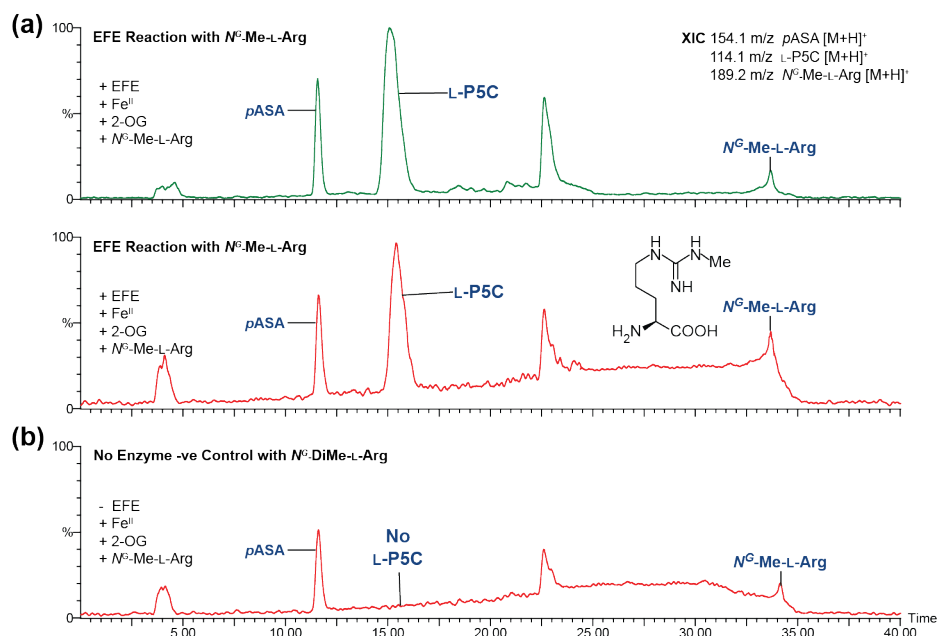


Figure 4.9: EFE reactions with a N^G -Me-L-Arg substrate: the LC/MS analysis of the reaction products.

(a) EFE reactions with 15 μM EFE, 100 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM 2-OG, 1 mM N^G -Me-L-Arg.

(b) A 'no-enzyme' (-ve) control with 100 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM 2-OG, 1 mM N^G -Me-L-Arg.

100 μL reactions were carried out at room temperature (25 $^\circ\text{C}$) for 14 h in 10 mM HEPES-NaOH, pH 7.5 buffer. Complete reactions were quenched with 100 μL methanol (1:1) containing a *para*-aminosalicylic acid (pASA) standard (0.25 mM), and then microcentrifuged to remove precipitate.

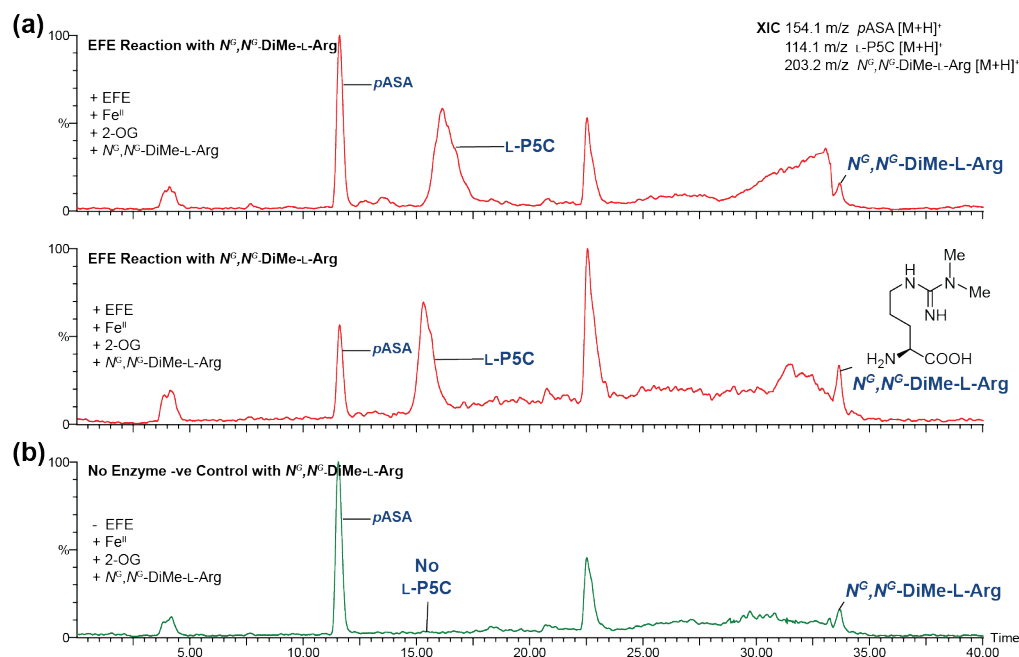


Figure 4.10: EFE Reactions with the N^G, N^G -DiMe-L-Arg substrate: the LC/MS analysis of the reaction products.

(a) EFE reactions with 15 μM EFE, 100 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM 2-OG, 1 mM N^G, N^G -DiMe-L-Arg.

(b) A ‘no-enzyme’ (-ve) control with 100 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM 2-OG, 1 mM N^G, N^G -DiMe-L-Arg.

100 μL reactions were carried out at room temperature (25 $^\circ\text{C}$) for 14 h in 10 mM HEPES-NaOH, pH 7.5 buffer. Complete reactions were quenched with 100 μL methanol (1:1) containing a *para*-aminosalicylic acid (pASA) standard (0.25 mM), and then microcentrifuged to remove precipitate.

N^G -Nitro-L-Arg was also tested for its EFE substrate activity; no L-P5C or other products were detected by LC/MS-analyses of the product mixtures (Figure 4.11 a) and the chromatograms matched those of the ‘no-enzyme’ control, suggesting that N^G -nitro-L-Arg is not an EFE substrate. The highly electron-withdrawing nitro group likely makes the guanidino group much less basic, and the *N*-nitro-guanidino group remains unprotonated (*i.e.* has no net charge); compared with L-Arg, this factor, together with the extra bulk from the nitro group, may affect the binding of the N^G -nitro-L-Arg to EFE. The apparently singly charged nature of N^G -nitro-L-Arg (*i.e.* $[\text{M}+\text{H}]^+$) does, however, allow it to be easily detected using a positive-mode ESI-MS detection method; the ion-extracted chromatograms for the N^G -nitro-L-Arg

EFE reactions contain a prominent peak at ~39 min ($m/z = 203.2$), showing the presence of N^G -nitro-L-Arg. The conclusion that N^G -nitro-L-Arg is probably not an EFE substrate is consistent with a previous study by Nagahama *et al.*,²⁴¹ which reported that no ethylene is produced by EFE reactions using a N^G -nitro-L-Arg.

To test the effect that the length of the arginine side-chain has on its substrate activity, a L-homoarginine substrate was tested for turnover by EFE; the side-chain of homoarginine is one carbon atom longer (*i.e.* an extra methylene group) than regular arginine. No L-P5C (or its 6-membered analogues) was detected in the LC/MS chromatograms of the L-homoarginine EFE reactions and detection of L-homoarginine was poor (like L-arginine).

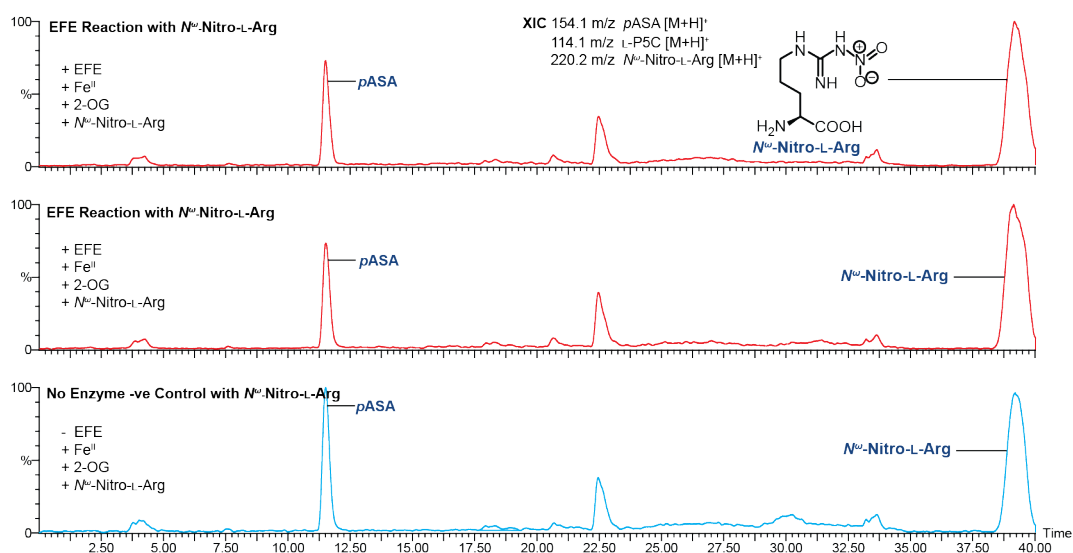


Figure 4.11: EFE Reactions with the N^G -Nitro-L-Arg substrate: the LC/MS analysis of the reaction products.

- (a) EFE reactions with 15 μM EFE, 100 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM 2-OG, 1 mM N^G -Nitro-L-Arg.
 (b) A 'no-enzyme' (-ve) control with 100 μM $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$, 1 mM 2-OG, 1 mM N^G -Nitro-L-Arg.

100 μL reactions were carried out at room temperature (25 $^\circ\text{C}$) for 14 h in 10 mM HEPES-NaOH, pH 7.5 buffer. Complete reactions were quenched with 100 μL methanol (1:1) containing a *para*-aminosalicylic acid (*pASA*) standard (0.25 mM), and then microcentrifuged to remove precipitate.

4.3.3 L-P5C Purification and Characterisation

In an attempt to further prove that the second product of the succinate-pathway of EFE is L-P5C, L-P5C was purified (at least partially) using an LC/MS (HPLC) method similar to that used to purify proline-analogues (see Section 5.2.2.3). EFE reactions using L-arginine and 2-oxoglutarate substrates were scaled up 20-fold, quenched with methanol (with no *p*ASA standard), diluted with deionised water, and then freeze-dried to remove all water/solvent; MeOH has a very low freezing point (-97.6 °C), so the quenched reactions were diluted with water before being lyophilised, to stop the samples melting during the freeze-drying process. Dry EFE reaction product sample then diluted in 20% (v/v) methanol in water, and 3 LC/MS runs were performed to purify the L-P5C. The L-P5C peaks (retention time = ~12 min) were clearly visible in both the TIC (total-ion count) (Figure 4.12) and XIC (extracted-ion count) (Figure 4.13 a) chromatograms, suggesting good L-P5C product yields, and these L-P5C fractions were collected; the major molecular-ion in the in these peaks appears to be a species with $m/z = \sim 114$ (Figure 4.13 b), consistent with the presence of L-P5C. The L-P5C fractions from the first LC/MS purification step were then repurified using the same LC/MS method used previously, further ‘polishing’ the L-P5C sample; the L-P5C eluted from the HPLC column at roughly the same retention time (slightly later at ~13 min) and was further resolved from impurity peaks.

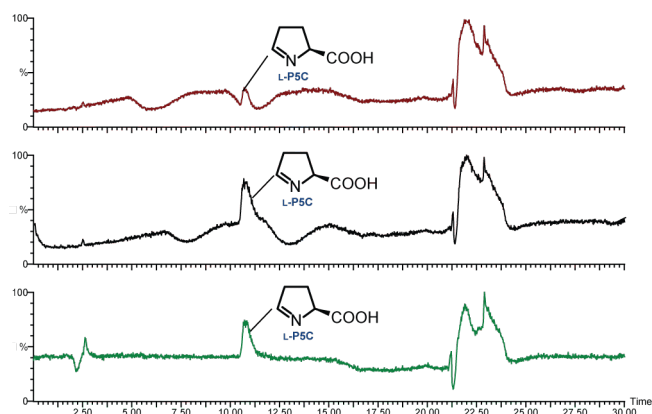


Figure 4.12: LC/MS Purification of L-P5C from EFE reactions: the TIC (total-ion count) chromatograms of each purification run.

2×1 mL reactions were carried out at room temperature (25 °C) for 14 h in 10 mM HEPES-NaOH, pH 7.5 buffer. Complete reactions were quenched with 1 mL methanol (1:1), and then microcentrifuged to remove precipitate. L-P5C eluted with a retention time of ~12 min.

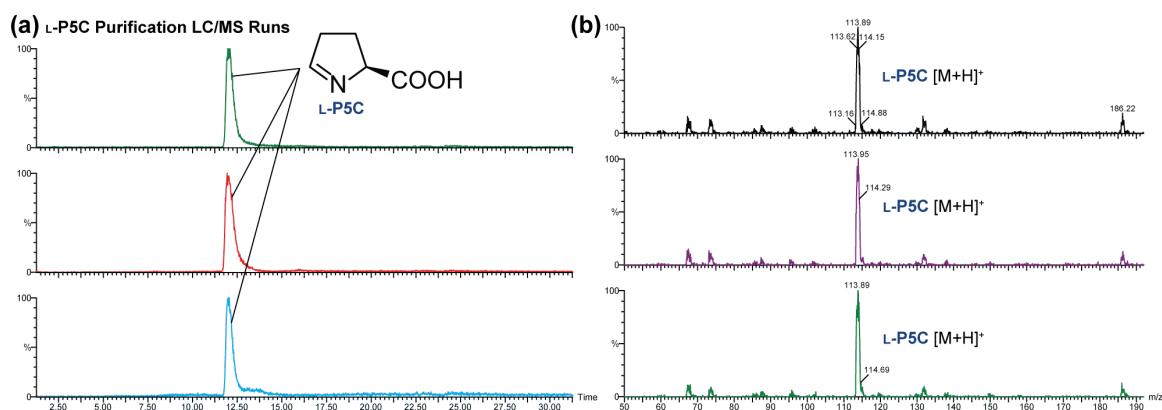


Figure 4.13: LC/MS Purification of L-P5C from EFE reactions: (a) the XIC (extracted-ion count) chromatograms of each purification run and (b) the mass spectra for the L-P5C peaks (for which the major peak $m/z = \sim 114$).

The purified L-P5C samples were dissolved in D₂O and then freeze-fried for NMR analysis; ¹H-NMR (Figure 4.14 a) showed that the substance purified had a spectrum with the same peaks (thought to be from L-P5C) as those found in previous ¹H-NMR and ge-1D TOCSY spectra of EFE reactions (Figures 4.14 c and b, respectively), confirming those peaks were, indeed, from L-P5C product peaks. The only peak not particularly visible in the L-P5C ¹H-spectrum (Figure 4.14 a) was that from the C-2 (‘α’) proton ($\delta_{\text{H}} = 4.451$ ppm), which was overlapping with the large, broad, HDO-solvent peak ($\delta_{\text{H}} = 4.701$ ppm), despite using a pulse sequence with solvent-suppression; zooming in with Bruker Topspin 3.0 analysis software shows this peak as a small shoulder (‘blip’) on the solvent peak at roughly 4.45 ppm, and is not well-resolved. The L-P5C sample was relatively weak; 1D TOCSY experiments, to isolate the L-P5C peaks from the slight contamination, were unsuccessful.

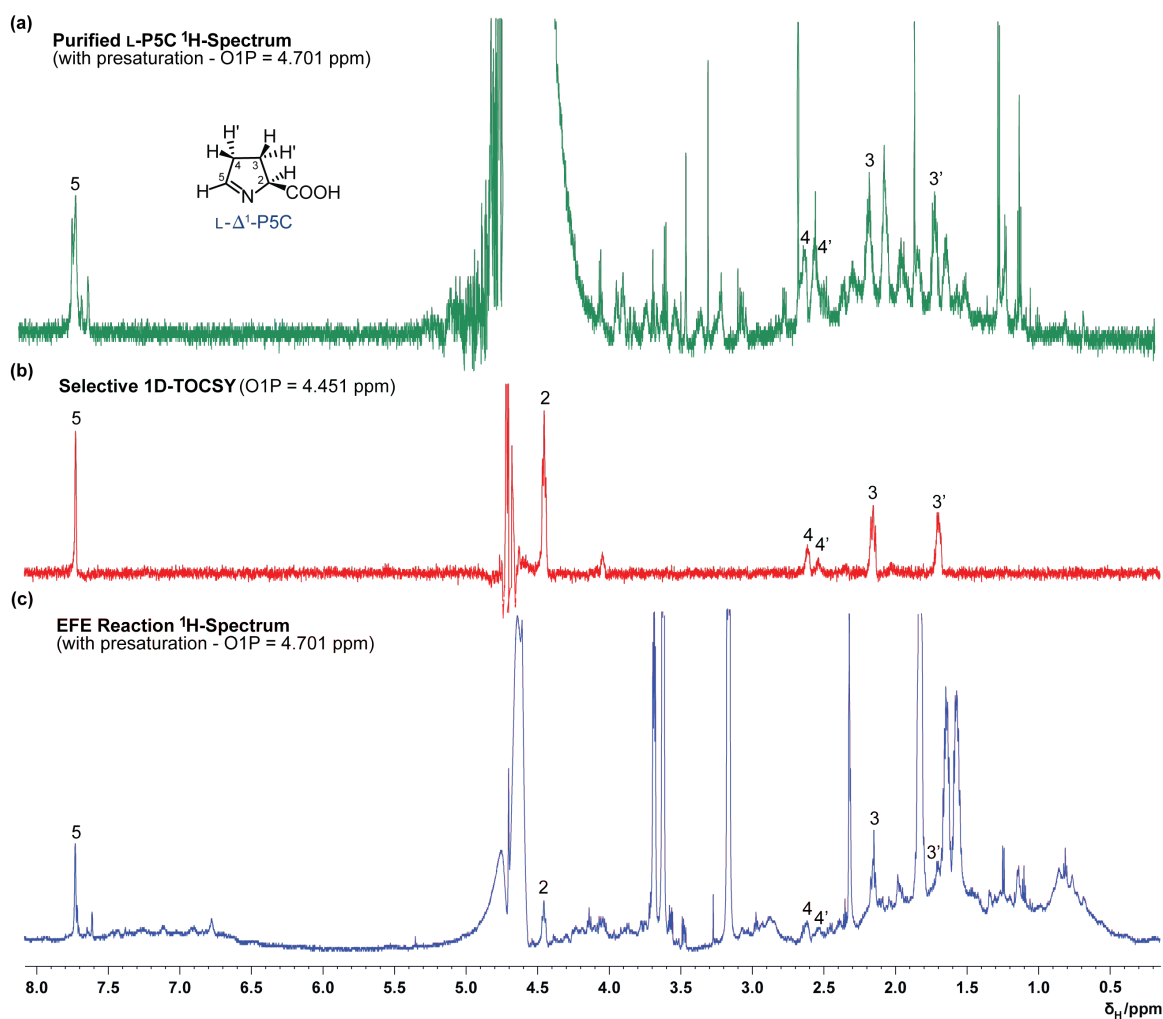


Figure 4.14: ^1H -NMR spectrum of LC/MS-purified L-P5C (compared EFE reaction spectra).

(a) ^1H -Spectrum of LC/MS-purified L-P5C.

(b) The ge-1D TOCSY spectrum of an EFE reaction (O1P = 4.451 ppm).

(c) ^1H -Spectrum of an EFE reaction with solvent suppression (O1P = 4.701 ppm).

Spectra were measured at 298 K on a 700 MHz Bruker AVIII 700 spectrometer (with a $^1\text{H}/^{13}\text{C}/^{15}\text{N}$ TCI CryoProbe) with samples in 1 mm NMR tubes. Samples were prepared in 16 μL of 99.99%-deuterated D_2O with 0.05% (w/v) TSP- d_4 standard.

4.3.4 Gas Chromatograph (GC) Assays

In order to detect ethylene being produced by EFE reactions, reactions were incubated in sealed vials (to trap gaseous products); the head-space gas was then extracted using a gas syringe and injected into a GC system with a flame-ionisation detector (FID) for analysis.

GC chromatograms (Figure 4.15 a) for EFE reactions with L-Arg, 2-OG, and Fe^{II} showed a large peak with a retention time of ~2.25 min, probably indicating the presence of ethylene (in the 'blue' and 'red' GC traces); an ethylene standard (prepared commercially - 10 ppm in N₂) was also analysed, but, unfortunately, was not seen above the background noise. The detection sensitivity of the equipment's FID is either insufficient for detection of the standard (requiring a more concentrated standard), or the standard was prepared poorly by the manufacturer. The very large peaks in the gas chromatograms, corresponding to ethylene products, are not present in controls with no enzyme or no L-Arg present, suggesting that formation of the substance is enzyme- and L-Arg-dependent. The addition of L-histidine to the reaction mixture, previously reported to increase the rate of ethylene production,^{241,242} appeared to have no impact on the yield of ethylene produced; the GC chromatogram for this reaction (blue) is almost identical to the spectrum measured for an EFE reaction without L-His (red) (Figure 4.15 a). *N*^G-Acetyl-L-Arg was also tested as an EFE substrate; a small peak (above noise levels) at the same retention time as ethylene (~2.25 min) was observed (green trace - Figure 4.15 b), suggesting *N*^G-acetyl-L-Arg could be a EFE substrate. Further testing, perhaps using a larger scale reaction (for higher turnover), is required to confirm whether this is the case, or whether *N*^G-Ac-L-Arg is being hydrolysed to L-Arg.

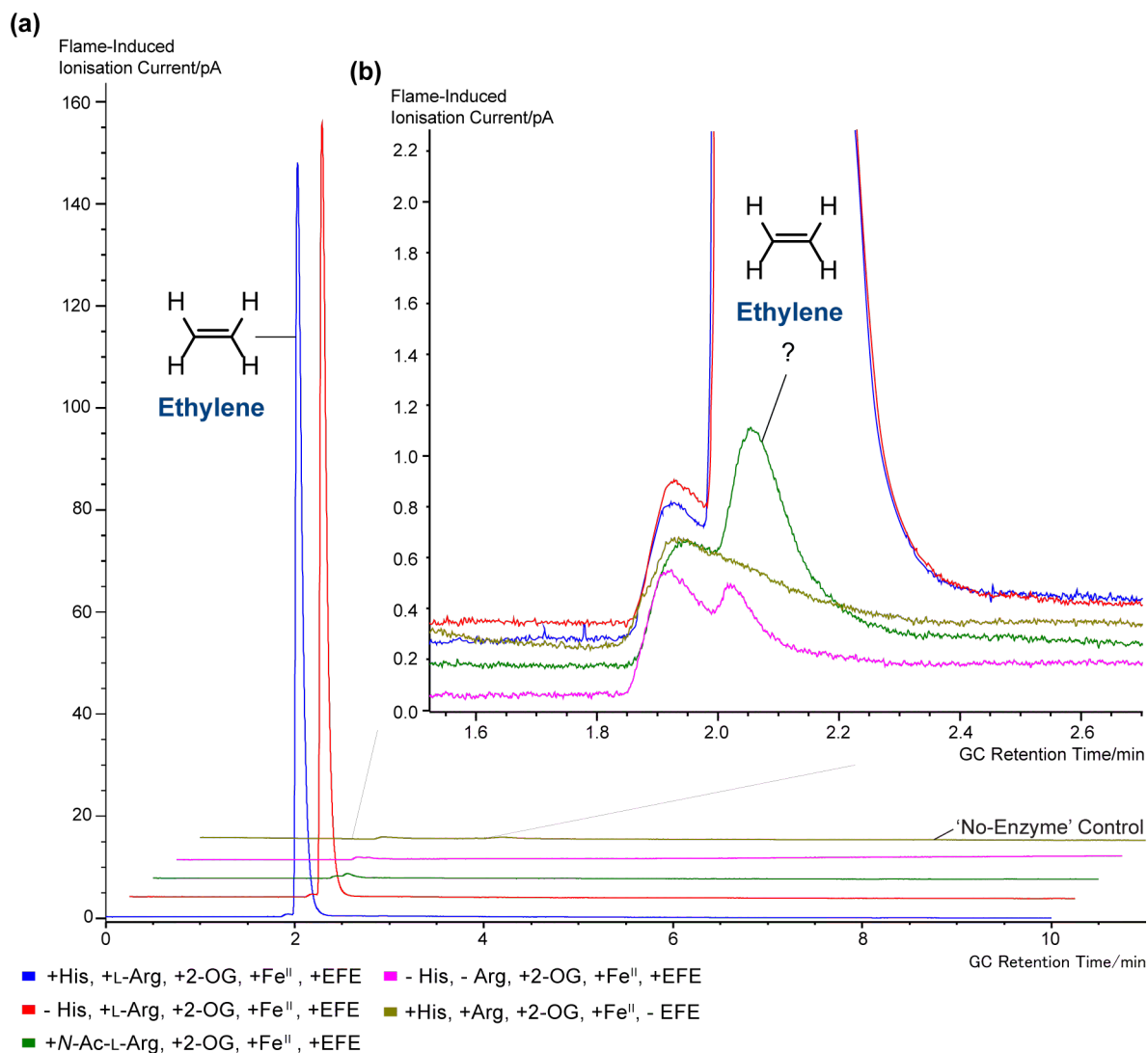


Figure 4.15: GC analysis of EFE reaction headspaces; signals were detected using an FID. These reactions were carried out with Dr Zhihong Zhang.

(a) EFE Reactions with various combinations of reactions components added.

(b) A small peak with the same retention time (~2.25 min) as ethylene was detected for a reaction using an *N*^G-Acetyl-L-Arg substrate, suggesting it may be a substrate of EFE.

A 2-oxoglutarate (2-OG) analogue, 2-oxoadipate (2-OA), was then tested for EFE turnover; the carbon skeleton this substance is one carbon atom longer than 2-OG. Interestingly, product peaks (retention time = ~ 3 min) were found by GC-analysis for reactions using 2-OA as a substrate (Figure 4.16), suggesting 2-OA may be a substrate of EFE; this peak was missing from ‘no-enzyme’ control mixture GC spectra, suggesting it is EFE-produced. Given that 2-OA is one carbon atom longer than 2-OG, it is proposed that the product is a 3-carbon species, such as cyclopropane, as opposed to 2-carbon ethylene. It is possible that the potential cyclopropane product could be formed by an analogous mechanism in which the C—COOH bond and C₂—C₃ bond of 2-OA align in a pseudo-antiperiplanar fashion (Figure 4.17 b), mimicking the antiperiplanar alignment of the corresponding bonds in 2-OG (Figure 4.17 a). If 2-OA were to align this way, the σ (C—COOH) HOMO (highest-occupied molecular orbital) could overlap with the σ^* (C₂—C₃) LUMO (lowest unoccupied molecular orbital), allowing a migration/fragmentation reaction to be stereoelectronically favourable (Figure 4.17 b); in the case of 2-OA, on decarboxylation, the C-5 could react with the C-3 position, forming a 3-membered, cyclopropane ring, rather than forming a double bond with the adjacent carbon atom, as occurs in the reaction of 2-OG with EFE (Figure 4.17 a). At the time of writing this thesis, another ethylene standard and another machine (GC/MS) is being considered for more detailed ethylene detection; it is hoped that the observations recorded in thesis are confirmed and further details are uncovered, furthering our understanding of the ethylene-forming pathway of EFE.

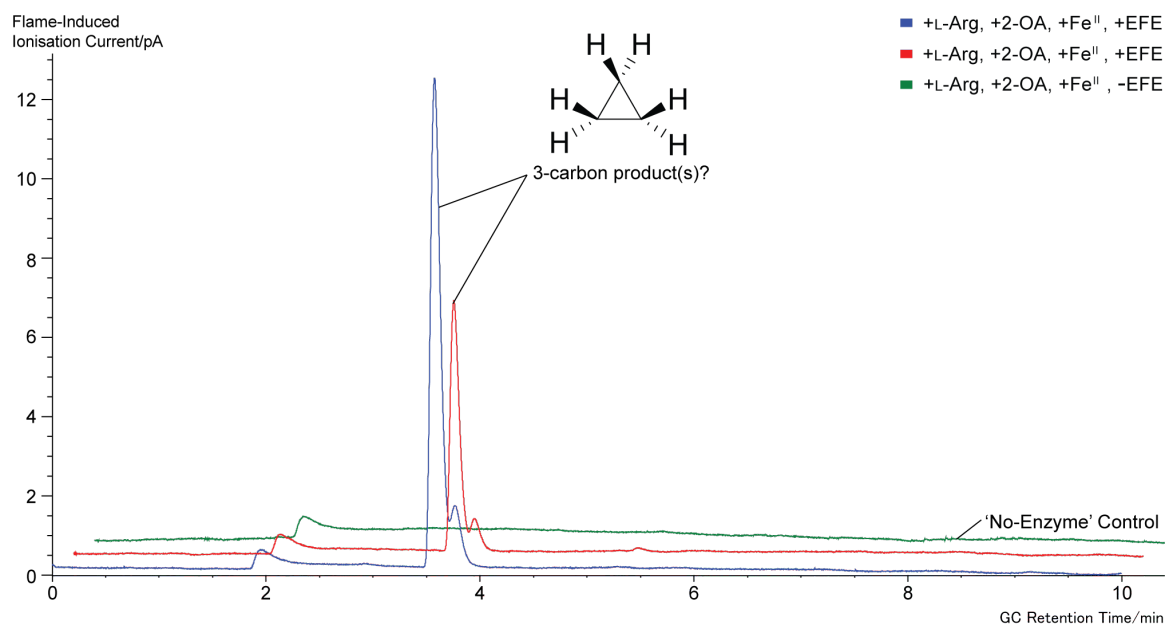


Figure 4.16: GC analysis of EFE reactions using 2-oxoadipate (2-OA): signals were detected using an FID. These reactions were carried out with Dr Zhihong Zhang.

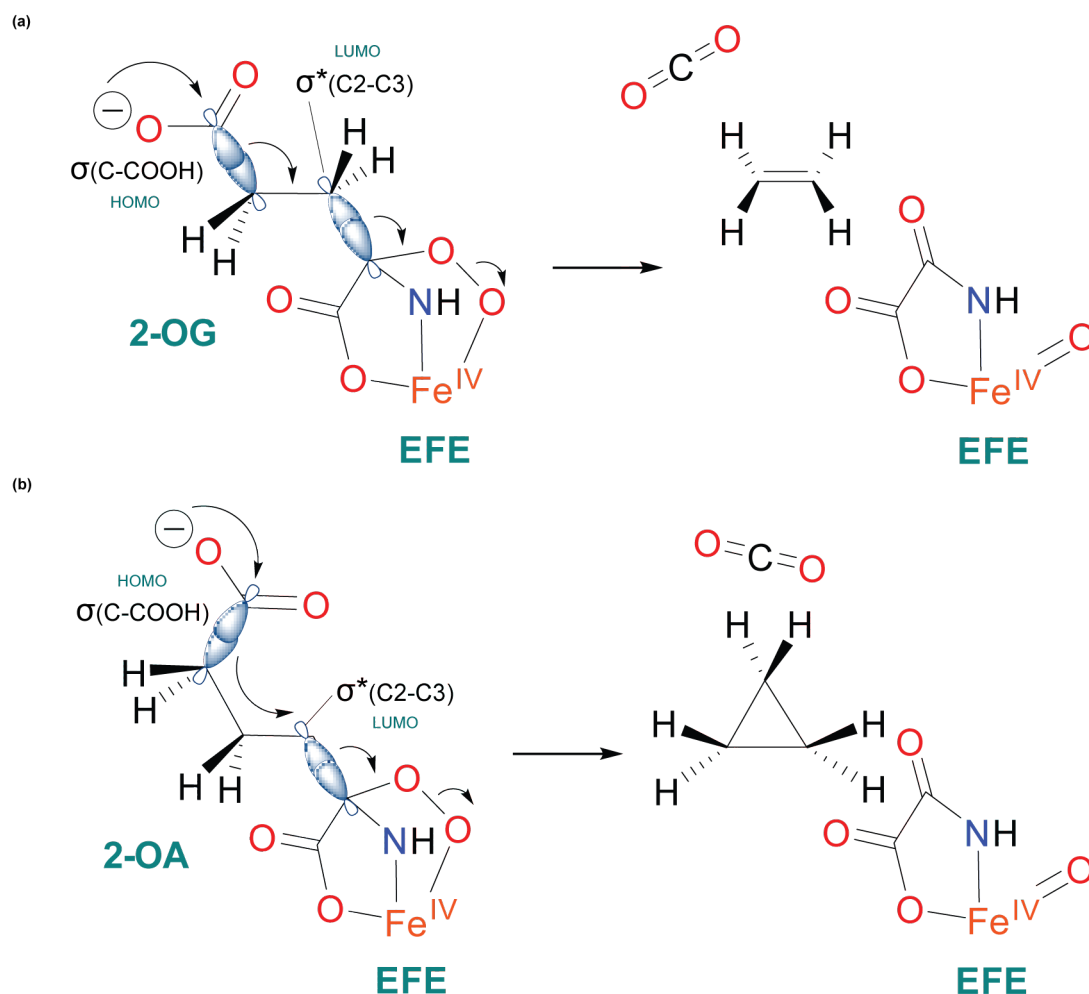


Figure 4.17: Proposed fragmentation mechanism of EFE-catalysed fragmentation of (a) 2-oxoglutarate and (b) 2-oxoadipate (2-OA) (a compound analogous to 2-OG): stereoelectronic considerations.

As well as 2-OG analogues of different lengths, it would be interesting to test whether EFE can accept 2-OG-based substrates with substitutions at the C3 and C4 positions (Figure 4.18); there is a chance that relatively small substituents may still be accommodated by EFE, if these substituents do not affect the binding mode of the substrate. Since the proposed, concerted decarboxylative fragmentation of 2-OG (to produce ethylene) requires antiperiplanar $\text{C}_2\text{—C}_3$ and $\text{C}_4\text{—C}_5$ bonds (on stereoelectronic grounds), it may be the case that EFE can produce unsaturated compounds with specific (E/Z) stereochemistry from C3/C4-substituted

2-OG compounds with particular stereochemistries (Figure 4.18). The products of experiments using C3/C4-substituted 2-OG analogues may therefore improve our mechanistic understanding of 2-OG fragmentation; products with specific stereochemistries produce from specific stereoisomers may suggest that the fragmentation of 2-OG occurs by a concerted mechanism; deviation from stereoselectivity by EFE may suggest a mechanism which is step-wise.

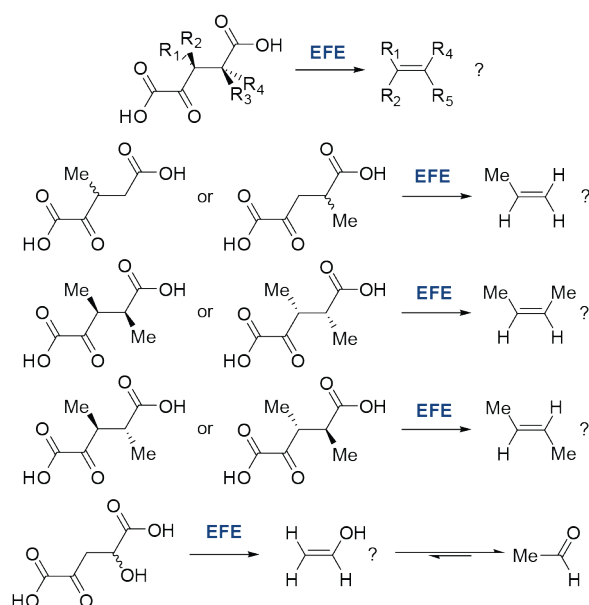


Figure 4.18: Proposed potential EFE substrates: different 2-OG analogues could be transformed into specific E/Z stereoisomers.

4.3.5 EFE Structural Studies

Protein crystallisation experiments were carried out in the hope that an X-ray crystal structure for EFE could be solved. EFE crystal structures may eventually also be useful tools in engineering EFE variants with different properties; for example, it may be possible to produce EFE variants with new substrate selectivities, accepting new substrates, or a variant which no longer requires L-Arg to catalyse ethylene production (such an EFE variant may be useful in industry, since fewer reagents would be required, saving resources and capital).

Initial EFE crystallisation trials were carried out using the purest available EFE protein batches to improve the chances of high-quality crystal formation (*i.e.* crystal which diffract well); a number of commercial screens, with a variety of crystallisation conditions, were tested using a sitting drop method (a crystallisation robot was used to set crystal trays). Unfortunately, most crystallisation drops contained precipitate, and those that didn't were mostly completely clear; microcrystals were formed in a few wells, however, suggesting EFE crystallisation may be possible. It was hoped that improved EFE sample purity would improve EFE crystallisation. Analysis of gel filtration UV_{280 nm}-absorbance chromatograms suggested that the purified EFE protein may be adopting more than one oligomerisation state (*i.e.* existing as monomers and dimers); EFE samples would be more homogeneous if EFE monomers were separated from EFE dimers, and would be more likely to crystallise.

Further improvements to the EFE purification and crystallisation method by my colleague, Dr Zhihong Zhang, lead to a high quality EFE protein crystals being formed; encouragingly, these crystals diffracted well (2.1 Å resolution) and a structure was solved by molecular replacement (by Dr Michael McDonough). Refinement of the EFE structure revealed that a BisTris molecule was bound to the Fe^{II} centre of the EFE active site (Figure 4.19), potentially mimicking the usual

bidentate 2-OG ligand binding mode in ODDs; BisTris buffer originated from the condition in which EFE crystallised.

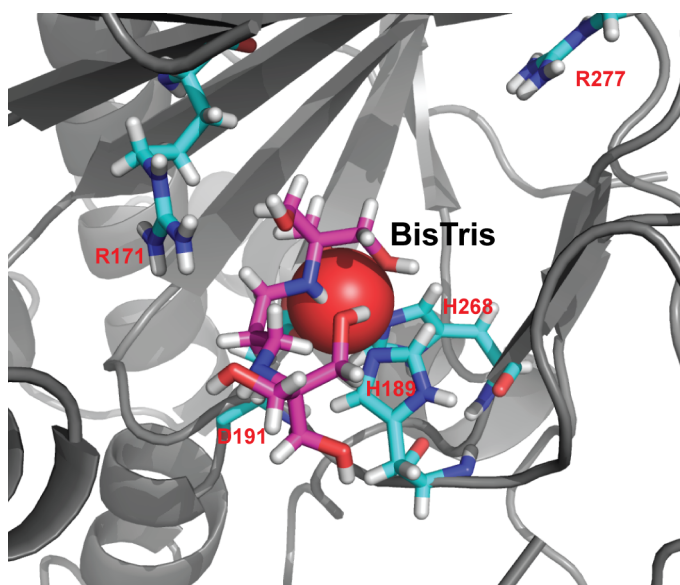


Figure 4.19: View from a crystal structure of EFE showing BisTris bound to active-site Fe^{II} .

Further EFE crystallisation screening by Dr Zhihong Zhang lead to the crystallisation of selenomethionine-containing EFE (to solve a structure from first principles - carried out by Dr Michael McDonough); the EFE protein was co-crystallised with 2-OG and Fe^{II} . Data processing and structure refinement produced an EFE structure showing 2-OG was bound to the active-site Fe^{II} in a bidentate manner (Figure 4.20). A comparison of the EFE structure containing a 2-OG ligand with the structure without 2-OG (and a BisTris molecule) shows that the binding of 2-OG to EFE may cause the side chain of an arginine residue, R171, to move (~ 3 Å) closer to 2-OG to form a hydrogen bond with C2-carbonyl of 2-OG (Figure 4.21). The further distance of R171's side chain from the Fe^{II} site in the BisTris EFE structure does not seem to occur to prevent a steric clash, since there would be sufficient space for R171 to adopt a similar position as in the EFE 2-OG structure; this suggests that 2-OG-binding induces a conformational change on more charge-based, chemical

grounds (*e.g.* inclusion of additional dipoles/charged). It is possible that the EFE arginine residue, R171, helps position the 2-OG cosubstrate in the correct orientation for ethylene production, or the maybe the charge distribution of a bound 2-OG molecule attracts L-Arg molecules, and promotes the binding of L-arginine substrate to EFE.

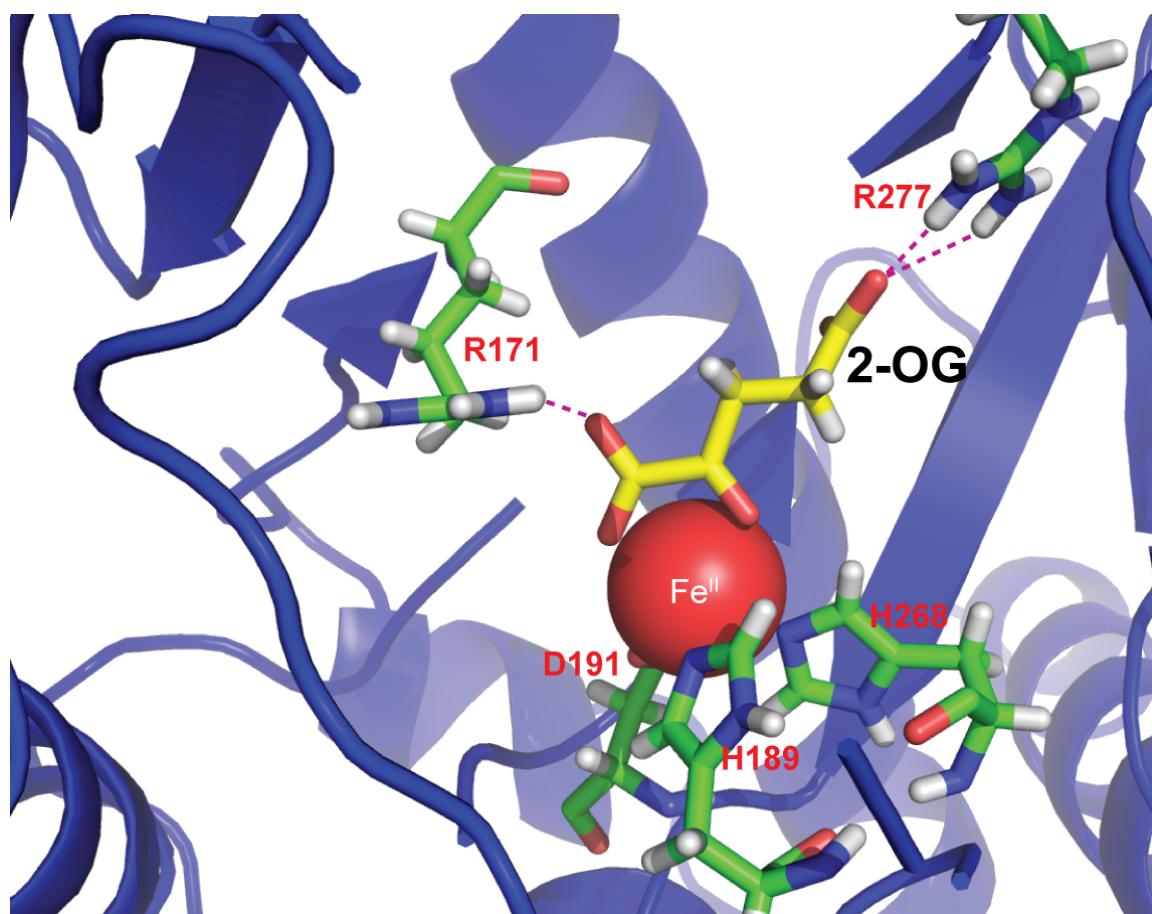


Figure 4.20: View of the active site of the EFE-Fe^{II}-(2-OG) complex crystal structure; key residues are shown.

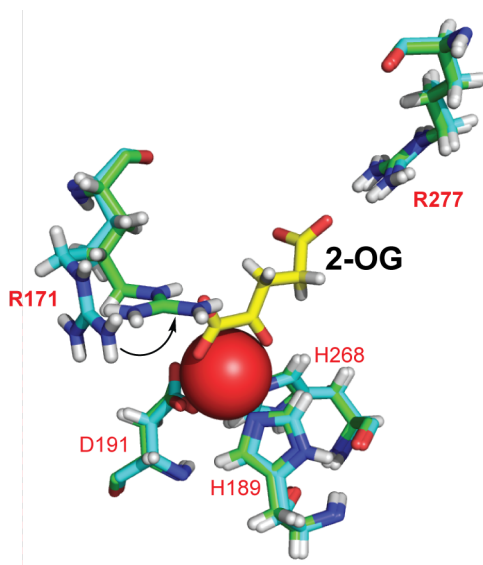


Figure 4.21: A comparison of active site residues of the EFE structures with 2-OG bound (green) and without 2-OG bound (blue): both structures were overlaid and analysed using PyMOL modelling software. Interestingly, R171 appears to move closer to the 2-OG site to form a hydrogen bond with 2-OG when 2-OG binds to EFE.

Interestingly, in the crystal structure of the EFE-Fe^{II}-(2-OG) complex, the sidechain of the 2-OG molecule did not adopt a coplanar conformation (*i.e.* the carbon and oxygen ‘backbone’ atoms were not all in the same plane) (Figure 4.23); a sterically favourable ‘flat’ 2-OG ligand is a common feature of many ODD enzymes, such as BBOX1, which shows a coplanar *N*-OG (*N*-oxalylglycine, a 2-OG analogue) in its crystal structure (PDB ID: 3MS5 - Figure 4.22). The 2-OG molecule in the EFE crystal structure deviates from planarity in two respects: the C₄—C₅OOH bond is twisted ~36° out of the C₁-C₂-C₃-C₄ plane of 2-OG (Figure 4.20 a), and the C₅-carboxylate group is twisted ~58° out of the C₃-C₄-C₅ plane of 2-OG (Figure 4.20 b). It is not fully clear whether these distortions of 2-OG away from planarity allow the carboxylate to interact with the guanidino group of R277, or whether these torsion angles lead to a more stable transition state for a Grob-like, decarboxylative fragmentation of 2-OG during ethylene formation. In terms of stereoelectronics, the decarboxylation of the C₅-carboxylate should be possible at any C₄—C₅OOH

bond-rotation angle, since the C5-lobe of the $\sigma^*(\text{C}_4\text{—C}_5\text{OOH})$ orbital is cylindrically symmetric about a line protruding out of the C5-end of the $\text{C}_4\text{—C}_5\text{OOH}$ bond (Figure 4.24); this means the carboxylate lone pair ($\text{n}(\text{O}^-)$) can overlap with the $\sigma^*(\text{C}_4\text{—C}_5\text{OOH})$ orbital to break the $\text{C}_4\text{—C}_5\text{OOH}$ bond at any carboxylate rotation angle (Figure 4.24).

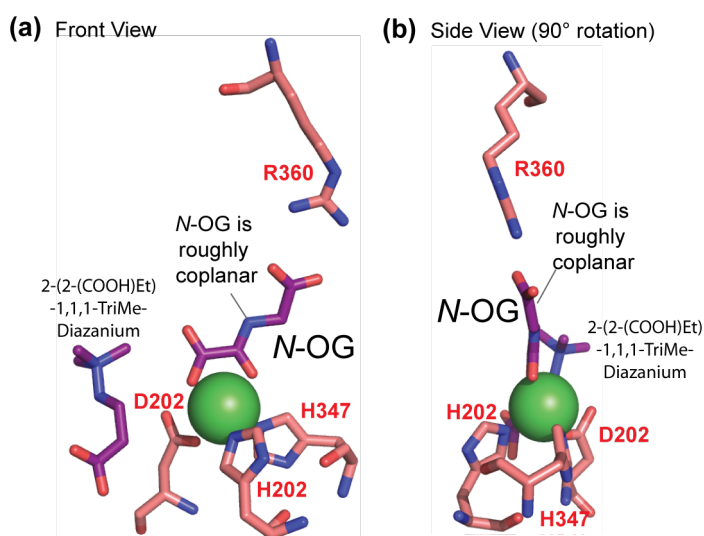


Figure 4.22: The active site residues of a BBOX1 as seen in a crystal structure (PDB ID: 3MS5) containing complexed *N*-OG (*N*-oxalylglycine, a 2-OG mimic), Ni^{2+} (in place of Fe^{2+}), and 2-(2-carboxyethyl)-1,1,1-trimethyl diazanium (2-(2-(COOH)Et)-1,1,1-TriMe-Diazanium, a substrate analogue).⁴¹² The *N*-OG is roughly coplanar, as can be seen if observed from the ‘front’ or (b) from a side-on view (90° rotation).

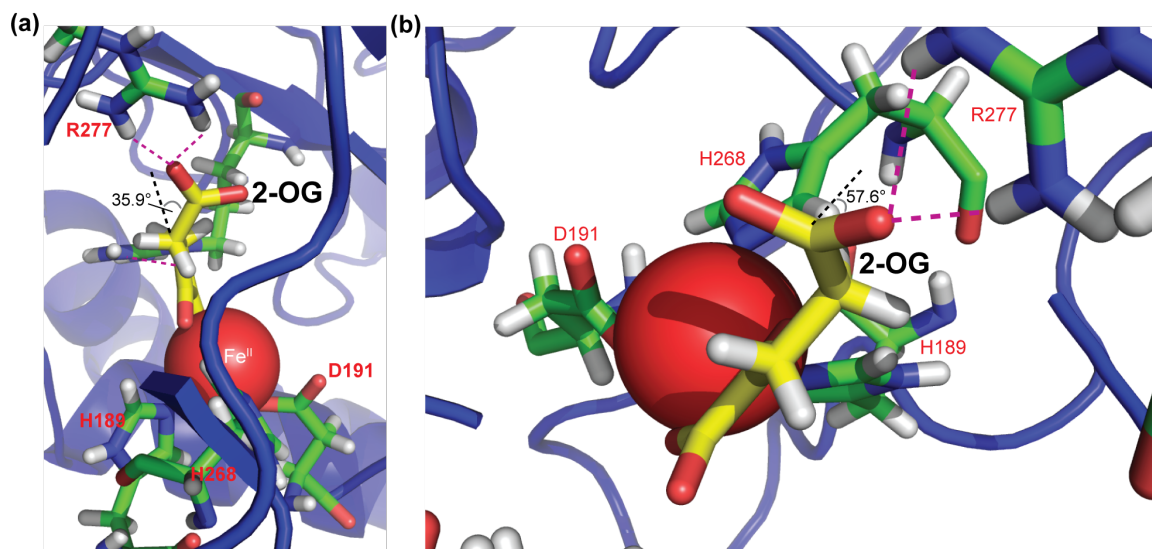


Figure 4.23: The deviation of EFE-bound 2-OG from a coplanar conformation: (a) the C_4-C_5OOH bond is twisted 35.9° out of the $C_1-C_2-C_3-C_4$ plane of 2-OG; (b) the C_5 -carboxylate group is twisted 57.6° out of the $C_3-C_4-C_5$ plane of 2-OG.

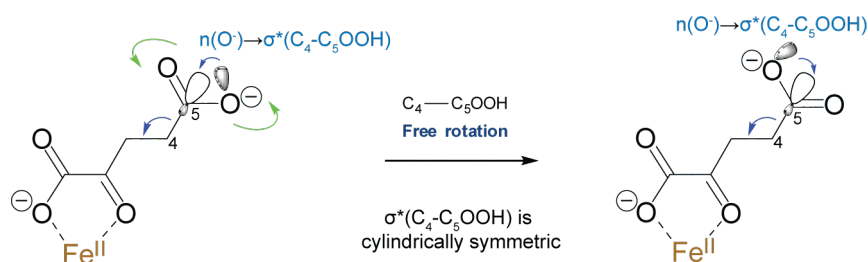


Figure 4.24: Stereoelectronic considerations of EFE-mediated 2-OG decarboxylation: the C_5 -lobe of the $\sigma^*(C_4-C_5OOH)$ orbital is cylindrically symmetric about a line protruding out of the C_5 -end of the C_4-C_5OOH bond, allowing the C_5 -carboxylate lone pair ($n(O^-)$) to overlap with the $\sigma^*(C_4-C_5OOH)$ orbital at any carboxylate rotation angle (to break the C_4-C_5OOH bond).

Although many ODD enzymes have ‘coplanar’ 2-OG molecules, there are other exceptions, such as the proline hydroxylase, *cisP4H*, from *Mesorhizobium loti* (Figure 4.25); in this *cisP4H* crystal structure (PDB ID: 4P7W),⁷ the C₃-C₄-C₅OOH plane of the EFE-bound 2-OG is almost 80° to the C₁-C₂-C₃ plane (Figure 4.25). The C5-carboxylate of 2-OG appears to be tilted to so that it points towards an arginine (R164) side chain to form an interaction in the *cisP4H* crystal structure, a similar feature to that seen in the 2-OG EFE structure (in which the same carboxylate points towards R277). If the 2-OG sidechain conformation affects the balance of succinate production versus ethylene production, it is that other ODDs make ethylene.

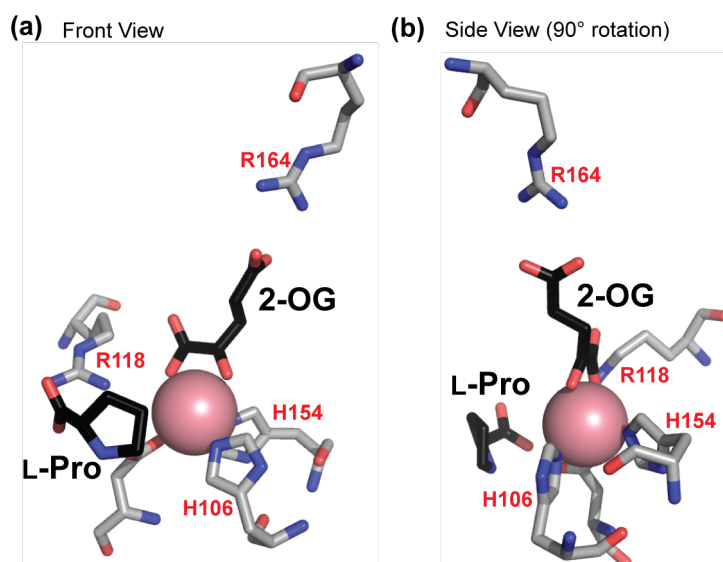


Figure 4.25: The active site residues of a *cisP4H* crystal structure (PDB ID: 4P7W) containing complexed 2-OG, Co²⁺ (in place of Fe²⁺), and *l*-proline.⁷ The 2-OG is not coplanar, as can be seen if observed from the ‘front’ or (b) from a side-on view (90° rotation).

Further crystallisation trials (carried out by Dr Zhihong Zhang) recently produced EFE crystals in complex with Fe²⁺, N-OG (a 2-OG analogue), and L-arginine; as before, the crystals were grown under anaerobic conditions to prevent oxidation of Fe(II) to Fe(III) by ambient molecular oxygen. X-ray diffraction experiments showed that these EFE crystals diffracted well (2.1 Å resolution), and an EFE structure with bound substrates was quickly solved (by Dr Michael McDonough) (Figure

4.26); molecular replacement calculations used previous EFE structures as a basis for structural models. The EFE-Fe(II)-(N-OG)-(L-Arg) complex exists as tetramer in the unit cell of its particular crystal form (Figure 4.26 a). Analysis of the active site residues of the EFE-substrates complex (Figure 4.26 b) reveals EFE reaction mechanistic details. Interestingly, Cys317, Val85, Thr86, and Arg316 help to anchor the L-arginine substrate in place in a manner, which points the C-5 position of L-Arg directly at the Fe(II) redox centre of EFE (Figure 4.26 b); this binding mode, therefore, explains the L-Arg C-5-hydroxylation selectivity of the EFE ferryl intermediate (in the succinate pathway). The EFE-substrates structure also shows Arg171 is still involved in 2-OG/N-OG binding, forming a hydrogen bond with the C-1 carbonyl oxygen atom of 2-OG/N-OG with its guanidino group [as it does in the EFE-Fe(II)-(2-OG) structure] (Figure 4.26 b); Arg171 does appear to move slightly to accommodate L-arginine binding, however. In moving, Arg171 can now form an interaction with Glu84, another residue which interacts with L-Arg, further stabilising the complex.

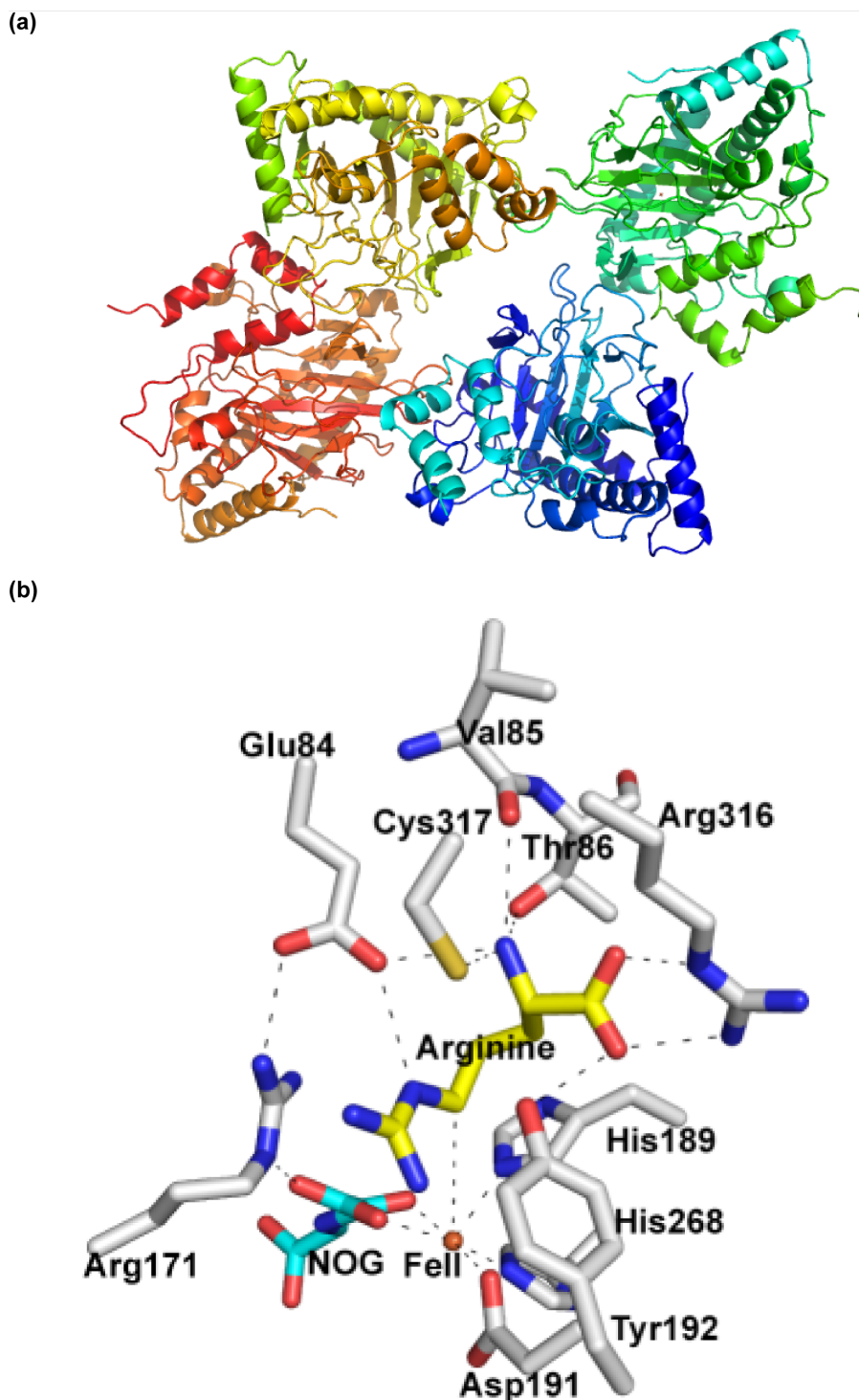
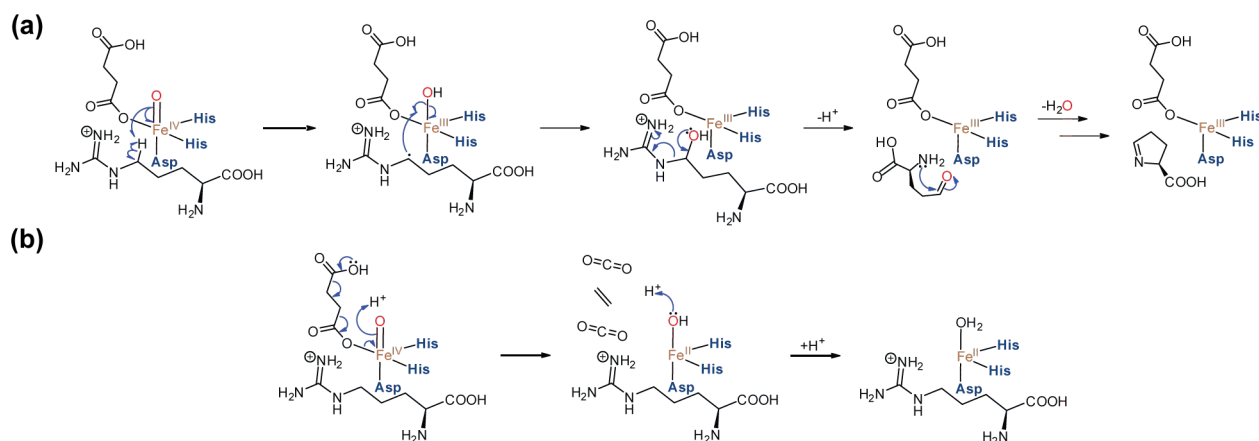


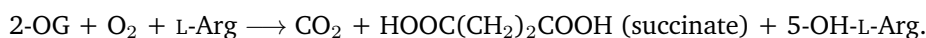
Figure 4.26: View from a structure of EFE in complex with Fe(II), *N*-OG, and L-Arg: (a) the unit cell EFE tetramer and (b) the active site residues showing the *N*-OG and L-Arg binding modes.

The crystal structure of EFE in complex with 2-OG and L-arginine suggests new mechanistic possibilities other than the ‘dual-circuit’ mechanism previously proposed²⁵⁵ given that: (i) there are dislike binding sites for 2-OG and L-arginine; (ii) the L-arginine is positioned for C-5-hydroxylation; and (iii) the 2-OG ‘sidechain’ conformation may be aligned for decarboxylation. The simplest mechanism that can be envisaged is probably one in which a ‘ferryl’ intermediate is formed, as occurs in the consensus 2-OG oxygenase mechanism (see Section 1.1.3).^{13–15} The ferryl intermediate then acts as a bifurcation point leading to two possible competing reaction pathways. In one instance, C-5-hydroxylation of L-arginine can occur, which, ultimately, results in the formation of L-P5C (Scheme 4.12 a); the ‘tetrahedral intermediate’ of 5-hydroxy-L-Arg collapses, eliminating the guanidino group, and cyclisation of the L-glutamate-5-semialdehyde produced subsequently generates L-P5C (Scheme 4.12 a). This ‘hydroxylation pathway’ has a ‘conventional’ ODD reaction mechanism. Alternatively, the succinate-ferryl complex can undergo a Grob-like, decarboxylative fragmentation reaction, which generates one ethylene molecule and two more CO₂ molecules (in addition to the one already produced from 2-OG) (Scheme 4.12 b); this reaction involves the concomitant reduction of Fe(IV) to Fe(II).

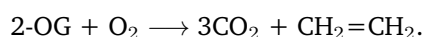


Scheme 4.12: Proposed EFE reaction pathways: (a) the formation of L-P5C via C-5-hydroxylation of L-Arg and (b) the Grob-like, decarboxylative fragmentation of succinate to produce CO₂ and ethylene.

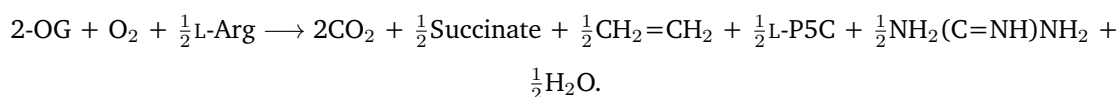
The stoichiometry in the first EFE process is thus proposed to be:



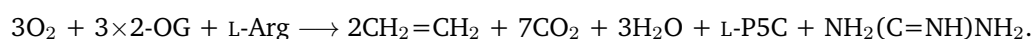
The 5-OH-L-Arg then cyclises to L-P5C, releasing guanidine ($\text{NH}_2(\text{C}=\text{NH})\text{NH}_2$) and water. In the second EFE process, in which L-Arg is a ‘cofactor’, the stoichiometry is:



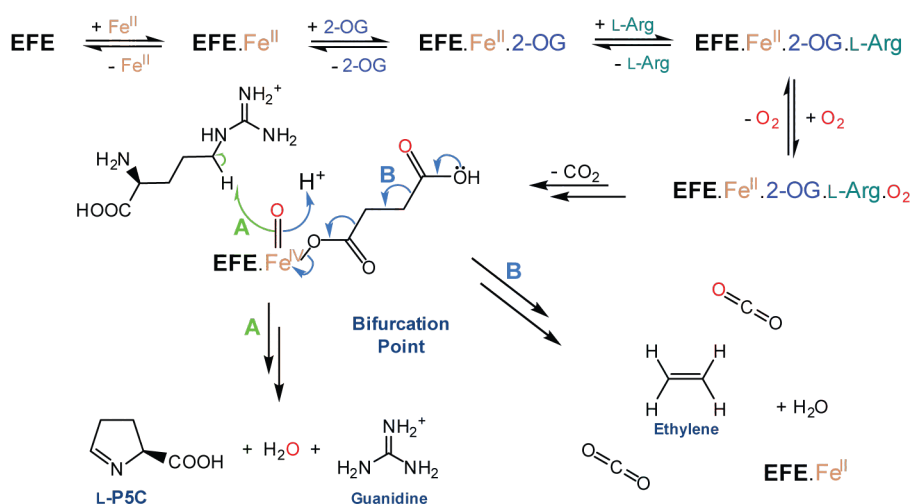
At present the relative rates of the two processes are not accurately known, but both reactions occur substantially. Assuming a 1:1 partition coefficient for the EFE pathways, the ‘average’ EFE reaction stoichiometry would be:



This differs from the reported overall stoichiometry for the EFE reactions:²⁵⁵



Therefore, it is proposed that EFE catalysis proceeds as follows (Scheme 4.13).



Scheme 4.13: Proposed binding order of substrates and cofactors in EFE bifurcating reactions: the EFE active site binds Fe^{2+} , then 2-OG, then L-Arg, then molecular O_2 .

Oxidative decarboxylation of 2-OG produces succinate and CO_2 ; the $\text{EFE} \cdot \text{Fe}^{\text{IV}} = \text{O} \cdot \text{Succinate}$ could then undergo two possible reactions: (a) the ferryl intermediate can hydroxylate C-5 of L-Arg, ultimately generating L-P5C and guanidine, or (b) the $\text{EFE} \cdot \text{Fe}^{\text{IV}} = \text{O} \cdot \text{Succinate}$ complex undergoes a Grob-like fragmentation reaction, forming 2CO_2 and ethylene.

4.4 Conclusions

Although more evidence needs to be collected, the work on EFE from *Pseudomonas syringae* pv. *phaseolicola* PK2 has provided substantial molecular insights into its mechanism. Firstly, the identity of the proposed L-P5C coproduct has been verified using MS and NMR data, and the succinate product has been detected by solution NMR; GC analysis (with FID for detection) also showed that EFE did, in fact, efficiently catalyse the production of ethylene. In addition, crystallographic studies produced 3 high-resolution (~ 2.1 Å) crystal structures of the same ethylene-forming enzyme (EFE) in complex with: Fe^{2+} (and BisTris); Fe^{2+} and 2-OG; and, Fe^{2+} , 2-OG, and L-Arg. Analysis of these EFE crystal structures has led to the proposal of a new mechanism, which involves the bifurcation of a ferryl intermediate. In one reaction pathway, the ferryl intermediate performs a radical C-5-hydroxylation reaction on L-Arg; this generates L-P5C (plus water) when guandine is lost from 5-OH-L-Arg, and the L-G-5-SA formed cyclises. Alternatively, in the second possible reaction pathway, the ferryl-bound succinate undergoes a Grob-like, decarboxylative fragmentation reaction, which produces ethylene and carbon dioxide (plus water). Discrete binding sites were identified for both L-arginine and 2-OG substrates, and the dual binding modes are consistent with the reported structure activity relationships,^{241, 242, 255, 407} as well as those described here; this includes the potential formation of cyclopropane from 2-oxoadipate (2-OA). The EFE structural work will form the basis for mutagenesis studies, which aim to improve the efficiency of ethylene production by EFE using a L-Arg cofactor (*i.e.* rational design); moreover, EFE variants may be generated, which produce ethylene without any L-Arg being required, or, which produce ethylene using another compound (*e.g.* amino acids) in place of L-Arg. The latter could be interesting with a view to ‘activating’ ethylene production or consumption 2-OG consumption in cells; such variants could be use-

ful as research tools for generating ethylene or controlling 2-OG levels. Finally, it would be interesting to generate EFE variants, which are specifically designed for producing alkenes other than ethylene; these may include industrially important compounds, such as vinyl chloride, propylene, or fluorinated alkenes.

Chapter 5

Materials and Methods

5.1 Materials

All materials were from the Sigma-Aldrich Chemical Co. unless stated otherwise. Analytical-grade water was produced from tap-fed water by a Millipore Elix 10 system, and purified further through a 0.22 μm filter by a Millipore MilliQ system.

5.1.1 Cloning and Molecular Biology

All molecular biology reactions were prepared carefully using autoclaved tips, sterile/ nuclease-free Eppendorf tubes whilst wearing gloves. Qiagen nuclease-free water was used as a reaction solvent for all molecular biology experiments. Nuclease-free water was prepared in a proprietary process, which yields deionized water that it is DNase, RNase, and nuclease-free; no chemical additives were used, such as diethylpyrocarbonate (DEPC).

Polymerase Chain Reaction (PCR)

Stratagene *PfuTurbo* DNA Polymerase, 10 $\times$ *PfuTurbo* buffer, and a 100 mM dNTP mix stock solution were used for all PCR experiments (*n.b.* the dNTP mix stock solution was diluted 10-fold for use).

Restriction Endonuclease Enzymes

All restriction endonucleases were provided by New England Biolabs Inc. (NEB) and stored at -20 °C. Below is a list of restriction enzymes used.

Enzyme	Sequence Cut
BamHI-HF	5' ... G↓GATC C... 3'
	3' ... C CTAG↑G... 5'
BsaI	5' ... GGTCTC(N) ₁ ↓... 3'
	3' ... CCAGAG(N) ₆ ↑... 5'
HindIII	5' ... AA↓AGC TT... 3'
	3' ... TT CGA↑AA... 5'
NdeI	5' ... CA↓TA TG... 3'
	3' ... GT AT↑AC... 5'
PstI-HF	5' ... C TGCA↓G... 3'
	3' ... G↑ACGT C... 5'

DNA Ligations T4 DNA Ligase and 10×Ligase buffer was provided by Promega.

Ligation Independent Cloning (LIC)

T4 DNA Polymerase and 100 mM dGTP was purchased from NEB.

5.1.1.1 DNA gel electrophoresis

Tank Running Buffer

The electrophoresis tank running buffer solution (TAE - Tris-Acetate-EDTA) was prepared as a 50×TAE stock solution by combining the following chemicals (Table 5.1) in 1 L (total volume) of filtered distilled water.

Chemical	Final Concentration/M	Stock Concentration/M	Quantity
Tris	2.0	-	242 g
Acetic Acid	1.0	17.4 M (Glacial)	57.1 mL
EDTA	0.05	0.5	100 mL

Table 5.1: 1 L 50×TAE Tank Running Buffer Formulation (pH ~8.6)

The 1×TAE running buffer was prepared by diluting the 50×TAE stock solution with MilliQ deionised water (Table 5.2). For 1 L, the following mixture is prepared.

Component	Volume/mL
50×TAE	20
MilliQ dH ₂ O	980
Total	1000 mL

Table 5.2: 1 L TAE Running Buffer Preparation.

Sample Loading Dye

The following recipe is that used to produce 6×NEB Gel Loading Dye (Blue) (Table 5.3).

Chemical	Concentration
Tris-HCl, pH 8.0 at 25 °C	3.3 mM
EDTA	11 mM
SDS	0.017% (w/v)
Ficoll 400	2.5% (w/v)
Bromophenol Blue	0.015% (w/v)

Table 5.3: 100 mL Sample Loading Dye Formulation.

Agarose Gels

Initially, the mould was constructed. Agarose was then dissolved in 1×TAE buffer by heating the solution with a microwave. For 60 mL 1×TAE buffer, the following quantities were added (Table 5.4).

[Agarose]	Mass of Agarose
1% (w/v)	0.60 g
0.8% (w/v)	0.48 g

Table 5.4: Agarose DNA Gel Formulation.

The solution was left to cool (roughly 50 °C) and then the liquid was poured into the gel mould. Invitrogen SYBR Safe DNA gel stain (10,000× solution in DMSO) was carefully added to the solution in the mould (6 µL for 60 mL) and a pipette tip was used to evenly mix the dye throughout the gel solution. The well comb was quickly placed into the agarose solution and the gel is left to set. Once set, the gel comb was carefully removed and the end caps of the mould removed. The gel could then be placed in a electrophoresis tank for use.

5.1.2 Bacterial Growth Media

All growth media were prepared using purified, filtered water and sterilised by autoclaving for 20 minutes at 121 °C. Kanamycin sulfate or sodium ampicillin (Apollo Scientific) was added to all media after autoclaving for antibiotic selection; a final concentration of 50 µg/mL kanamycin sulfate or 100 µg/mL ampicillin (sodium salt) (50 mg/mL and 100 mg/mL stocks, respectively) was used for starter cultures and most other growths unless otherwise stated. Stratagene BL21-Gold (DE3) *E. coli* competent cells were used for protein expression and cloning, and Stratagene

XL10-Gold Ultracompetent Cells (*E. coli*) for cloning and mutagenesis. Competent cells were stored at -80 °C.

5.1.2.1 Bacterial Growth Media

The following ingredients were dissolved in MilliQ water, adjusted to pH 7.0, and then autoclaved; the quantities are for 1 L growth medium.²⁶¹

Lysogeny Broth (LB)

Substance	Quantity/g
Tryptone	10.0 g
Yeast Extract	5.0 g
NaCl	10.0 g

Table 5.5: 1 L LB Formulation.

2-YT Broth

Substance	Quantity/g
Tryptone	16.0 g
Yeast Extract	10.0 g
NaCl	10.0 g

Table 5.6: 1 L 2-YT Broth Formulation.

Super Optimal Broth (SOB)

Super Optimal Broth (SOB medium) is a nutrient-rich bacterial growth medium used for microbiological culture, generally of *Escherichia coli*. It is an adjusted version of LB medium (Table 5.7), which produces higher transformation efficiencies of plasmids when recovering bacteria are grown in it.⁴¹³

Substance	Quantity/g
Tryptone	20.0 g
Yeast Extract	5.0 g
NaCl	0.5 g

Table 5.7: 1 L SOB Formulation.

The mixture was autoclaved. Subsequently, 10 mL filter-sterilised 1 M MgCl_2 and 10 mL filter-sterilised MgSO_4 were added.

Super Optimal Broth with Catabolite Repression (SOC)

SOC is essentially SOB with 20 mM D-glucose;⁴¹³ it can be prepared by adding a small amount of filter-sterilised, concentrated D-glucose [usually 2M/20% (w/v)] to an SOB stock [1 mL of 2 M D-glucose or 2 mL of 20% (w/v) D-glucose in 100 mL SOB].

Terrific Broth(TB)

Substance	Quantity/g
Tryptone	20.0 g
Yeast Extract	5.0 g
NaCl	0.5 g
10×TB Salts (1 L)	
KH_2PO_4	23.12 g
K_2HPO_4	125.41 g

Table 5.8: 1 L TB Formulation. The tryptone, yeast extract, and NaCl were dissolved in 900 mL MilliQ dH_2O and autoclaved, and 100 mL 10×TB salts were added just before use. The TB salts were dissolved in 1 L MilliQ dH_2O and autoclaved.

ZYM-5052 Medium for Auto-induction (without trace elements)

It is recommended to use this medium (Table 5.9) for auto-induction of *E. coli* protein expression.²⁸³ The amount of kanamycin sulfate and ampicillin (sodium salt) (when appropriate) used should be 100 µg/mL and 200 µg/mL, respectively.

Medium Component	Substance	Concentration	Volume Add/mL
ZY	Tryptone	1% (w/v)	958
	Yeast Extract	0.5% (w/v)	
50×M	Na ₂ HPO ₄	50 mM	20.0
	KH ₂ PO ₄	50 mM	
	NH ₄ Cl	50 mM	
	Na ₂ SO ₄	5 mM	
50×5052	Glycerol	0.5% (w/v)	20.0
	D-glucose	0.05% (w/v)	
	α-Lactose	0.2% (w/v)	
1 M MgSO ₄	MgSO ₄	2 mM	2.00

Table 5.9: 1 L ZYM-5052 Auto-induction Medium Formulation. The components of the medium are prepared separately MilliQ dH₂O and autoclaved.

The recipes for the constituent parts of the medium are below. They are prepared separately and combined prior to inoculation with *E. coli*.

Medium Component	Substance	Concentration
ZY	1% (w/v) Tryptone	10.0 g/l
	0.5% (w/v) Yeast Extract	5.0 g/l

Table 5.10: 1 L ZF Formulation.

Tryptone and yeast extract were dissolved in the necessary amount of water (958 mL) to make up 1 L total of ZYM-5052 medium.

Medium Component	Substance	Concentration
50×M	2.5 M Na ₂ HPO ₄	340 g/l
	2.5 M KH ₂ PO ₄	355 g/l
	2.5 M NH ₄ Cl	340 g/l
	0.25 M Na ₂ SO ₄	35.5 g/l

Table 5.11: 1 L 50×M Solution Formulation.

The powders were added in sequence (Table 5.11) to 700 mL water and stirred until all salts have been dissolved. The solution was then topped up to 1 L and autoclaved. Occasionally, crystals appeared, but they were re-dissolved in the microwave. The pH of a 20-fold dilution in water should be pH ~6.75.

Medium Component	Substance	Concentration
50×5052	25% (w/v) Glycerol	250 g/l
	2.5% (w/v) D-Glucose	25 g/l
	10% (w/v) α-Lactose	100 g/l

Table 5.12: 1 L 50×5052 Formulation.

The compounds (Table 5.12) were added in sequence to 730 mL water and stirred until all components had been dissolved. Lactose is slow to dissolve and it can take two hours or more at room temperature; the process was often sped up by heating the solution in a microwave. The medium was then filter sterilised.

Medium Component	Substance	Concentration
500×MgSO ₄	1 M MgSO ₄	24.65 g/(100 mL)

Table 5.13: 1 L MgSO₄ Formulation.

For a 100 mL stock solution, 24.65 g MgSO₄ was dissolved in 87 mL water, topped up to 100 mL with MilliQ dH₂O, and then filter sterilised or autoclaved.

5.1.3 Protein Purification

5.1.3.1 Buffers: Stock solutions

All buffers were prepared freshly before each purification using stock solutions; their pH was checked with an electrical pH meter, before being vacuum filtered through a 0.22 μm cellulose membrane.

Stock Solution	Concentration/M	pH
Tris-HCl	1.0	7.5
HEPES-NaOH	1.0	7.5
MES-HCl	0.5	6.0
MES-HCl	0.5	6.5
Sodium Phosphate	1.0	7.0
NaCl	4.0	-

Table 5.14: Stock solutions for buffer preparation. Buffer stocks were pH adjusted dropwise with concentrated or dilute hydrochloric acid (HCl) and 5 M Sodium hydroxide (NaOH), apart from phosphate buffers which were prepared by adding the appropriate ratios of phosphate salts to the solutions.

5.1.3.2 SDS-PAGE Gel Formulations

Gels were cast using Biorad plates and casting stands. The separating gel mixtures (Table 5.15) were combined carefully with gloves on and the polymerisation was initiated with the addition of ammonium persulfate (APS) solution. After a quick swirl of the polymerisation mixture, the solution was pipetted into the space between the casting plates with a P1000 pipette (when the plates were in the casting stands), leaving a gap of air ~ 1 cm from the top of the gel-casting volume. Iso-propanol was added to the air space at the top to flatten the gel solution to ensure

that the gel boundary would be flat. The separating gel was left to set for ~15 minutes. When the separating gel was fully set, the isopropanol layer could be poured off to waste, leaving a the flat edge of the rectangular separating gel exposed. The stacking gel solution was then prepared (again polymerisation was initiated by APS) and quickly pipetted on top of the separating gel layer, filling the remaining space at the top of the casting volume. A well comb was then quickly, but gently, inserted into the stacking gel solution, using a paper towel to wipe away excess, overflowing stacking gel solution. The stacking gel was then left to set with the comb in place to mould the sample wells (typically ~15 minutes).

Component	Volume/mL	
	12.5% (w/v) Acrylamide	15.0% (w/v) Acrylamide
MilliQ dH ₂ O	1.4	1.1
30% (w/v) Acrylamide	2.1	2.5
1.5 M Tris-HCl, pH 8.8	1.3	1.3
10% (w/v) SDS	0.05	0.05
10% (w/v) APS	0.05	0.05
TEMED	0.002	0.002

Table 5.15: Tris-Glycine SDS-PAGE Separating Gel Formulations. The volumes are approximately for 1 gel (5 mL); if multiple gels are being casts, the volumes can be scaled appropriately.

Component	Volume/mL
MilliQ dH ₂ O	1.4
30% (w/v) Acrylamide	0.33
1.5 M Tris-HCl, pH 8.8	0.25
10% (w/v) SDS	0.02
10% (w/v) APS	0.02
TEMED	0.002

Table 5.16: 5% (w/v) Acrylamide Tris-Glycine SDS-PAGE Stacking Gel Formulation. The volumes are approximately for 1 gel (2 mL) (plus a small excess); if multiple gels are being casts, the volumes can be scaled appropriately.

5.1.3.2.1 Gel Band Visualisation: Staining and Destaining

An SDS-PAGE gel was removed from its plates and cleaned with MilliQ dH₂O. It was then placed into a gently-stirring box/beaker containing Coomassie stain (Table 5.17) until the gel was evenly stained a deep blue. The solution could then be carefully heated (*e.g.* using a microwave) before being left the stir for a while in order to speed up the process. The gel was then removed from stain, rinsed with MilliQ dH₂O to remove excess stain solution, and then placed into fresh destain solution until most of the gel is clear (leaving only stained bands).

Coomassie Stain Solution

Chemical	Quantity
Coomassie Brilliant Blue R-250	1.0 g
Glacial Acetic Acid	50 mL
Methanol	500 mL
MilliQ dH ₂ O	450 mL

Table 5.17: 1 L Coomassie Blue Stain Formulation.

Destain Solution

Chemical	Quantity
Glacial Acetic Acid	100 mL
Methanol	300 mL
MilliQ dH ₂ O	600 mL

Table 5.18: 1 L Destain Solution Formulation: 30%(v/v) MeOH, 10% (v/v) AcOH in dH₂O.

5.1.3.3 FPLC: Columns and Equipment

5.1.3.3.1 Protein Liquid Chromatography (FPLC)

Amersham Pharmacia Biotech ÄKTA FPLC systems were used to purify recombinant proteins from the soluble fraction of *E. coli* cell lysate. UNICORN 5.1 software was used to control the FPLC equipment. Two very similar system configuration were used:

1. Amersham Pharmacia Biotech ÄKTA Purifier FPLC System

The system comprises:

- (a) **pH/C-900** - an on-line monitor of pH and electrical conductivity.
- (b) **UV-900** - an on-line monitor of UV absorption.
- (c) **P-900** - a liquid pump system for chromatography.
- (d) **Frac-950** - an automated fraction collector with advanced functionality.
- (e) **INV-907 Valve** - a motorised, rotary, 7-port valve (3 positions) for loading a sample loop, injecting the sample onto the column, and washing the system pump.

- (f) **3×PV-908 Valves** - motorised, rotary, 8-port valves for selection of columns, samples, or buffers.

2. Amersham Pharmacia Biotech ÄKTA FPLC System

The system comprises:

- (a) **UPC-900** - a high precision, on-line monitor of UV absorption, electrical conductivity, and pH.
- (b) **P-920** - a liquid pump system for chromatography.
- (c) **Frac-920** - an automated, circular fraction collector.
- (d) **INV-907 Valve** - a motorised, rotary, 7-port valve (3 positions) for loading a sample loop, injecting the sample onto the column, and washing the system pump.
- (e) **2×PV-908 Valves** - motorised, rotary, 8-port valves for selection of columns, samples, or buffers.

5.1.3.3.2 Columns

The columns and/or the resins used were either manufactured by GE Healthcare or Pharmacia.

GE Healthcare 5 mL HisTrap Fast Flow

A GE Healthcare 5 mL HisTrap FF is a ready-to-use column, prepacked with precharged Ni Sepharose 6 Fast Flow for preparative purification of histidine-tagged recombinant proteins by immobilized metal ion affinity chromatography (IMAC). This column has a high protein binding capacity (approx. 40 mg/mL medium) and there is negligible leakage of Ni^{2+} ions. The resin also allows for high flow rates and aids small scale-up.

The following buffers were used to strip the resin (for cleaning purposes) and recharge the resin (Table 5.19).

Buffer Type	[NiSO ₄]/mM	pH	[EDTA]/mM
Charge	100	-	0.0
Strip	0.0	8.0	50.0

Table 5.19: GE Healthcare HisTrap FF (Ni-Sepharose) Maintenance Buffers.

Stripped resin was occasionally cleaned with 5 mL NaOH (hydrolyse any non-specifically-bound protein impurities) or 2 M NaCl (remove hydrophobically-bound impurities). The column was then washed with a large excess of MilliQ deionised water (~35 mL) before being recharged. The column can be equilibrated with 5-10 column volumes of buffer (~35 mL was typically used).

GE Healthcare 65 mL HiLoad 26/10 Phenyl Sepharose High Performance

The HiLoad™ 26/10 Phenyl Sepharose HP column is prepacked for preparative hydrophobic interaction chromatography. Phenyl Sepharose HP is based on a matrix of highly cross-linked agarose beads, which have been derivatized with phenyl groups. These are attached to the agarose matrix via uncharged, chemically stable ether linkages. The result is a hydrophobic medium with no ionic properties. The column is cleaned by washing it (maximum flow of 4.3 mL/min) with two column volumes of water, four column volumes of 1 M NaOH (removal of precipitated proteins), followed by four column volumes of water. The column is equilibrated using 1 column volume of low ionic strength buffer followed by two column volumes of high ionic strength buffer (binding/start buffer).

Pharmacia (30 mL & 250 mL) DEAE Sepharose Fast Flow

The DEAE Sepharose Fast Flow is a weak anion-exchanger based on the well established Sepharose Fast Flow ion-exchange platform, extensively used for preparative

protein separations. The hydrophilic nature of the base matrix ensures low levels of non-specific binding leading to low levels of host cell-derived impurities in the elution pool. DEAE Sepharose Fast Flow resin has high chemical stability, allowing well proven cleaning-in-place (CIP) and sanitization protocols. The column is cleaned by washing it (maximum flow of 10 mL/min) with two column volumes of water, two column volumes of 1 M NaOH (removal of precipitated proteins), followed by three column volumes of water. The column is equilibrated using 1 column volume of high ionic strength buffer followed by two column volumes of low ionic strength buffer (binding/start buffer).

Pharmacia (8 mL) Mono Q Anion-Exchange Column

Mono Q is a strong anion-exchanger prepacked with MonoBeads in a Tricorn column. The column is an excellent choice for small-scale polishing in purification of proteins, peptides, and other biomolecules when high purity is required; monodisperse 10 μ m porous beads yield high resolution, dynamic capacity, reproducibility, and durability.

Pharmacia (320 mL) HiLoad 26/60 Superdex 75

The Pharmacia (320 mL) HiLoad 26/60 Superdex 75 is a prep-scale size exclusion column with a resin that resolves biomolecules in the molecular weight range of 3-70 kDa. Superdex prep grade resin is a composite matrix of dextran and highly cross-linked agarose. The steep selectivity of dextran and the high chemical and physical stability of agarose enable high resolution separations. The column is cleaned by washing it with 1 column volume of water, 1 column volume of 1 M NaOH, and then another column volume of water. The column is equilibrated with at least 1 column volume of gel filtration buffer (if already in water) prior to use.

5.1.3.3.3 Purification Buffer Formulations

Purification buffers were typically prepared in beakers by diluting stock solutions using MilliQ deionised water, pH-adjusted with an electronic pH meter (using NaOH or HCl solutions), and then 0.22 μM vacuum-filtered into a clean Schott bottle (glass, screw cap bottle; some buffers used for gel filtration experiments were ‘de-gassed’ by being placed under vacuum whilst a magnetic stirrer agitated the solution to help liberate bubbles from the solution.

pH Measurements were collected at 25 °C with either a Hanna Instruments HI 9321 Microprocessor pH Meter (5 mm diameter electrode) or a Jenway pH Meter 3305; instruments were calibrated to pH 7.0 and pH 10.0 with commercial buffer solution (for solutions being pH-adjusted in the pH 7.0-10.0), or to pH 7.0 and pH 4.0 (for being pH-adjusted in the pH 4.0-7.0 range). All electrodes and stirring equipment were heavily rinsed with deionised water between measurements to ensure contamination was not an issue.

Deuterated buffers for NMR experiments (50 mM Tris- d_{11} -DCl in D_2O) were carefully prepared in 100 mL quantities and stored at 4 °C in parafilm-sealed 15 mL falcon tubes; all equipment was heavily cleaned with deionised water prior to pH adjustments being made, and the electrode was carefully dried to prevent inclusion of water contamination into the solution.

Purification buffers were prepared using the following formulations.

Buffer Type	[MES]/mM	pH	[NaCl]/mM	[EDTA]/mM	[Glycerol]/mM
Binding/Lysis	50.0	6.0	0.0	0.2	10% (v/v)
Elution	50.0	6.0	500	0.2	10% (v/v)

Table 5.20: DEAE Sepharose FF Anion-Exchange Chromatography Buffer Components: *cis*P3H Type I Purification Step 1 (and intermediate steps for experimental *cis*P4H/*trans*P4H).

Buffer Type	[HEPES]/mM	pH	[NaCl]/mM	[Imidazole]/mM	[Glycerol]	[Tween 20]
Binding	50.0	7.5	500	10	20% (v/v)	0.1% (v/v)
Wash	50.0	7.5	500	30	10% (v/v)	0.0% (v/v)
Elution	50.0	7.5	500	400	10% (v/v)	0.0% (v/v)

Table 5.21: Proline Hydroxylase (*cis*P4H/*trans*P4H) and EFE Purification Buffers for Immobilised Metal (Ni^{2+}) Affinity Chromatography (IMAC).

Buffer Type	[Tris]/mM	[MES]/mM	pH	[NaCl]/mM
GF Buffer (NMR Studies)	50.0	-	7.5	200
GF Buffer	-	50.0	6.5	200

Table 5.22: Size-Exclusion Chromatography Buffer Components - Proline Hydroxylases

Buffer Type	[Tris]/mM	[HEPES]/mM	pH	[NaCl]/mM
GF Buffer (NMR Studies)	50.0	-	7.0	200
GF Buffer	-	50.0	7.0	200

Table 5.23: Size-Exclusion Chromatography Buffer Components - Ethylene-Forming Enzyme

Buffer Type	[Tris]/mM	pH	[NaCl]/mM
Binding	50.0	7.5	0.0
Elution	50.0	7.5	500

Table 5.24: MonoQ Ion-Exchange Chromatography Buffer Components - Proline Hydroxylases

Buffer Type	[Tris]/mM	pH	[NaCl]/M
Binding	50.0	7.5	3.0
Elution	50.0	7.5	0.0

Table 5.25: Phenyl Sepharose Hydrophobic Interaction Chromatography Buffer Components - Proline Hydroxylases

5.1.4 Crystallography

Solutions for crystal screens were prepared using a Perkin Elmer Janus 8 Gripper liquid handler. Crystal plates were laid using an Art Robbins Instruments Phoenix/RE machine; crystallisation drops were typically 300 nL total. Plates were incubated and monitored in a Rigaku CrystalMation storage and imaging station. X-ray diffraction data collection was carried out at the Diamond Light Source (Didcot, UK) using the tunable macromolecular crystallography (MX) beamline (I03). Data processing and crystal indexing was carried out using HKL-2000 software; molecular replacement solutions were generated using CCP4 software. Structure refinement was achieved using refmac (from the CCP4 Program Suite), CNS, and Phenix software.

5.1.5 Mass Spectrometry

5.1.5.1 LC/MS Equipment

Both analytical and preparative LC/MS was carried out using a system made up of a Waters 1525 μ Binary HPLC Pump connected to a Waters 2777 Sample Manager (autosampler unit with injection valve systems), and solution ESI-MS detection was performed using a Micromass Quattro *micro* API MS system. Samples were typically injected onto the system from mass 400 μ L spectrometry vials using a 100 μ L syringe controlled by the 2777 autosampler. Sample HPLC was performed using a SiELC Primesep 100 column (250 mm $\times$ 4.6 mm, 10 μ m pore size), or using a C18/NH₂ 5 μ m Spherisorb mixed-mode (250 mm $\times$ 9.4 mm) column in some preparative cases.

The mass spectrometer component of the system typically operated under the following settings (scanning for species in the range 50-1000 m/z).

Component	Voltages
Capillary/kV	2.00
Cone/V	15
Extractor/V	2
Temperatures/°C	
Source Temp	90-100
Desolvation Temp	350
Gas Flow/L h ⁻¹	
Desolvation	250
Cone	25.0

Table 5.26: LC/MS ES-Source Settings

5.1.5.2 GC Detection Equipment

Gas chromatography was performed on the Agilent Technologies 7820A GC System with a RESTEK Rt-Msieve 5A column (30 m×320 μm×10 μm film thickness) using an FID detector (Heater: 350 °C; Air flow: 400 mL/min; H₂ fuel flow: 30 mL/min; He makeup flow: 25 mL/min; Column He flow: 1.2 mL/min; Flame: 26 pA).

The 11 min method (Figure 5.1) starts with an oven temperature of 50 °C (held for 4 min); the temperature is then increased to 150 °C (at rate = 50 °C min⁻¹), and held here for a further 5 min (total run time = 11 min). The carrier gas (He) pressure is held at 7.72 psi for 4 min, increases to 11.13 psi over 2 min, and is then held for a further 5 min. The carrier gas (He) flow rate is held at 1.2 mL/min throughout the run. The split ratio is 5:1, with a split flow of 6 mL/min.

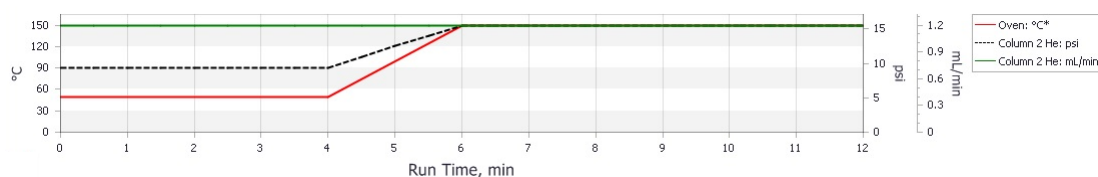


Figure 5.1: GC Method for Ethylene Detection: outline of 11 min GC profile.

Green line = He carrier gas flow rate (mL/min); Blue dashed line = He carrier gas pressure/psi; Red line = Oven temperature/°C.

5.1.6 NMR Spectroscopy

5.1.6.1 Real-Time Detection

^1H -NMR monitoring assays (*e.g.* enzyme reaction time courses) were performed using a 700 MHz Bruker AVIII 700 spectrometer equipped with an inverse TCI cryoprobe, optimised for ^1H -observation or a 600 MHz Bruker AVIII 600 NMR spectrometer (equipped with a 5 mm broadband (BBO)- $^{19}\text{F}/^1\text{H}$ Prodigy N_2 CryoProbe); instrument control and spectral processing was achieved using Bruker TopSpin 2.1 and then later TopSpin 3.0.

Early end-point assays were transferred to 5 mL tubes and analysed using a 500 MHz (^1H)/470.4 MHz (^{19}F) Bruker AVII 500 NMR spectrometer [equipped with a 5 mm TFI- $^1\text{H}/^{19}\text{F}(^{13}\text{C})$ probe].

5.2 Methods

5.2.1 Microbiology

Standard aseptic techniques were implemented throughout, using a Biological Safety Hood (Heraeus KJ12) to avoid contamination. All materials used were sterilised by autoclaving at 121 °C for 20 minutes or 0.22 µm filter sterilised (if any materials were heat sensitive). Tryptone, yeast extract, and agar (bacteriological) for use in culture media were obtained from Oxoid. Waste media was decanted off and treated with Antec Virkon to sterilise it.

5.2.1.1 *E. coli* Competent Cell Preparation

This procedure involves growth of *E. coli* in the absence of antibiotic (unless there is antibiotic resistance in your particular strain), so special care is taken to ensure everything is sterile. Before starting the procedure, Eppendorf vials are freshly autoclaved for cell storage, pipette tips for cell resuspension etc., as well as ~500 mL competent cell buffer [100 mM CaCl₂, 20% (v/v) glycerol]. Sterile 50 mL Falcon tubes for cell centrifugation.

Cells should be tested for sufficient transformation efficiency before using them for subcloning experiments.

Protocol

1. Streak a source of the *E. coli* cells being made competent onto an LB/2-YT-agar plate and grow colonies at 37 °C overnight.
2. Use one colony to inoculate 100 mL sterile growth medium (e.g. 2TY) and grow in a shaking incubator (220 rpm) at 37 °C overnight.

3. Pre-chill sterile 50 mL Falcon tubes for cell harvesting.
4. Centrifuge the cell cultures at 750 g and 4 °C for 5 minutes, and gently resuspend in 50 mL sterile competent cell buffer [containing 100 mM CaCl₂ and 20% (v/v) glycerol].
5. Incubate the cell suspensions on ice for 20 minutes.
6. Centrifuge the cell cultures again at 750 g and 4 °C for 5 minutes, and gently resuspend the cells in 2 mL of sterile competent cell buffer [100 mM CaCl₂ and 20% (v/v) glycerol].
7. Quickly, but gently, aliquot the cells in appropriate volumes (*e.g.* 40 µL) in sterile Eppendorf vials on ice.
8. Flash freeze the fresh competent cells with liquid nitrogen and immediately store the aliquots at -80 °C.

5.2.1.2 Heat Shock DNA Transformation Protocol

DNA fragments (*e.g.* plasmids, ligation products etc.) are forcibly translocated into bacteria by disrupting the integrity of their cell membranes; this is achieved with a sudden increase in temperature (a ‘heat shock’). The cells are then rapidly cooled after a specific period of time (depends on the cells) and kept cold for a couple of minutes in a cell recovery step.

Protocol: Stratagene XL-10 Gold Ultracompetent Cells

1. Cells are gently thawed on ice.
2. Gently pipette 40 μ L cells into a sterile, thin-walled PCR tube with autoclaved tips.
3. Add 2 μ L β -mecaptoethanol to the cells and swirl carefully.
4. Add 1-2 μ L (~20-50 ng) of transformant (DNA ligation product, DNA plasmid, or mutagenesis PCR DNA) and gently swirl the mixture.
5. Incubate the mixture on ice for 30min.
6. Heat pulse the cells at 42 °C for 30s using a water bath and then immediately placed on ice for cell recovery.
7. Incubate the calls on ice for 2 minutes.
8. Transfer the cells into a 15 mL falcon tube containing 360 μ L 2YT broth (or SOC medium for mutagenesis transformations), resuspending the cells.
9. Incubate the cell culture broth for 1 hour at 37 °C with 220 rpm shaking.
10. Plate the cells on a 2YT agar plate containing the appropriate antibiotic for the application [50 μ g/mL kanamycin sulfate or 50-100 μ g/mL ampicillin (sodium salt)] and/or other necessary substances [Plates used for pNIC-28-Bsa4 cloning contain 5% (w/v) sucrose]. *n.b.* Keep any ligation reactions at 4 °C (Cold room).
11. Incubate the plate at 37 °C.
12. Inspect the plate for colonies and screen colonies for successfully-cloned, correct sequences. If there are no colonies from ligation reactions, retrans-

form the chilled (4 °C) ligation reaction mixture (which was left to continue overnight).

Stratagene BL21-Gold (DE3)Competent Cells

The same procedure was used as above except that 45 μ L competent cells were used, and no β -mecaptoethanol mix was used. SOC medium was also used, heat pulses lasted 20s, and all cells were plated after centrifuge concentrating the cell recovery culture at (~180 g for 1 min and remove excess medium).

Colonies

Colonies were picked using sterile, autoclaved tips and used to inoculate 10 mL LB broth with 50 μ g/mL kanamycin sulfate/100 μ g/mL ampicillin (10 μ L 1000 $\times$ antibiotic stock solutions added to 10 mL LB/2-YT broth).

Glycerol stocks

100 μ L autoclaved glycerol was added to 900 μ L overnight culture and frozen at -80 °C. Glycerol stocks used to inoculate starter cultures were kept on dry ice to prevent thawing (freeze-thawing cycles make the cells unhealthy, reducing protein production).

5.2.1.3 Recombinant *E. coli* Growth for Enzyme Purification

Cells grown on terrific broth (TB) or ZYM-5052 auto-induction media (with kanamycin or ampicillin selection) were typically grown until roughly OD₆₀₀ 0.6, before dropping the cell incubation temperature to 18 °C (for pET-controlled overexpressions) or 15 °C (for pCOLD I, cold-shock-controlled expressions); induced cell cultures were then incubated overnight for a further 14 h before cell harvesting. TB-grown cells used 0.5 mM ITPG induction.

5.2.1.3.1 Cell Growth for Large-Scale Production of *cisP3H* Type I

5 L TB media (100 µg/mL kanamycin) cultures were inoculated with 10 mL of overnight 100 mL 2YT broth starter culture and incubated at 37 °C with 220 rpm shaking until the OD₆₀₀ reached roughly 0.65. Protein expression was then induced with 0.5 mM IPTG and cultures were then incubated at 20 °C for 16 h for *cisP3H* Type I [pET-24a(+)]. Cells were harvested using 1L centrifuge tubes and a Beckman Coulter™ Avanti J25 centrifuge with a JLA 9.1000 rotor spinning at 7,500 rpm for 7 min at 3 °C. Cell pellet was frozen at -80 °C. 4L-worth of pellet was resuspended in 150 mL FPLC binding buffer containing 0.5 mM TCEP and Roche EDTA-free Protease Inhibitor Cocktail Tablets at 4 °C (Cold room). Cells were lysed on ice by sonication using a Sonics Vibra Cell, applying ten 30 second bursts interrupted by 30 second pauses for 10 min total. A Beckman Coulter Avanti J25 centrifuge with a JLA-25.50 rotor was used to pellet the cell debris, spinning at 23,000 rpm for 20 min at 4 °C. Lysate containing *cisP3H* was treated with 5% (w/v) PEI solution on ice in the cold room (4 °C), adding it dropwise until the concentration reached 0.01% (w/v). The lysate was, once again, clarified by centrifugation in a Beckman Coulter™ Avanti J25 centrifuge with a JLA-25.50 rotor spinning at 23,000 rpm for 20min at 4 °C (pelleting DNA ppt and precipitated nucleoproteins. Lysate was filtered through 0.22 µm filters using 10 mL syringe.

5.2.1.3.2 Large-Scale Production of *cisP4H*, *transP4H*, and EFE

The cell growth methods for recombinant *cisP4H* and *transP4H* overexpression were the same as for *cisP3H* Type I, except that cultures were grown with 100 µg/mL ampicillin (sodium salt) (instead of kanamycin sulfate), and expression was induced at 15 °C with 0.5 mM IPTG; after 24 h of incubation, the induced cell were harvested by centrifugation. Additionally, the cell lysates containing *cisP4H* and

*trans*P4H proteins were not treated with PEI, as in the case of *cis*P3H Type I (it was deemed unnecessary and unlikely to improve protein purity), and lysates were just filtered (0.22 μ m pores) prior to being loaded onto the Ni-sepharose, 5 mL HisTrap FF column.

5.2.1.4 Purification Methods

5.2.1.4.1 Ni-Affinity Chromatography

5 mL GE Healthcare HisTrap FF chromatography columns were charged with 20 mL of 50 mM NiSO₄, washed with 10 mL of MilliQ distilled water, and washed with 50 mL of binding buffer prior to loading the cell lysate at 1.5 mL/min (230 mL). The column was then washed with binding buffer until the UV absorbance reached the baseline and then with 60 mL wash buffer. When all residual products of bacterial fermentation were eluted from the column, the protein of interest was eluted with a linear gradient of elution buffer (0-100% over 100 mL - 30mM to 400mM imidazole). Eluted protein was collected and fractionated; fractions containing purified protein (as determined by SDS-PAGE gels) were concentrated using Millipore Amicon Ultracell-10k concentrators in a Beckmann Allegra 21R centrifuge, spinning 4000 rpm at 4 °C.

5.2.1.4.2 Anion-Exchange Chromatography: DEAE Sepharose FF

After all the lysate/protein fractions were loaded onto the column, the column was washed with binding buffer until the UV-absorbance reached the baseline. The column was then washed with 45% elution buffer and 65% binding buffer (225 mM NaCl) to remove impurities. The washing occurred until the UV-absorbance reduced back to the baseline. The *cis* P3H Type I was then eluted from the column using a linear gradient of elution buffer from 45% to 100% over 2 and 1/3 column volumes

(roughly 80 mL) and 5 mL fractions were collected. The purity of protein was assessed using SDS-PAGE [15% (w/v) acrylamide], and the purest fractions were pooled and concentrated using 10 kDa cut off spin concentrators.

5.2.1.4.3 Anion-Exchange Chromatography: MonoQ

Protein was buffer exchanged with spin-concentrators to desalt the protein buffers solutions and then loaded onto a pre-equilibrated column. After all the protein fractions were loaded onto the column, the column was washed with binding buffer until the UV-absorbance reached the baseline. A linear, increasing NaCl gradient (0-500 mM - 0-100% elution buffer) was then applied to the column to elute the protein.

5.2.1.4.4 Size-Exclusion Chromatography

Gel filtration columns (Superdex 75, 200 resins) were cleaned with a column volume of distilled water and then equilibrated with a column volume of buffer prior to use. Concentrated fractions from previous purification steps (typically from Ni-affinity steps and DEAE Sepharose weak anion-exchange steps) were loaded into a 2 mL sample loop with a volume ≤ 1 mL (*i.e.* half the sample loop volume). The sample loop valve was altered, so that the flow of pumping buffer pushed the sample protein in the loop onto the column; roughly 25 loop-volumes were flowed through the loop (to loop all sample thoroughly) before the injection valve was altered to bypass the sample loop (and reduce system backpressure). 5 mL fractions were collected from 1/3 column volume onwards (typically at a retention time of 100 mL) with a fraction collector to collect protein samples, transitions being achieved by linear gradients.

5.2.2 Enzymology: Detection Assays

5.2.2.1 LC/MS-Detection Assays

The following formulations were prepared in Eppendorf vials and then later transferred to mass spectrometry vials after enzyme precipitation.

Reaction Component	Stock Concentration	Final Concentration	Volume/ μ L
2-Oxoglutarate (2-OG)	100 mM	10.0 mM	10.0 μ L
L-Proline/Substrate	50/100 mM	1.0 mM	2.0/1.0 μ L
Sodium Ascorbate	100 mM	1.0 mM	1.0 μ L
$(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$	5 mM	0.5 mM	10.0 μ L
<i>cis</i> P3H Type I/ <i>cis</i> P4H/ <i>trans</i> P4H	51/71/25 mg/mL	15 μ M (0.5 mg/mL)	1.0/2.0/0.7 μ L
50 mM MES-NaOH, pH 6.5	50.0 mM	-	(76.0/77.0) / (75.0/76.0) / (76.3/77.3) μ L
Total			100 μ L

Table 5.27: Proline hydroxylase LC/MS assay conditions for L-Pro-analogue detection.

Reaction Component	Stock Concentration	Final Concentration	Volume/ μ L
2-Oxoglutarate (2-OG)	100 mM	1.0 mM	1.0 μ L
L-Arginine	100 mM	1.0 mM	1.0 μ L
$(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$	5.0 mM	100 μ M	2.0 μ L
EFE	22 mg/mL	15.0 μ M	4.0 μ L
50 mM Tris-HCl/10 mM HEPES-NaOH, pH 7.5	50.0/10.0 mM	-	92.0 μ L
Total			100 μ L

Table 5.28: EFE LC/MS assay conditions for L-P5C detection.

Methanol-quenched reactions were chilled for 15 min at -20°C to aid protein precipitation; the precipitant was then pelleted by microcentrifugation (5 min at 9000 g), and transferred to a mass spectrometry vial for analysis.

Typically, 20-80 μ L of sample (from 100 μ L) was injected was analysis per run. The HPLC method involved the following transitions in solvent conditions.

Time/min	Flow/mL min ⁻¹	% Buffer A	% Buffer B
0.00	1.00	95.0	5.0
15.00	1.00	30.0	70.0
24.00	1.00	0.0	100.0
29.00	1.00	0.0	100.0
30.00	1.00	95.0	5.0
40.00	1.00	95.0	5.0

Table 5.29: SiELC Primesep 100 HPLC Method for LC/MS analysis of assay samples.

Buffer A: 0.1% (v/v) FA in H₂O

Buffer B: 0.1% (v/v) in ACN.

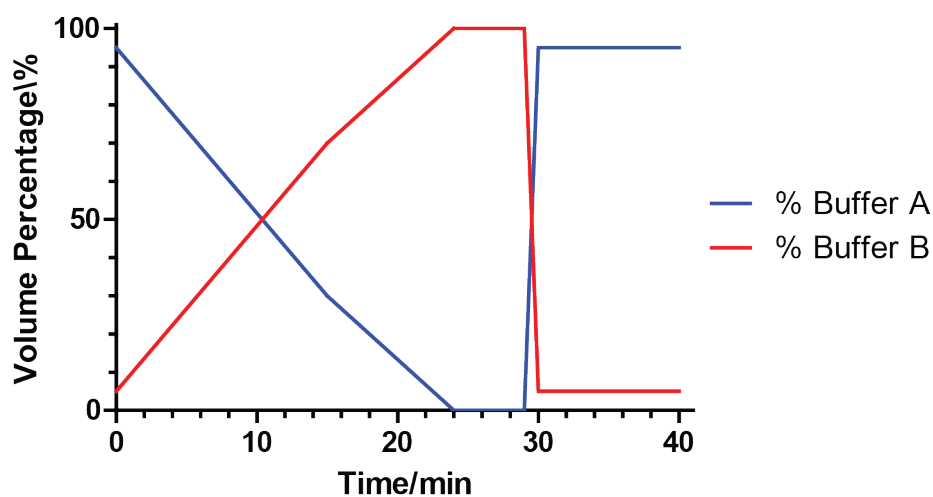


Figure 5.2: SiELC Primesep 100 HPLC Method for LC/MS analysis of assay samples.

Buffer A: 0.1% (v/v) FA in H₂O

Buffer B: 0.1% (v/v) in ACN.

5.2.2.2 GC-Detection Assay

The following EFE reaction mixture was prepared in an Eppendorf vial and quickly transferred to a sealed mass spectrometry vial for analysis. A gas syringe was used to inject gaseous samples into the GC system for analysis.

Reaction Component	Stock Concentration	Final Concentration	Volume/ μ L
2-Oxoglutarate (2-OG)	100 mM	10.0 mM	15.0 μ L
L-Arginine	100 mM	2.0 mM	3.0 μ L
DTT	100 mM	2.0 mM	3.0 μ L
(NH ₄) ₂ Fe(SO ₄) ₂	10 mM	0.2 mM	3.0 μ L
EFE	22 mg/mL	15.0 μ M	4.0 μ L
40 mM HEPES-NaOH, pH 7.5	40.0 mM	-	122 μ L
Total			150 μ L

Table 5.30: EFE GC assay conditions for ethylene detection.

5.2.2.3 Preparative Scale Reactions: Product Purification

Reactions were typically scaled up 2×10 -fold (2 mL total); protein from complete reactions was precipitated by adding cold methanol in a 1:1 ratio with the assay mixture (as in the analytical-scale reactions); the solutions were diluted 2-fold with water (to raise the freezing point), and then freeze-dried. Products were redissolved in 300 μ L of 20% (v/v) MeOH in water. Samples were then purified in 3/4 runs using 75-100 μ L, depending on sample retention properties (usually a lower volume was loaded for a weak binder).

The following HPLC methods were used for ‘polar’ (*e.g.* hydroxypipelic acids) and ‘very polar’ substances (*e.g.* hydroxyfluoro prolines), respectively.

Time/min	Flow/mL min ⁻¹	% Buffer A	% Buffer B
0.00	1.50	99.0	1.0
10.00	1.50	99.0	1.0
15.00	1.50	90.0	10.0
20.00	1.50	0.0	100.0
21.00	1.50	99.0	1.0
30.00	1.50	99.0	1.0

Table 5.31: SiELC Primesep 100 HPLC method for LC/MS purification of ‘polar’ proline analogues.

Buffer A: 0.05% (v/v) FA in H₂O

Buffer B: 0.1% (v/v) in ACN.

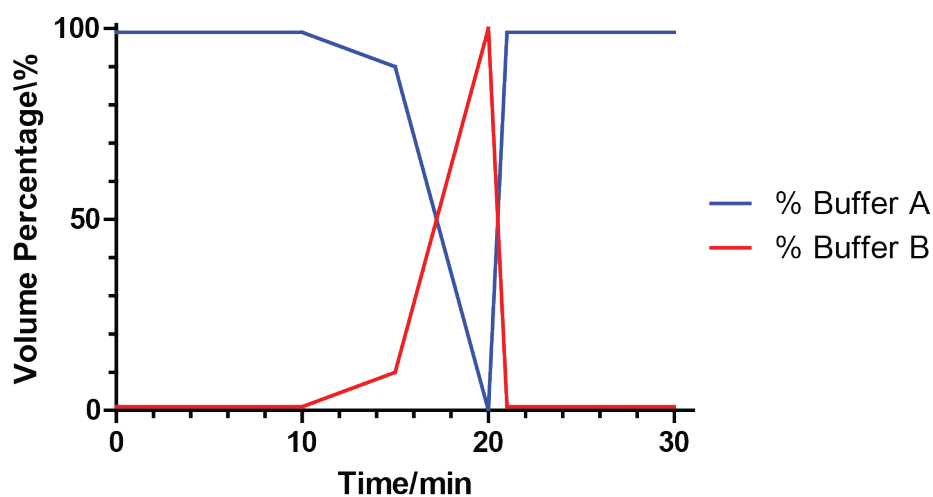


Figure 5.3: SiELC Primesep 100 HPLC method for LC/MS purification of ‘polar’ proline analogues.

Buffer A: 0.05% (v/v) FA in H₂O

Buffer B: 0.1% (v/v) in ACN.

Time/min	Flow/mL min ⁻¹	% Buffer A	% Buffer B
0.00	1.50	99.9	0.1
10.00	1.50	99.9	0.1
15.00	1.50	90.0	10.0
20.00	1.50	0.0	100.0
21.00	1.50	99.9	0.1
30.00	1.50	99.9	0.1

Table 5.32: SiELC Primesep 100 HPLC method for LC/MS purification of ‘very polar’ proline analogues.

Buffer A: 0.05% (v/v) FA in H₂O

Buffer B: 0.1% (v/v) in ACN.

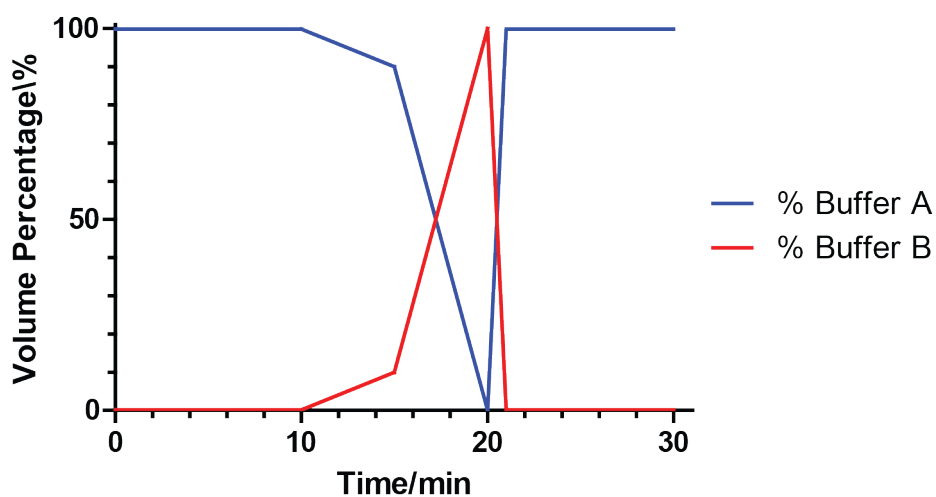


Figure 5.4: SiELC Primesep 100 HPLC method for LC/MS purification of ‘polar’ proline analogues.

Buffer A: 0.05% (v/v) FA in H₂O

Buffer B: 0.1% (v/v) in ACN.

5.2.2.4 NMR-Detection Assays

The following formulations were prepared in Eppendorf vials and then later transferred to NMR tubes for analysis.

Reaction Component	Stock Concentration	Final Concentration	Volume/ μ L
2-Oxoglutarate (2-OG)	100 mM	5.0/0.5 mM	5.0/0.5 μ L
L-Proline/Substrate	50 mM	5.0 mM	50.0 μ L
Sodium Ascorbate	100 mM	2.0/1.0 mM	10.0/5.0 μ L
(NH ₄) ₂ Fe(SO ₄) ₂	5 mM	0.5/0.25 mM	50.0/25.0 μ L
<i>cis</i> P3H Type I/ <i>cis</i> P4H/ <i>trans</i> P4H	51/71/25 mg/mL	15 μ M (0.5 mg/mL)	5.0/10.0/3.5 μ L
50 mM Tris- <i>d</i> ₁₁ -DCl, pD 7.5, D ₂ O	50.0 mM	-	(380/414.5) / (375/409.5) / (381.5/416) μ L
Total			500 μ L

Table 5.33: Proline hydroxylase NMR end-point assay conditions for starting material and product detection.

Reaction Component	Stock Concentration	Final Concentration	Volume/ μ L
2-Oxoglutarate (2-OG)	100 mM	1.0 mM	1.5 μ L
L-Proline/Substrate	50 mM	1.0 mM	3.0 μ L
Sodium Ascorbate	100 mM	1.0 mM	1.5 μ L
(NH ₄) ₂ Fe(SO ₄) ₂	5 mM	0.5/0.25 mM	10.0/5.0 μ L
<i>cis</i> P3H Type I/ <i>cis</i> P4H/ <i>trans</i> P4H	51/71/25 mg/mL	15/10 μ M	(1.5/1.0) / (1.4/0.7) / (3.0/2.0) μ L
50 mM Tris- <i>d</i> ₁₁ -DCl, pD 7.5, D ₂ O	50.0 mM	-	(132.5/138.0) / (132.6/138.3) / (131.0/137.0) μ L
Total			150 μ L

Table 5.34: Proline hydroxylase NMR timecourse assay conditions for starting material and product detection.

Reaction Component	Stock Concentration	Final Concentration	Volume/ μ L
2-Oxoglutarate (2-OG)	100 mM	5.0 mM	16.0 μ L
L-Proline/Substrate	50 mM	5.0 mM	8.0 μ L
Sodium Ascorbate	100 mM	1.0 mM	1.6 μ L
(NH ₄) ₂ Fe(SO ₄) ₂	5 mM	100 μ L	3.2 μ L
<i>cis</i> P3H Type I/ <i>cis</i> P4H/ <i>trans</i> P4H	51/71/25 mg/mL	5 μ M (0.5 mg/mL)	0.36 / 0.25 / 0.72 μ L
50 mM Tris- <i>d</i> ₁₁ -DCl, pD 7.5, D ₂ O	50.0 mM	-	130.8 / 131.0 / 130.5 μ L
Total			160 μ L

Table 5.35: Proline hydroxylase NMR timecourse assay conditions for starting material and product detection.

CHAPTER 5. MATERIALS AND METHODS

Reaction Component	Stock Concentration	Final Concentration	Volume/ μL
2-Oxoglutarate (2-OG)	100 mM	250 μM	0.8 μL
L-Proline/Substrate	50 mM	250 μM	0.4 μL
Sodium Ascorbate	100 mM	500 μM	0.8 μL
$(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$	5 mM	50 μL	1.6 μL
<i>cis</i> P3H Type I/ <i>cis</i> P4H/ <i>trans</i> P4H	5.1/7.1/2.5 mg/mL	1/0.5 μM	(1.07/0.53) / (0.75/0.37) / (2.13/1.07) μL
50 mM Tris- d_{11} -DCl, pD 7.5, D_2O	50.0 mM	-	(155.3/155.9) / (155.7/156.0) / (154.3/155.3) μL
Total			160 μL

Table 5.36: Proline hydroxylase NMR timecourse assay conditions for starting material and product detection.

Appendix A

Cloning, Sequences, and Purification

A.1 Purification Methods - IMPACT™-CN System

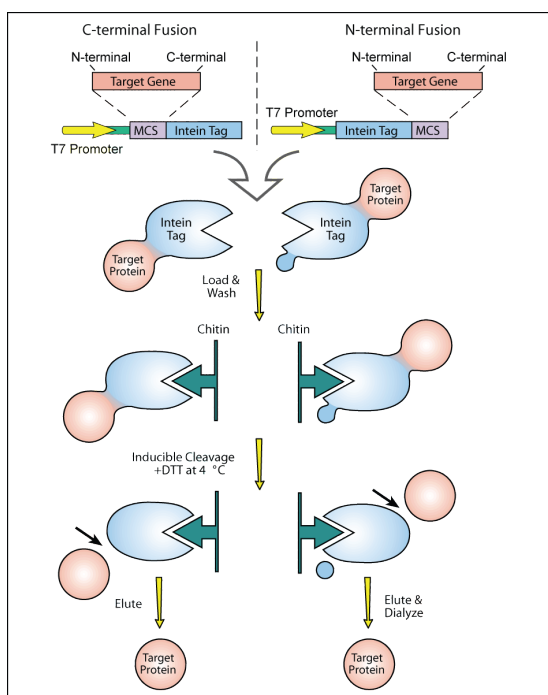


Figure A.1: Schematic illustration of the IMPACT™-CN System.²⁶⁶

A.2 Proline Hydroxylase Sequences

The following sequences are those which were used to produce the proline hydroxylase enzymes in this project.

A.2.1 *cis*P3H Type I Sequences

Sequences for full-length *Streptomyces* sp. L-Proline *cis*-3-Hydroxylase Type I construct as inserted in a Novagen pET-24-a(+) vector.

DNA Sequence

ATG CGC TCG CAC ATA CTC GGC CGC ATT GAA CTC GAC CAG GAA CGC TTA GGC AGG GAC CTC
GAA TAT CTC GCC ACG GTG CCC ACC GTG GAA GAG GAG TAC GAC GAG TTC AGC AAC GGG TTC
TGG AAG AAC ATC CCG CTG TAC AAC GCG AGC GGC GGC AGC GAG GAC CGG CTG TAC CGC
GAC CTC GAG GGC TCA CCG GCG CAG CCC ACC AAA CAC GCC GAG CAG GTT CCG TAC CTC AAC
GAG ATC ATC ACC ACG GTC TAC AAC GGC GAG CGG CTC CAG ATG GCG CGT ACG CGG AAC CTG
AAG AAC GCC GTC GTC ATC CCG CAC CGC GAC TTC GTG GAG CTC GAC CGC GAA CTC GAC CAG
TAC TTC CGC ACC CAT TTG ATG CTT GAG GAC AGC CCG CTG GCC TTC CAC TCG GAC GAC GAC
ACC GTC ATC CAC ATG CGG GCC GGC GAG ATC TGG TTC CTC GAC GCG GCC GCC GTC CAC TCG
GCC GTC AAC TTC GCC GAG TTC AGC AGG CAG TCG CTC TGC GTC GAC CTC GCC TTC GAC GGC
GCG TTC GAC GAG AAG GAA GCC TTC GCG GAC GCC ACG GTC TAC GCC CCG AAC CTC AGC CCC
GAC GTC CGC GAA CGC AAG CCG TTC ACC AAG GAG CGG GAG GCC GGG ATC CTC GCC CTG
TCC GGC GTG ATC GGA CGC GAG AAC TTC CGG GAC ATC CTC TTT CTG CTG TCC AAG GTC CAC
TAC ACC TAC GAC GTC CAT CCG GGT GAA ACC TTC GAG TGG CTC GTG AGC GTC TCC AAG GGT
GCG GGA GAC GAC AAG ATG GTG GAG AAG GCC GAG CGG ATC AGG GAC TTC GCC ATC GGC
GCA CGG GCA CTC GGC GAG CGT TTC TCG CTG ACC ACC TGG TAG

Amino Acid Sequence

10	20	30	40	50	60
MRSHILGRIE	LDQERLGRDL	EYLATVPTVE	EEYDEFSNGF	WKNIPLYNAS	GGSEDRLYRD
70	80	90	100	110	120
LEGSPAQPTK	HAEQVPYLN	IITTVYNGER	LQMARTRNK	NAVVIPHRDF	VELDRELDQY
130	140	150	160	170	180
FRTHLMLED	PLAFHSDDDT	VIHMRAGEIW	FLDAAAVHSA	VNFAEFSRQS	LCVDLAFDGA
190	200	210	220	230	240
FDEKEAFADA	TVYAPNLSPD	VRERKPFTKE	REAGILALSG	VIGRENFRDI	LFLLSKVHYT
250	260	270	280	290	
YDVHPGETFE	WLVSVSKGAG	DDKMVEKAER	IRDFAIGARA	LGERFSLTTW	

Theoretical Information

Number of Amino Acids: 290

Molecular Weight: 33,154.1 Da

Theoretical pI: 4.93

Total Number of Negatively Charged Residues (Asp + Glu): 51

Total Number of Positively Charged Residues (Arg + Lys): 33

Formula: C₁₄₈₁H₂₂₇₁N₄₀₇O₄₄₉S₆

Total number of atoms: 4614

Extinction coefficients are predicted for protein solutions in water using UV light ($\lambda = 280$ nm).

Ext. Coefficient = 36,900 M⁻¹ cm⁻¹ (Cys residue pairs all form cystines)

Abs 0.1% ($\equiv 1$ g l⁻¹) = 1.113

Ext. Coefficient = 36,900 M⁻¹ cm⁻¹ (Cys residues are all reduced)

Abs 0.1% ($\equiv 1$ g l⁻¹) = 1.113

Instability Index (II): 34.20 (*i.e.* Protein is stable)

Aliphatic Index: 83.10

Grand Average of Hydropathicity (GRAVY): -0.408

A.2.2 *cis*P4H Sequences

Sequence for full-length *Sinorhizobium meliloti* L-Proline *cis*-4-Hydroxylase construct. The DNA sequence was codon-optimised for the *E. coli* host.

DNA Sequence

ATG AGC ACC CAT TTT CTG GGC AAA GTG AAA TTT GAT GAA GCA CGT CTG GCA GAA GAT CTG
AGC ACC CTG GAA GTT GCA GAA TTC AGC AGC GCA TAT AGC GAT TTT GCA TGT GGT AAA TGG
GAA GCA TGT GTT CTG CGT AAT CGT ACC GGT ATG CAA GAA GAA GAT ATT GTT GTT AGC CAT
AAT GCA CCG GCA CTG GCA ACA CCG CTG AGC AAA AGC CTG CCG TAT CTG AAT GAA CTG GTT
GAA ACC CAT TTT GAT TGT AGC GCA GTT CGT TAT ACC CGT ATT GTT CGT GTT AGC GAA AAT
GCC TGT ATT ATT CCG CAT AGC GAT TAT CTG GAA CTG GAC GAA ACC TTT ACC CGT CTG CAT
CTG GTT CTG GAT ACC AAT AGC GGT TGT GCA AAT ACC GAA GAG GAT AAA ATC TTT CAC ATG
GGT CTG GGC GAA ATT TGG TTT CTG GAT GCA ATG CTG CCG CAT AGT GCA GCA TGT TTT AGC
AAA ACA CCG CGT CTG CAC CTG ATG ATT GAT TTT GAA GCA ACC GCA TTT CCG GAA TCA TTT
CTG CGC AAT GTT GAA CAG CCG GTT ACC ACC CGT GAT ATG GTT GAT CCG CGT AAA GAA CTG
ACC GAT GAA GTT ATT GAA GGT ATT CTG GGC TTT AGC ATC ATT ATT AGC GAA GCC AAT TAT
CGC GAA ATC GTT AGC ATT CTG GCA AAA CTG CAC TTC TTC TAT AAA GCA GAT TGC CGC AGC
ATG TAT GAT TGG CTG AAA GAA ATT TGT AAA CGT CGT GGT GAT CCG GCA CTG ATT GAA AAA
ACC GCA AGC CTG GAA CGT TTT TTT CTG GGT CAT CGT GAA CGT GGT GAA GTG ATG ACC TAT
TAA

Amino Acid Sequence

10	20	30	40	50	60
MSTHFLGKVK	FDEARLAEDL	STLEVAEFSS	AYSDFACGKW	EACVLRNRTG	MQEEDIVVSH
70	80	90	100	110	120
NAPALATPLS	KSLPYLNEIV	ETHFDCSAVR	YTRIVRVSEN	ACIIPHSDDL	ELDETFTRLH
130	140	150	160	170	180
LVLDTNSGCA	NTEEDKIFHM	GLGEIWFLDA	MLPHSAACFS	KTPRLHLMID	FEATAFPESF
190	200	210	220	230	240
LRNVEQPVT	RDMVDPRKEL	TDEVIEGILG	FSIIIEANY	REIVSILAKL	HFFYKADCRS
250	260	270	280		
MYDWLKEICK	RRGDPALIEK	TASLERFFLG	HRERGEVMTY		

Theoretical Properties

Number of Amino Acids: 272

Molecular Weight: 29,715.8 Da

Theoretical pI: 5.78

Total Number of Negatively Charged Residues (Asp + Glu): 34

Total Number of Positively Charged Residues (Arg + Lys): 27

Formula: C₁₃₁₉H₂₀₇₈N₃₈₄O₃₈₄S₈

Total number of atoms: 4173

Extinction coefficients are predicted for protein solutions in water using UV light ($\lambda = 280$ nm).

Ext. Coefficient = 32,555 M⁻¹ cm⁻¹ (Cys residue pairs all form cystines)

Abs 0.1% ($\equiv 1$ g l⁻¹) = 1.096

Ext. Coefficient = 32,430 M⁻¹ cm⁻¹ (Cys residues are all reduced)

Abs 0.1% ($\equiv 1$ g l⁻¹) = 1.091

Instability Index (II): 41.69 (*i.e.* Protein is unstable)

Aliphatic Index: 93.68

Grand Average of Hydropathicity (GRAVY): -0.196

A.2.3 *trans*P4H Sequences

Sequences for full-length *Dactylosporangium* sp. (strain RH1) L-Proline *trans*-4-Hydroxylase.

DNA Sequence

ATG CTG ACC CCG ACC GAA CTG AAA CAG TAT CGT GAA GCA GGT TAT CTG CTG ATT GAA GAT
 GGT CTG GGT CCG CGT GAA GTT GAT TGT CTG CGT CGT GCA GCA GCA GCA CTG TAT GCA CAG
 GAT AGT CCG GAT CGT ACC CTG GAA AAA GAT GGT CGT ACC GTT CGT GCA GTT CAT GGT TGT
 CAT CGT CGT GAT CCG GTT TGC CGT GAT CTG GTT CGT CAT CCG CGT CTG CTG GGT CCG GCA
 ATG CAG ATT CTG AGC GGT GAT GTT TAT GTG CAC CAG TTT AAA ATC AAT GCC AAA GCA CCG
 ATG ACC GGT GAT GTG TGG CCG TGG CAT CAG GAT TAT ATC TTT TGG GCA CGC GAA GAT GGT
 ATG GAT CGT CCG CAT GTT GTT AAT GTT GCA GTT CTG CTG GAT GAA GCA ACC CAT CTG AAT
 GGT CCG CTG CTG TTT GTT CCG GGT ACA CAT GAA CTG GGT CTG ATT GAT GTT GAA CGT CGT
 GCC CCT GCC GGT GAT GGT GAT GCA CAG TGG CTG CCG CAG CTG AGC GCA GAT CTG GAT TAT
 GCA ATT GAT GCC GAT CTG CTG GCA CGT CTG ACC GCA GGT CGT GGT ATT GAA AGC GCA ACC
 GGT CCG GCA GGT AGC ATT CTG CTG TTC GAT AGC CGT ATT GTT CAT GGT AGC GGC ACC AAT
 ATG AGT CCG CAT CCG CGT GGT GTT GTT CTG GTT ACC TAT AAT CGT ACC GAT AAT GCA CTG
 CCT GCA CAG GCA GCA CCG CGT CCG GAA TTT CTG GCA GCA CGT GAT GCA ACA CCG CTG GTT
 CCG CTG CCT GCA GGT TTT GCA CTG GCA CAG CCG GTT **TAA**

Amino Acid Sequence

10	20	30	40	50	60
MLTPTELKQY	REAGYLLIED	GLGPREDVCL	RRAAAALYAQ	DSPDRTLEKD	GRTVRAVHGC
70	80	90	100	110	120
HRRDPVCRDL	VRHPRLLGPA	MQILSGDVYV	HQFKINAKAP	MTGDVWPWHQ	DYIFWAREDG
130	140	150	160	170	180
MDRPHVVNVA	VLLDEATHLN	GPLLFPVPGTH	ELGLIDVERR	APAGDGDAQW	LPQLSADLDY
190	200	210	220	230	240
AIDADLLARL	TAGRGIESAT	GPAGSILLFD	SRIVHGSGTN	MSPHPRGVVL	VTYNRTDNAL
250	260	270			
PAQAAPRPEF	LAARDATPLV	PLPAGFALAQ	PV		

Theoretical Properties

Number of Amino Acids: 272

Molecular Weight: 29,715.8 Da

Theoretical pI: 5.78

Total Number of Negatively Charged Residues (Asp + Glu): 34

Total Number of Positively Charged Residues (Arg + Lys): 27

Formula: C₁₃₁₉H₂₀₇₈N₃₈₄O₃₈₄S₈

Total number of atoms: 4173

Extinction coefficients are predicted for protein solutions in water using UV light ($\lambda = 280 \text{ nm}$).

Ext. Coefficient = $32,555 \text{ M}^{-1} \text{ cm}^{-1}$ (Cys residue pairs all form cystines)

Abs 0.1% ($\equiv 1 \text{ g l}^{-1}$) = 1.096

Ext. Coefficient = $32,430 \text{ M}^{-1} \text{ cm}^{-1}$ (Cys residues are all reduced)

Abs 0.1% ($\equiv 1 \text{ g l}^{-1}$) = 1.091

Instability Index (II): 41.69 (*i.e.* Protein is unstable)

Aliphatic Index: 93.68

Grand Average of Hydropathicity (GRAVY): -0.196

A.2.3.1 Takara pCOLD I N-His₆-Tag Sequence (5'-*Nde*I and 3'-*Bam*HI Cloning Sites)

DNA Sequence

ATGAAT CAC AAA GTG CAT CAT CAT CAT CAT ATC GAA GGT AGG CAT ...

Amino Acid Sequence

MNHKV HHHHH HIEGR H...

Molecular Weight: 2,048.28 Da

A.2.3.1.1 N-His₆-cisP4H Theoretical Properties

Number of Amino Acids: 296

Molecular Weight: 34,046.7 Da

Theoretical pI: 5.68

Total Number of Negatively Charged Residues (Asp + Glu): 46

Total Number of Positively Charged Residues (Arg + Lys): 32

Formula: C₁₅₁₅H₂₃₄₀N₄₁₆O₄₄₅S₁₇

Total number of atoms: 4733

Extinction coefficients are predicted for protein solutions in water using UV light ($\lambda = 280 \text{ nm}$).

Ext. Coefficient = $28,920 \text{ M}^{-1} \text{ cm}^{-1}$ (Cys residue pairs all form cystines)

Abs 0.1% ($\equiv 1 \text{ g l}^{-1}$) = 0.849

Ext. Coefficient = $28,420 \text{ M}^{-1} \text{ cm}^{-1}$ (Cys residues are all reduced)

Abs 0.1% ($\equiv 1 \text{ g l}^{-1}$) = 0.835

Instability Index (II): 43.18 (*i.e.* Protein is unstable)

Aliphatic Index: 93.68

Grand Average of Hydropathicity (GRAVY): -0.246

A.2.3.1.2 N-His₆-*trans*P4H Theoretical Properties

Number of Amino Acids: 288

Molecular Weight: 31,741.0 Da

Theoretical pI: 6.29

Total Number of Negatively Charged Residues (Asp + Glu): 35

Total Number of Positively Charged Residues (Arg + Lys): 29

Formula: C₁₄₀₆H₂₂₀₃N₄₂₁O₄₀₃S₉

Total number of atoms: 4442

Extinction coefficients are predicted for protein solutions in water using UV light ($\lambda = 280$ nm).

Ext. Coefficient = 32,555 M⁻¹ cm⁻¹ (Cys residue pairs all form cystines)

Abs 0.1% ($\equiv 1$ g l⁻¹) = 1.026

Ext. Coefficient = 32,430 M⁻¹ cm⁻¹ (Cys residues are all reduced)

Abs 0.1% ($\equiv 1$ g l⁻¹) = 1.022

Instability Index (II): 44.17 (*i.e.* Protein is unstable)

Aliphatic Index: 90.83

Grand Average of Hydropathicity (GRAVY): -0.292

A.2.4 EFE Sequences

Sequences for full-length *Pseudomonas syringae* pv. *phaseolicola* Ethylene-Forming Enzyme (EFE).

DNA Sequence

ATG ACC AAT CTG CAA ACC TTT GAA CTG CCG ACC GAA GTT ACC GGT TGT GCA GCA GAT ATT
AGC CTG GGT CGT GCA CTG ATT CAG GCA TGG CAG AAA GAT GGT ATT TTT CAG ATT AAA ACC
GAT AGC GAG CAG GAT CGT AAA ACC CAA GAA GCA ATG GCA GCA AGC AAA CAG TTT TGT AAA
GAA CCG CTG ACC TTT AAA AGC AGC TGT GTT AGC GAT CTG ACC TAT AGC GGT TAT GTT GCA
AGC GGT GAA GAA GTG ACC GCA GGT AAA CCG GAT TTT CCT GAA ATC TTT ACC GTT TGC AAA
GAT CTG AGC GTT GGT GAT CAG CGT GTT AAA GCA GGT TGG CCG TGT CAT GGT CCG GTT CCG
TGG CCG AAT AAT ACC TAT CAG AAA AGC ATG AAA ACC TTT ATG GAA GAA CTG GGT CTG GCA
GGC GAA CGT CTG CTG AAA CTG ACC GCA CTG GGT TTT GAG CTG CCG ATT AAT ACA TTT ACC
GAT CTG ACA CGT GAT GGC TGG CAT CAT ATG CGT GTT CTG CGT TTT CCG CCT CAG ACC AGC
ACC CTG AGC CGT GGT ATT GGT GCA CAT ACC GAT TAT GGT CTG CTG GTT ATT GCA GCA CAG
GAT GAT GTT GGT GGT CTG TAT ATT CGT CCG CCT GTT GAA GGT GAA AAA CGT AAT CGT AAT
TGG CTG CCT GGT GAA AGC AGC GCA GGT ATG TTT GAA CAT GAT GAA CCG TGG ACC TTT GTT
ACC CCG ACA CCG GGT GTT TGG ACC GTT TTT CCG GGT GAT ATT CTG CAA TTT ATG ACC GGT
GGT CAG CTG CTG AGC ACA CCG CAT AAA GTT AAA CTG AAT ACC CGT GAA CGT TTT GCC TGT

APPENDIX A. CLONING, SEQUENCES, AND PURIFICATION

GCC TAT TTT CAT GAA CCG AAT TTT GAA GCA AGC GCC TAT CCG CTG TTT GAA CCG AGC GCA
AAT GAA CGT ATT CAT TAT GGT GAA CAC TTT ACC AAC ATG TTC ATG CGT TGT TAT CCG GAT
CGT ATT ACC ACC CAG CGT ATT AAC AAA GAA AAT CGT CTG GCA CAT CTG GAA GAT CTG AAA
AAA TAC AGC GAT ACC CGT GCA ACC GGT AGC TAA

Amino Acid Sequence

10	20	30	40	50	60
MTNLQTFELP	TEVTGCAADI	SLGRALIQAW	QKDGIFQIKT	DSEQDRKTQE	AMAASKQFCK
70	80	90	100	110	120
EPLTFKSSCV	SDLTYSGYVA	SGEEVTAGKP	DFPEIFTVCK	DLSVGDQRVK	AGWPCHGPVP
130	140	150	160	170	180
WPNNTYQKSM	KTFMEELGLA	GERLLKLTAL	GFELPINTFT	DLTRDGWHHM	RVLRFPPQTS
190	200	210	220	230	240
TLSRGIGAHT	DYGLLVIAAQ	DDVGGLYIRP	PVEGEKRNRN	WLPGESSAGM	FEHDEPWTFV
250	260	270	280	290	300
TPTPGVWTVF	PGDILQFMTG	GQLLSTPHKV	KLNTRERFAC	AYFHEPNFEA	SAYPLFEPSA
310	320	330	340	350	
NERIHYGEHF	TNMFMRCPD	RITTQRINKE	NRLAHLEDLK	KYSDTRATGS	

Theoretical Properties

Number of Amino Acids: 350

Molecular Weight: 39,444.6 Da

Theoretical pI: 5.96

Total Number of Negatively Charged Residues (Asp + Glu): 43

Total Number of Positively Charged Residues (Arg + Lys): 37

Formula: C₁₇₆₁H₂₆₉₆N₄₇₈O₅₂₃S₁₆

Total number of atoms: 5474

Extinction coefficients are predicted for protein solutions in water using UV light ($\lambda = 280$ nm).

Ext. Coefficient = 53,775 M⁻¹ cm⁻¹ (Cys residue pairs all form cystines)

Abs 0.1% ($\equiv 1$ g l⁻¹) = 1.363

Ext. Coefficient = 53,400 M⁻¹ cm⁻¹ (Cys residues are all reduced)

Abs 0.1% ($\equiv 1$ g l⁻¹) = 1.354

Instability Index (II): 30.09 (*i.e.* Protein is stable)

Aliphatic Index: 66.63

Grand Average of Hydropathicity (GRAVY): -0.472

A.2.4.1 Takara pCOLD I N-His₆-Tag Sequence (5'-*SacI* and 3'-*PstI* Cloning Sites)

DNA Sequence

ATG AAT CAC AAA GTG CAT CAT CAT CAT CAT ATC GAA GGT AGG CAT ATG GAG CTC ...

Amino Acid Sequence

MNHNKV HHHHH HIEGH MEL...

Molecular Weight: 2,260.71 Da

A.2.4.1.1 N-His₆-EFE Theoretical Properties

Number of Amino Acids: 368

Molecular Weight: 41,687.1 Da

Theoretical pI: 6.22

Total Number of Negatively Charged Residues (Asp + Glu): 45

Total Number of Positively Charged Residues (Arg + Lys): 32

Formula: C₁₈₅₈H₂₈₃₆N₅₁₄O₅₄₆S₁₈

Total number of atoms: 4733

Extinction coefficients are predicted for protein solutions in water using UV light ($\lambda = 280$ nm).

Ext. Coefficient = 53,775 M⁻¹ cm⁻¹ (Cys residue pairs all form cystines)

Abs 0.1% ($\equiv 1$ g l⁻¹) = 1.290

Ext. Coefficient = 53,400 M⁻¹ cm⁻¹ (Cys residues are all reduced)

Abs 0.1% ($\equiv 1$ g l⁻¹) = 1.281

Instability Index (II): 31.90 (i.e. Protein is unstable)

Aliphatic Index: 66.28

Grand Average of Hydropathicity (GRAVY): -0.514

References

- [1] Pauling, L. *The Nature of the Chemical Bond* (Cornell University Press, Ithaca, NY, 1960), 3rd Edn.
- [2] Jia, C., Kitamura, T. & Fujiwara, Y. Catalytic Functionalization of Arenes and Alkanes via C—H Bond Activation. *Acc. Chem. Res.* **34**, 633–639 (2001).
- [3] Dach, R., Song, J. H. J., Roschangar, F., Samstag, W. & Senanayake, C. H. The Eight Criteria Defining a Good Chemical Manufacturing Process. *Org. Process Res. Dev.* **16**, 1697–1706 (2012).
- [4] Castro, G. R. & Knubovets, T. Homogeneous Biocatalysis in Organic Solvents and Water-Organic Mixtures. *Crit. Rev. Biotechnol.* **23**, 195–231 (2003).
- [5] Fasan, R. Tuning P450 Enzymes as Oxidation Catalysts. *ACS Catalysis* **2**, 647–666 (2012).
- [6] Julsing, M. K., Cornelissen, S., Buhler, B. & Schmid, A. Heme-Iron Oxygenases: Powerful Industrial Biocatalysts? *Curr. Opin. Chem. Biol.* **12**, 177–186 (2008).
- [7] Zhang, K. D., Shafer, B. M., Demars, M. D., Stern, H. A. & Fasan, R. Controlled Oxidation of Remote sp(3) C—H Bonds in Artemisinin via P450 Catalysts with Fine-Tuned Regio- and Stereoselectivity. *J. Am. Chem. Soc.* **134**, 18695–18704 (2012).
- [8] Meunier, B., de Visser, S. P. & Shaik, S. Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes. *Chem. Rev.* **104**, 3947–3980 (2004).
- [9] Guengerich, F. P. Mechanisms of Cytochrome P450 Substrate Oxidation: MiniReview. *J. Biochem. Mol. Toxicol.* **21**, 163–168 (2007).
- [10] Guengerich, F. P., Sohl, C. D. & Chowdhury, G. Multi-Step Oxidations Catalyzed by Cytochrome P450 Enzymes: Processive vs. Distributive Kinetics and the Issue of Carbonyl Oxidation in Chemical Mechanisms. *Arch. Biochem. Biophys.* **507**, 126–134 (2011).

REFERENCES

- [11] Huijbers, M. M., Montersino, S., Westphal, A. H., Tischler, D. & van Berkel, W. J. Flavin Dependent Monooxygenases. *Arch. Biochem. Biophys.* **544**, 2–17 (2014).
- [12] Hausinger, R. P. FeII/ α -Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68 (2004).
- [13] Clifton, I. J. *et al.* Structural Studies on 2-Oxoglutarate Oxygenases and Related Double-Stranded β -Helix Fold Proteins. *J. Inorg. Biochem.* **100**, 644–669 (2006).
- [14] Flashman, E. & Schofield, C. J. The Most Versatile of All Reactive Intermediates? *Nat. Chem. Biol.* **3**, 86–87 (2007).
- [15] Loenarz, C. & Schofield, C. J. Expanding Chemical Biology of 2-Oxoglutarate Oxygenases. *Nat. Chem. Biol.* **4**, 152–156 (2008).
- [16] Farrow, S. C. & Facchini, P. J. Dioxygenases Catalyze O-Demethylation and O,O-Demethylenation with Widespread Roles in Benzyloquinoline Alkaloid Metabolism in Opium Poppy. *J. Biol. Chem.* **288**, 28997–29012 (2013).
- [17] Tarhonskaya, H. *et al.* Studies on Deacetoxycephalosporin C Synthase Support a Consensus Mechanism for 2-Oxoglutarate Dependent Oxygenases. *Biochemistry* **53**, 2483–2493 (2014).
- [18] De Carolis, E. & De Luca, V. 2-Oxoglutarate-Dependent Dioxygenase and Related Enzymes: Biochemical Characterization. *Phytochemistry* **36**, 1093–1107 (1994).
- [19] Martens, S., Preuss, A. & Matern, U. Multifunctional Flavonoid Dioxygenases: Flavonol and Anthocyanin Biosynthesis in *Arabidopsis thaliana* L. *Phytochemistry* **71**, 1040–1049 (2010).
- [20] Lando, D. *et al.* FIH-1 is an Asparaginyl Hydroxylase Enzyme that Regulates the Transcriptional Activity of Hypoxia-Inducible Factor. *Genes Dev.* **16**, 1466–1471 (2002).
- [21] Hewitson, K. S. *et al.* Hypoxia-Inducible Factor (HIF) Asparagine Hydroxylase is Identical to Factor Inhibiting HIF (FIH) and is Related to the Cupin Structural Family. *J. Biol. Chem.* **277**, 26351–26355 (2002).
- [22] Mori, H., Shibasaki, T., Uozaki, Y., Ochiai, K. & Ozaki, A. Detection of Novel Proline 3-Hydroxylase Activities in *Streptomyces* and *Bacillus* spp. By Regio- and Stereospecific Hydroxylation of L-Proline. *Appl. Environ. Microbiol.* **62**, 1903–1907 (1996).
- [23] Mori, H., Shibasaki, T., Yano, K. & Ozaki, A. Purification and Cloning of a Proline 3-Hydroxylase, A Novel Enzyme Which Hydroxylates Free L-Proline to *cis*-3-Hydroxy-L-Proline. *J. Bacteriol.* **179**, 5677–5683 (1997).

-
- [24] Watanabe, M. *et al.* A Phosphonate-Induced Gene which Promotes Penicillium-Mediated Bioconversion of *cis*-Propenylphosphonic Acid to Fosfomycin. *Appl. Environ. Microbiol.* **65**, 1036–1044 (1999).
- [25] Hashimoto, T. & Yamada, Y. Hyoscyamine 6 β -Hydroxylase, a 2-Oxoglutarate-Dependent Dioxygenase, in Alkaloid-Producing Root Cultures. *Plant Physiol.* **81**, 619–625 (1986).
- [26] Vaillancourt, F. H., Yin, J. & Walsh, C. T. Syrb2 in Syringomycin E Biosynthesis Is a Nonheme Fe(II)-, α -Ketoglutarate-, and O₂-Dependent Halogenase. *Proc. Natl. Acad. Sci. U S A* **102**, 10111–10116 (2005).
- [27] Saito, K., Kobayashi, M., Gong, Z., Tanaka, Y. & Yamazaki, M. Direct Evidence for Anthocyanidin Synthase as a 2-Oxoglutarate-Dependent Oxygenase: Molecular Cloning and Functional Expression of cDNA from a Red Forma of *Perilla frutescens*. *Plant J.* **17**, 181–189 (1999).
- [28] Wilmouth, R. C. *et al.* Structure and Mechanism of Anthocyanidin Synthase from *Arabidopsis thaliana*. *Structure* **10**, 93–103 (2002).
- [29] McGowan, S. J. *et al.* Analysis of Bacterial Carbapenem Antibiotic Production Genes Reveals a Novel β -Lactam Biosynthesis Pathway. *Mol. Microbiol.* **22**, 415–426 (1996).
- [30] Baldwin, J. E. *et al.* Purification and Initial Characterization of an Enzyme with Deacetoxycephalosporin C Synthetase and Hydroxylase Activities. *Biochem. J.* **245**, 831–841 (1987).
- [31] Dotzlaf, J. E. & Yeh, W. K. Copurification and Characterization of Deacetoxycephalosporin C Synthetase/Hydroxylase from *Cephalosporium acremonium*. *J. Bacteriol.* **169**, 1611–1618 (1987).
- [32] Samson, S. M. *et al.* Isolation, Sequence Determination, and Expression in *Escherichia coli* of the Isopenicillin N Synthetase Gene from *Cephalosporium acremonium*. *Nature* **318**, 191–194 (1985).
- [33] Busby, R. W., Chang, M. D. T., Busby, R. C., Wimp, J. & Townsend, C. A. Expression and Purification of 2 Isozymes of Clavamate Synthase and Initial Characterization of the Iron-Binding Site - General Error Analysis in Polymerase Chain-Reaction Amplification. *J. Biol. Chem.* **270**, 4262–4269 (1995).
- [34] Janc, J. W., Egan, L. A. & Townsend, C. A. Purification and Characterization of Clavamate Synthase from *Streptomyces antibioticus*. A Multifunctional Enzyme of Clavam Biosynthesis. *J. Biol. Chem.* **270**, 5399–5404 (1995).
- [35] Eichhorn, E., Van Der Ploeg, J. R., Kertesz, M. A. & Leisinger, T. Characterization of α -Ketoglutarate-Dependent Taurine Dioxygenase from *Escherichia coli*. *J. Biol. Chem.* **272**, 23031–23036 (1997).
-

REFERENCES

- [36] Whetstine, J. R. *et al.* Reversal of Histone Lysine Trimethylation by the JMJD2 Family of Histone Demethylases. *Cell* **125**, 467–481 (2006).
- [37] Klose, R. J. *et al.* The Transcriptional Repressor JHDM3A Demethylates Trimethyl Histone H3 Lysine 9 and Lysine 36. *Nature* **442**, 312–316 (2006).
- [38] Price, J. C., Barr, E. W., Tirupati, B., Bollinger, J. M. & Krebs, C. The First Direct Characterization of a High-Valent Iron Intermediate in the Reaction of an α -Ketoglutarate-Dependent Dioxygenase: A High-Spin Fe(IV) Complex in Taurine/ α -Ketoglutarate Dioxygenase (TauD) from *Escherichia coli*. *Biochemistry* **42**, 7497–7508 (2003).
- [39] Chelikani, P., Fita, I. & Loewen, P. C. Diversity of Structures and Properties Among Catalases. *Cell. Mol. Life Sci.* **61**, 192–208 (2004).
- [40] Prescott, A. G. & John, P. Dioxygenases: Molecular Structure and Role in Plant Metabolism. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 245–271 (1996).
- [41] Hufton, S. E., Jennings, I. G. & Cotton, R. G. H. Structure and Function of the Aromatic Amino-Acid Hydroxylases. *Biochem. J.* **311**, 353–366 (1995).
- [42] Fitzpatrick, P. F. The Aromatic Amino Acid Hydroxylases. *Adv. Enzymol. Relat. Areas Mol. Biol.* **74**, 235–294 (2000).
- [43] Zurfluh, M. R. *et al.* Molecular Genetics of Tetrahydrobiopterin-Responsive Phenylalanine Hydroxylase Deficiency. *Hum. Mutat.* **29**, 1079–1079 (2008).
- [44] Finocchiaro, L. M. E. *et al.* Serotonin and Melatonin Synthesis in Peripheral-Blood Mononuclear-Cells - Stimulation by Interferon-Gamma as Part of an Immunomodulatory Pathway. *J. Interferon Res.* **8**, 705–716 (1988).
- [45] Mohammad-Zadeh, L. F., Moses, L. & Gwaltney-Brant, S. M. Serotonin: A Review. *J. Vet. Pharmacol. Ther.* **31**, 187–199 (2008).
- [46] Fitzpatrick, P. F. Tetrahydropterin-Dependent Amino Acid Hydroxylases. *Annu. Rev. Biochem.* **68**, 355–381 (1999).
- [47] Lakey, J. H., Lea, E. J., Rudd, B. A., Wright, H. M. & Hopwood, D. A. A New Channel-Forming Antibiotic from *Streptomyces coelicolor* A3(2) Which Requires Calcium for Its Activity. *J. Gen. Microbiol.* **129**, 3565–3573 (1983).
- [48] Kempter, C. *et al.* CDA: Calcium-Dependent Peptide Antibiotics from *Streptomyces coelicolor* A3(2) Containing Unusual Residues. *Angew. Chem. Int. Ed. Engl.* **36**, 498–501 (1997).
- [49] Hojati, Z. *et al.* Structure, Biosynthetic Origin, and Engineered Biosynthesis of Calcium-Dependent Antibiotics from *Streptomyces coelicolor*. *Chem. Biol.* **9**, 1175–1187 (2002).

-
- [50] Neary, J. M. *et al.* An Asparagine Oxygenase (AsnO) and a 3-Hydroxyasparaginyl Phosphotransferase (HasP) Are Involved in the Biosynthesis of Calcium-Dependent Lipopeptide Antibiotics. *Microbiology* **153**, 768–776 (2007).
- [51] Strieker, M., Kopp, F., Mahlert, C., Essen, L. O. & Marahiel, M. A. Mechanistic and Structural Basis of Stereospecific C- β -Hydroxylation in Calcium-Dependent Antibiotic, A Daptomycin-Type Lipopeptide. *ACS Chem. Biol.* **2**, 187–196 (2007).
- [52] Katz, E., Kamal, F. & Mason, K. Biosynthesis of *trans*-4-Hydroxy-L-Proline by *Streptomyces griseoviridus*. *J. Biol. Chem.* **254**, 6684–6690 (1979).
- [53] Onishi, M., Okumura, Y., Okamoto, R. & Ishikura, T. Proline Hydroxylation by Cell Free Extract of a Streptomycete. *Biochemical and Biophysical Research Communications* **120**, 45–51 (1984).
- [54] Yin, X., O'Hare, T., Gould, S. J. & Zabriskie, T. M. Identification and Cloning of Genes Encoding Viomycin Biosynthesis from *Streptomyces vinaceus* and Evidence for Involvement of a Rare Oxygenase. *Gene* **312**, 215–224 (2003).
- [55] Yin, X. H. & Zabriskie, T. M. VioC is a Non-Heme Iron, α -Ketoglutarate-Dependent Oxygenase That Catalyzes the Formation of 3S-Hydroxy-L-Arginine During Viomycin Biosynthesis. *ChemBioChem* **5**, 1274–1277 (2004).
- [56] Segre, A. *et al.* The Structure of Syringomycins A1, E, and G. *FEBS Lett.* **255**, 27–31 (1989).
- [57] Scaloni, A. *et al.* Structure, Conformation and Biological Activity of a Novel Lipodepsipeptide from *Pseudomonas corrugata*: Cormycin A. *Biochem. J.* **384**, 25–36 (2004).
- [58] Ballio, A. *et al.* Novel Bioactive Lipodepsipeptides from *Pseudomonas syringae*: the Pseudomycins. *FEBS Lett.* **355**, 96–100 (1994).
- [59] Fukuchi, N., Isogai, A., Nakayama, J. & Suzuki, A. Structure of Syringotoxin B, a Phytotoxin Produced by Citrus Isolates of *Pseudomonas syringae* *pv. syringae*. *Agric. Biol. Chem. (Tokyo)* **54**, 3377–3379 (1990).
- [60] Sorensen, K. N., Kim, K. H. & Takemoto, J. Y. *In vitro* Antifungal and Fungicidal Activities and Erythrocyte Toxicities of Cyclic Lipodepsinonapeptides Produced by *Pseudomonas syringae* *pv. syringae*. *Antimicrob. Agents Chemother.* **40**, 2710–2713 (1996).
- [61] Feil, H. *et al.* Comparison of the Complete Genome Sequences of *Pseudomonas syringae* *pv. syringae* B728a and *pv. tomato* DC3000. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 11064–11069 (2005).
-

REFERENCES

- [62] Singh, G. M., Fortin, P. D., Koglin, A. & Walsh, C. T. β -Hydroxylation of the Aspartyl Residue in the Phytotoxin Syringomycin E: Characterization of Two Candidate Hydroxylases AspH and SyrP in *Pseudomonas syringae*. *Biochemistry* **47**, 11310–11320 (2008).
- [63] Fowden, L., Pratt, H. M. & Smith, A. 4-Hydroxyisoleucine from Seed of *Trigonella foenum-graecum*. *Phytochemistry* **12**, 1707–1711 (1973).
- [64] Alcock, N. W. *et al.* Stereochemistry of the 4-Hydroxyisoleucine from *Trigonella foenum-graecum*. *Phytochemistry* **28**, 1835–1841 (1989).
- [65] Al-Habori, M. & Raman, A. Antidiabetic and Hypcholesterolaemic Effects of Fenugreek. *Phytother. Res.* **12**, 233–242 (1998).
- [66] Sauvaire, Y. *et al.* 4-Hydroxyisoleucine: A Novel Amino Acid Potentiator of Insulin Secretion. *Diabetes* **47**, 206–210 (1998).
- [67] Broca, C. *et al.* 4-Hydroxyisoleucine: Experimental Evidence of its Insulinotropic and Antidiabetic Properties. *Am. J. Physiol.* **277**, E617–E623 (1999).
- [68] Broca, C. *et al.* Insulinotropic Agent ID-1101 (4-Hydroxyisoleucine) Activates Insulin Signaling in Rat. *Am. J. Physiol. Endocrinol. Metab.* **287**, E463–E471 (2004).
- [69] Jackson, J. E. & Bressler, R. Clinical Pharmacology of Sulphonylurea Hypoglycaemic Agents: Part 1. *Drugs* **22**, 211–245 (1981).
- [70] Jennings, A. M., Wilson, R. M. & Ward, J. D. Symptomatic Hypoglycemia in NIDDM Patients Treated with Oral Hypoglycemic Agents. *Diabetes Care* **12**, 203–208 (1989).
- [71] Wang, Q., Quazzani, J., Sasaki, N. A. & Potier, P. A Practical Synthesis of (2S,3R,4S)-4-Hydroxyisoleucine, a Potent Insulinotropic α -Amino Acid from Fenugreek. *Eur. J. Org. Chem.* **2002**, 834–839 (2002).
- [72] Smirnov, S. V. *et al.* A Novel Strategy for Enzymatic Synthesis of 4-Hydroxyisoleucine: Identification of an Enzyme Possessing Hmcp (4-Hydroxy-3-Methyl-2-Keto-Pentanoate) Aldolase Activity. *FEMS Microbiol. Lett.* **273**, 70–77 (2007).
- [73] Ogawa, J. *et al.* Synthesis of 4-Hydroxyisoleucine by the Aldolase-Transaminase Coupling Reaction and Basic Characterization of the Aldolase from *Arthrobacter simplex* AKU 626. *Biosci., Biotechnol., Biochem.* **71**, 1607–1615 (2007).
- [74] Haefele, C., Bonfils, C. & Sauvaire, Y. Characterization of a Dioxygenase from *Trigonella foenum-graecum* Involved in 4-Hydroxyisoleucine Biosynthesis. *Phytochemistry* **44**, 563–566 (1997).

-
- [75] Kodera, T. *et al.* A Novel L-Isoleucine Hydroxylating Enzyme, L-Isoleucine Dioxygenase from *Bacillus thuringiensis*, Produces (2*S*,3*R*,4*S*)-4-Hydroxyisoleucine. *Biochem. Biophys. Res. Commun.* **390**, 506–510 (2009).
- [76] Hibi, M. *et al.* Characterization of *Bacillus thuringiensis* L-Isoleucine Dioxygenase for Production of Useful Amino Acids. *Appl. Environ. Microbiol.* **77**, 6926–6930 (2011).
- [77] Smirnov, S. V. *et al.* Metabolic Engineering of *Escherichia coli* to Produce (2*S*, 3*R*, 4*S*)-4-Hydroxyisoleucine. *Appl. Microbiol. Biotechnol.* **88**, 719–726 (2010).
- [78] Weber, M. M., French, C. L., Barnes, M. B., Siegele, D. A. & McLean, R. J. A Previously Uncharacterized Gene, *yjfO* (*bsmA*), Influences *Escherichia coli* Biofilm Formation and Stress Response. *Microbiology* **156**, 139–147 (2010).
- [79] Smirnov, S. V. *et al.* A Novel Family of Bacterial Dioxygenases That Catalyse the Hydroxylation of Free L-Amino Acids. *FEMS Microbiol. Lett.* **331**, 97–104 (2012).
- [80] Smirnov, S. V. *et al.* A Novel L-Isoleucine-4'-Dioxygenase and L-Isoleucine Dihydroxylation Cascade in *Pantoea ananatis*. *MicrobiologyOpen* **2**, 471–481 (2013).
- [81] Golakoti, T., Yoshida, W. Y., Chaganty, S. & Moore, R. E. Isolation and Structures of Nostopeptolides A1, A2, and A3 from the Cyanobacterium *Nostoc sp.* GSV224. *Tetrahedron* **56**, 9093–9102 (2000).
- [82] Hoffmann, D., Hevel, J. M., Moore, R. E. & Moore, B. S. Sequence Analysis and Biochemical Characterization of the Nostopeptolide A Biosynthetic Gene Cluster from *Nostoc sp.* GSV224. *Gene* **311**, 171–180 (2003).
- [83] Becker, J. E., Moore, R. E. & Moore, B. S. Cloning, Sequencing, and Biochemical Characterization of the Nostocyclopeptide Biosynthetic Gene Cluster: Molecular Basis for Imine Macrocyclization. *Gene* **325**, 35–42 (2004).
- [84] Ota, Y. & Gomi, S. L-Leucine Hydroxylase and DNA Encoding the Enzyme (2008). JP Patent: WO/2008/013262 (Filed 2007); Kyowa Hakko Kogyo Co., Ltd., Japan.
- [85] Hibi, M. *et al.* L-Leucine 5-Hydroxylase of *Nostoc punctiforme* is a Novel Type of Fe(II)/ α -Ketoglutarate-Dependent Dioxygenase that is Useful as a Biocatalyst. *Appl. Microbiol. Biotechnol.* **97**, 2467–2472 (2013).
- [86] Yadav, J., Chandrasekhar, S., Ravindra Reddy, Y. & Rama Rao, A. V. Synthesis of (2*S*, 3*R*)-3-Hydroxy Leucine: A Constituent of Lysobactin. *Tetrahedron* **51**, 2749–2754 (1995).
-

REFERENCES

- [87] Makino, K., Okamoto, N., Hara, O. & Hamada, Y. Efficient Enantioselective Synthesis of (2R,3R)- and (2S,3S)-3-Hydroxyleucines and their Diastereomers Through Dynamic Kinetic Resolution. *Tetrahedron: Asymmetry* **12**, 1757–1762 (2001).
- [88] Kimura, T., Vassilev, V. P., Shen, G. J. & Wong, C. H. Enzymatic Synthesis of β -Hydroxy- α -Amino Acids Based on Recombinant D- and L-Threonine Aldolases. *J. Am. Chem. Soc.* **119**, 11734–11742 (1997).
- [89] Hibi, M. *et al.* A Novel Fe(II)/ α -Ketoglutarate-Dependent Dioxygenase from *Burkholderia ambifaria* Has β -Hydroxylating Activity of N-Succinyl-L-Leucine. *Lett. Appl. Microbiol.* **55**, 414–419 (2012).
- [90] Finlay, A. C. *et al.* Viomycin a New Antibiotic Active Against Mycobacteria. *Am. Rev. Tuberc.* **63**, 1–3 (1951).
- [91] Herr, E. B., Haney, M. E., Pittenger, G. E. & Higgins, C. E. Isolation and Characterization of a New Peptide Antibiotic. *Proc. Indiana Acad. Sci.* **69**, 134 (1960).
- [92] Nagata, A., Ando, T., Izumi, R., Sakakibara, H. & Take, T. Studies on Tuberactinomycin (Tuberactin), A New Antibiotic. I. Taxonomy of Producing Strain, Isolation and Characterization. *J. Antibiot. (Tokyo)* **21**, 681–687 (1968).
- [93] Liou, Y. F. & Tanaka, N. Dual Actions of Viomycin on the Ribosomal Functions. *Biochem. Biophys. Res. Commun.* **71**, 477–483 (1976).
- [94] Marrero, P., Cabanas, M. J. & Modolell, J. Induction of Translational Errors (Misreading) by Tuberactinomycins and Capreomycins. *Biochem. Biophys. Res. Commun.* **97**, 1047–1052 (1980).
- [95] Modolell, J. & Vázquez, D. The Inhibition of Ribosomal Translocation by Viomycin. *Eur. J. Biochem.* **81**, 491–497 (1977).
- [96] Ju, J. H., Ozanick, S. G., Shen, B. & Thomas, M. G. Conversion of (2S)-Arginine to (2S,3R)-Capreomycinidine by VioC and VioD from the Viomycin Biosynthetic Pathway of *Streptomyces* sp. Strain ATCC11861. *ChemBioChem* **5**, 1281–1285 (2004).
- [97] Magarvey, N. A., Haltli, B., He, M., Greenstein, M. & Hucul, J. A. Biosynthetic Pathway for Mannopectimycins, Lipoglycopeptide Antibiotics Active Against Drug-Resistant Gram-Positive Pathogens. *Antimicrob. Agents Chemother.* **50**, 2167–2177 (2006).
- [98] Haltli, B. *et al.* Investigating β -Hydroxyenduracididine Formation in the Biosynthesis of the Mannopectimycins. *Chem. Biol.* **12**, 1163–1168 (2005).

-
- [99] Singh, M. P. *et al.* Mannopeptimycins, New Cyclic Glycopeptide Antibiotics Produced by *Streptomyces hygroscopicus* LL-AC98: Antibacterial and Mechanistic Activities. *Antimicrob. Agents Chemother.* **47**, 62–69 (2003).
- [100] Seyden-Penne, J. *Chiral Auxiliaries and Ligand in Asymmetric Synthesis* (Wiley, New York, 1995).
- [101] Davies, J. S. *Amino Acids and Peptides* (Chapman and Hall, London, 1985).
- [102] Barrett, G. C. *Chemistry and Biochemistry of the Amino Acids* (Chapman and Hall, London, 1985).
- [103] Jones, J. H. Amino Acids and Peptides. In *Specialist Periodical Report*, Vol. 23. The Royal Society of Chemistry, London (The Royal Society of Chemistry, London, 1992).
- [104] Hoekstra, W. J. The Chemistry and Biology of β -Amino Acids. *Curr. Med. Chem.* **6**, 905–1004 (1999).
- [105] Bewley, C. A. & Faulkner, D. J. Lithistid Sponges: Star Performers or Hosts to the Stars. *Angew. Chem. Int. Ed. Engl.* **37**, 2,162–2,178 (1998).
- [106] Berks, A. H. Preparations of Two Pivotal Intermediates for the Synthesis of 1- β -Methyl Carbapenem Antibiotics. *Tetrahedron* **52**, 331–375 (1996).
- [107] Trost, B. M. & Brindle, C. S. The Direct Catalytic Asymmetric Aldol Reaction. *Chemical Society Reviews* **39**, 1600–1632 (2010).
- [108] Zlotin, S. G., Kucherenko, A. S. & Beletskaya, I. P. Organocatalysis of Asymmetric Aldol Reaction. Catalysts and Reagents. *Russ. Chem. Rev.* **78**, 737–784 (2009).
- [109] Kuttan, R. & Radhakrishnan, A. N. Biochemistry of the Hydroxyprolines. *Adv. Enzymol. Relat. Areas Mol. Biol.* **37**, 273–347 (1973).
- [110] Wichmann, C. F., Liesch, J. M. & Schwartz, R. E. L-671,329, a New Antifungal Agent. II. Structure Determination. *J. Antibiot. (Tokyo)* **42**, 168–173 (1989).
- [111] Brockmann, H., Pampus, G. & Manegold, J. H. Actinomycine. 21. Antibiotica Aus Actinomyceten. 43. Actinomycin-X-O- β -Zur Systematik Und Nomenklatur. *Chemische Berichte-Recueil* **92**, 1294–1302 (1959).
- [112] Morris, M. I. & Villmann, M. Echinocandins in the Management of Invasive Fungal Infections, Part 1. *Am. J. Health Syst. Pharm.* **63**, 1693–1703 (2006).
- [113] Morris, M. I. & Villmann, M. Echinocandins in the Management of Invasive Fungal Infections, Part 2. *Am. J. Health Syst. Pharm.* **63**, 1813–1820 (2006).
-

REFERENCES

- [114] Hollstein, U. Actinomycin. Chemistry and Mechanism of Action. *Chem. Rev.* **74**, 625–652 (1974).
- [115] Ivan, M. *et al.* HIF α Targeted for VHL-Mediated Destruction by Proline Hydroxylation: Implications for O₂ Sensing. *Science* **292**, 464–468 (2001).
- [116] Jaakkola, P. *et al.* Targeting of HIF-1 α to the von Hippel-Lindau Ubiquitylation Complex by O₂-Regulated Prolyl Hydroxylation. *Science* **292**, 468–472 (2001).
- [117] Masson, N., Willam, C., Maxwell, P. H., Pugh, C. W. & Ratcliffe, P. J. Independent Function of Two Destruction Domains in Hypoxia-Inducible Factor- α Chains Activated by Prolyl Hydroxylation. *EMBO J.* **20**, 5197–5206 (2001).
- [118] Yu, F., White, S. B., Zhao, Q. & Lee, F. S. HIF-1 α Binding to VHL Is Regulated by Stimulus-Sensitive Proline Hydroxylation. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9630–9635 (2001).
- [119] Smith, T. G., Robbins, P. A. & Ratcliffe, P. J. The Human Side of Hypoxia-Inducible Factor. *Br. J. Haematol.* **141**, 325–334 (2008).
- [120] Gleadle, J. M. & Ratcliffe, P. J. Hypoxia and the Regulation of Gene Expression. *Mol. Med. Today* **4**, 122–129 (1998).
- [121] Semenza, G. L. HIF-1 and Mechanisms of Hypoxia Sensing. *Curr. Opin. Cell. Biol.* **13**, 167–171 (2001).
- [122] Semenza, G. L. Hypoxia-Inducible Factor 1: Oxygen Homeostasis and Disease Pathophysiology. *Trends Mol. Med.* **7**, 345–350 (2001).
- [123] Benizri, E., Ginouves, A. & Berra, E. The Magic of the Hypoxia-Signaling Cascade. *Cell. Mol. Life Sci.* **65**, 1133–1149 (2008).
- [124] Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. Hypoxia-Inducible Factor 1 is a Basic-Helix-Loop-Helix-PAS Heterodimer Regulated by Cellular O₂ Tension. *Proc. Natl. Acad. Sci. U S A* **92**, 5510–5514 (1995).
- [125] Cockman, M. E. *et al.* Hypoxia Inducible Factor- α Binding and Ubiquitylation by the von Hippel-Lindau Tumor Suppressor Protein. *J. Biol. Chem.* **275**, 25733–25741 (2000).
- [126] Maxwell, P. H. *et al.* The Tumour Suppressor Protein VHL Targets Hypoxia-Inducible Factors for Oxygen-Dependent Proteolysis. *Nature* **399**, 271–275 (1999).
- [127] Salceda, S. & Caro, J. Hypoxia-Inducible Factor 1 α (HIF-1 α) Protein Is Rapidly Degraded by the Ubiquitin-Proteasome System under Normoxic Conditions. Its Stabilization by Hypoxia Depends on Redox-Induced Changes. *J. Biol. Chem.* **272**, 22642–22647 (1997).

-
- [128] Huang, L. E., Gu, J., Schau, M. & Bunn, H. F. Regulation of Hypoxia-Inducible Factor 1 α Is Mediated by an O₂-Dependent Degradation Domain via the Ubiquitin-Proteasome Pathway. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 7987–7992 (1998).
- [129] Kallio, P. J., Wilson, W. J., O'Brien, S., Makino, Y. & Poellinger, L. Regulation of the Hypoxia-Inducible Transcription Factor 1 α by the Ubiquitin-Proteasome Pathway. *J. Biol. Chem.* **274**, 6519–6525 (1999).
- [130] Sutter, C. H., Laughner, E. & Semenza, G. L. Hypoxia-Inducible Factor 1 α Protein Expression Is Controlled by Oxygen-Regulated Ubiquitination That Is Disrupted by Deletions and Missense Mutations. *Proc. Natl. Acad. Sci. U S A* **97**, 4748–4753 (2000).
- [131] Epstein, A. C. *et al.* *C. elegans* EGL-9 and Mammalian Homologs Define a Family of Dioxygenases that Regulate HIF by Prolyl Hydroxylation. *Cell* **107**, 43–54 (2001).
- [132] Bruick, R. K. & McKnight, S. L. A Conserved Family of Prolyl-4-Hydroxylases That Modify HIF. *Science* **294**, 1337–1340 (2001).
- [133] Pugh, C. W., O'Rourke, J. F., Nagao, M., Gleadle, J. M. & Ratcliffe, P. J. Activation of Hypoxia-Inducible Factor-1; Definition of Regulatory Domains within the α -Subunit. *J. Biol. Chem.* **272**, 11205–11214 (1997).
- [134] Flashman, E. *et al.* Kinetic Rationale for Selectivity toward N- and C-Terminal Oxygen-Dependent Degradation Domain Substrates Mediated by a Loop Region of Hypoxia-Inducible Factor Prolyl Hydroxylases. *J. Biol. Chem.* **283**, 3808–3815 (2008).
- [135] Min, J. H. *et al.* Structure of an HIF-1 α -pVHL Complex: Hydroxyproline Recognition in Signaling. *Science* **296**, 1886–1889 (2002).
- [136] Hon, W. C. *et al.* Structural Basis for the Recognition of Hydroxyproline in HIF-1 α by pVHL. *Nature* **417**, 975–978 (2002).
- [137] Kadler, K. E., Baldock, C., Bella, J. & Boot-Handford, R. P. Collagens at a Glance. *J. Cell Sci.* **120**, 1955–1958 (2007).
- [138] Myllyharju, J. & Kivirikko, K. I. Collagens and Collagen-Related Diseases. *Ann. Med.* **33**, 7–21 (2001).
- [139] Kivirikko, K. I., Myllyla, R. & Pihlajaniemi, I. *Post-Translational Modifications of Proteins* (CRC Press, Boca Raton, FL, 1992).
- [140] Kivirikko, K. I. & Pihlajaniemi, T. Collagen Hydroxylases and the Protein Disulfide Isomerase Subunit of Prolyl 4-Hydroxylases. *Advances in Enzymology* **72**, 325–398 (1998).
-

REFERENCES

- [141] Smith, T. G. & Talbot, N. P. Prolyl Hydroxylases and Therapeutics. *Antioxid. Redox Signaling* **12**, 431–433 (2010).
- [142] Berg, R. A. & Prockop, D. J. The Thermal Transition of a Non-Hydroxylated Form of Collagen - Evidence for a Role for Hydroxyproline in Stabilizing Triple-Helix of Collagen. *Biochem. Biophys. Res. Commun.* **52**, 115–120 (1973).
- [143] Privalov, P. L. Stability of Proteins - Proteins Which Do Not Present a Single Cooperative System. *Adv. Protein Chem.* **35**, 1–104 (1982).
- [144] Beck, K. & Brodsky, B. Supercoiled Protein Motifs: The Collagen Triple-Helix and the α -Helical Coiled Coil. *J. Struct. Biol.* **122**, 17–29 (1998).
- [145] Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. Crystal and Molecular Structure of a Collagen-Like Peptide at 1.9 Å Resolution. *Science* **266**, 75–81 (1994).
- [146] Ramachandran, G. N. & Kartha, G. Structure of Collagen. *Nature* **174**, 269–270 (1954).
- [147] Ramachandran, G. N. & Kartha, G. Structure of Collagen. *Nature* **176**, 593–595 (1955).
- [148] Ramakrishnan, C. *In Memoriam: Professor G.N. Ramachandran (1922–2001)*. *Protein Sci.* **10**, 1689–1691 (2001).
- [149] Ramachandran, G. N., Bansal, M. & Bhatnagar, R. S. A Hypothesis on the Role of Hydroxyproline in Stabilizing Collagen Structure. *Biochim. Biophys. Acta* **322**, 166–171 (1973).
- [150] Inouye, K., Sakakibara, S. & Prockop, D. J. Effects of the Stereo-Configuration of the Hydroxyl Group in 4-Hydroxyproline on the Triple-Helical Structures Formed by Homogenous Peptides Resembling Collagen. *Biochim. Biophys. Acta* **420**, 133–141 (1976).
- [151] Vitagliano, L., Berisio, R., Mazzarella, L. & Zagari, A. Structural Bases of Collagen Stabilization Induced by Proline Hydroxylation. *Biopolymers* **58**, 459–464 (2001).
- [152] Berisio, R. *et al.* Effects of Microgravity on the Crystal Quality of a Collagen-Like Polypeptide. *Acta Crystallogr. D Biol. Crystallogr.* **56**, 55–61 (2000).
- [153] Shoulders, M. D. & Raines, R. T. Collagen Structure and Stability. *Annu. Rev. Biochem.* **78**, 929–958 (2009).
- [154] Deerinck, T. Collagen Fibrils, SEM (C022/2892). URL <http://www.sciencephoto.com/media/620521/view>.

-
- [155] Hutton, J. J., A., K. M. & Udenfriend, S. Conversion of the Amino Acid Sequence Gly-Pro-Pro in Protein to Gly-Pro-Hyp by Collagen Proline Hydroxylase. *Arch. Biochem. Biophys.* **121**, 384–391 (1967).
- [156] Kivirikko, K. I. & Prockop, D. J. Partial Characterization of Protocollagen from Embryonic Cartilage. *Biochem. J.* **102**, 432–442 (1967).
- [157] Gorres, K. L. & Raines, R. T. Prolyl 4-Hydroxylase. *Crit. Rev. Biochem. Mol. Biol.* **45**, 106–124 (2010).
- [158] Prockop, D. J., Berg, R. A., Kivirikko, K. I. & Uitto, J. *Biochemistry of Collagen*, Chap. Intracellular Steps in the Biosynthesis of Collagen, 163–273 (Springer US, 1976).
- [159] Atreya, P. L. & Ananthanarayanan, V. S. Interaction of Prolyl 4-Hydroxylase with Synthetic Peptide Substrates: A Conformational Model for Collagen Proline Hydroxylation. *J. Biol. Chem.* **266**, 2852–2858 (1991).
- [160] Cardinale, G. J. & Udenfriend, S. Prolyl Hydroxylase. *Adv. Enzymol. Relat. Areas Mol. Biol.* **41**, 245–300 (1974).
- [161] Wijffels, G. L. *et al.* The Secreted Cathepsin L-Like Proteinases of the Trematode, *Fasciola hepatica*, Contain 3-Hydroxyproline Residues. *Biochem J.* **299** (Pt 3), 781–790 (1994).
- [162] Bozas, S. E. & Spithill, T. W. Identification of 3-Hydroxyproline Residues in Several Proteins of *Fasciola hepatica*. *Exp. Parasitol.* **82**, 69–72 (1996).
- [163] Risteli, J., Tryggvason, K. & Kivirikko, K. I. Prolyl 3-Hydroxylase: Partial Characterization of the Enzyme from Rat Kidney Cortex. *Eur. J. Biochem./FEBS* **73**, 485–492 (1977).
- [164] Tryggvason, K., Risteli, J. & Kivirikko, K. I. Separation of Prolyl 3-Hydroxylase and 4-Hydroxylase Activities and the 4-Hydroxyproline Requirement for Synthesis of 3-Hydroxyproline. *Biochem. Biophys. Res. Commun.* **76**, 275–281 (1977).
- [165] Singleton, R. S. *et al.* OGFOD1 Catalyzes Prolyl Hydroxylation of RPS23 and is Involved in Translation Control and Stress Granule Formation. *Proc. Natl. Acad. Sci. U S A* **111**, 4031–4036 (2014).
- [166] Hughes, B. T. & Espenshade, P. J. Oxygen-Regulated Degradation of Fission Yeast SREBP by Ofd1, a Prolyl Hydroxylase Family Member. *EMBO J.* **27**, 1491–1501 (2008).
- [167] Keeling, K. M., Salas-Marco, J., Osherovich, L. Z. & Bedwell, D. M. Tpa1p is Part of an mRNP Complex that Influences Translation Termination, mRNA Deadenylation, and mRNA Turnover in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **26**, 5237–5248 (2006).
-

REFERENCES

- [168] Henri, J. *et al.* Structural and Functional Insights into *Saccharomyces cerevisiae* Tpa1, a Putative Prolylhydroxylase Influencing Translation Termination and Transcription. *J. Biol. Chem.* **285**, 30767–30778 (2010).
- [169] Kim, H. S. *et al.* Crystal Structure of Tpa1 from *Saccharomyces cerevisiae*, a Component of the Messenger Ribonucleoprotein Complex. *Nucleic Acids Res.* **38**, 2099–2110 (2010).
- [170] Wehner, K. A., Schütz, S. & Sarnow, P. OGFOD1, a Novel Modulator of Eukaryotic Translation Initiation Factor 2 α Phosphorylation and the Cellular Response to Stress. *Molecular and Cellular Biology* **30**, 2006–2016 (2010).
- [171] Saito, K., Adachi, N., Koyama, H. & Matsushita, M. OGFOD1, a Member of the 2-Oxoglutarate and Iron Dependent Dioxygenase Family, Functions in Ischemic Signaling. *FEBS Lett.* **584**, 3340–3347 (2010).
- [172] Heinemann, B. *et al.* Etamycin, a New Antibiotic. *Antibiot. Ann.* **1954-1955**, 728–732 (1955).
- [173] Bartz, Q. R. *et al.* Griseoviridin and Viridogrisein: Isolation and Characterisation. *Antibiot. Ann.* **1954-1955**, 777–784 (1954).
- [174] Arnold, R. B., Johnson, A. W. & Mauger, A. B. The Structure of Viridogrisein (Etamycin). *J. Chem. Soc.* **0**, 4466–4470 (1958).
- [175] Sheehan, J. C., Zachau, H. G. & Lawson, W. B. The Structure of Etamycin. *J. Am. Chem. Soc.* **80**, 3349–3355 (1958).
- [176] Dickison, H. L., Cull, K. M. & Tisch, D. E. Pharmacologic Studies with Etamycin. *Antibiot. Ann., Medical Encyclopedia Inc.* **1954-1955**, 733–744 (1955).
- [177] Ehrlich, J. *et al.* Griseoviridin and Viridogrisein: Biological Studies. *Antibiot. Ann., Medical Encyclopedia Inc.* **1954-1955**, 790 (1955).
- [178] Sheehan, J. C. & Ledis, S. L. Total synthesis of a monocyclic peptide lactone antibiotic, etamycin. *J. Am. Chem. Soc.* **95**, 875–879 (1973).
- [179] Hook, D. J. & Vining, L. C. Biosynthetic Precursors of Etamycin, a Peptidolactone Antibiotic from *Streptomyces griseoviridus*. *Can. J. Biochem.* **51**, 1630–1637 (1973).
- [180] Bodanszky, M. & Perlman, D. Origin of D-Amino-Acids in Microbial Peptides: Rule of α -Epimerization. *Nature* **218**, 291–292 (1968).
- [181] Bycroft, B. W. Structural Relationships in Microbial Peptides. *Nature* **224**, 595–597 (1969).

-
- [182] Okumura, Y., Ohnishi, M. & Okamoto, R. Some Effectors on the Biosynthesis of *cis*-4-Hydroxy-D-Proline in Viridogrisein. *Agric. Biol. Chem.* **46**, 3063–3068 (1982).
- [183] Haste, N. M. *et al.* Activity of the Streptogramin Antibiotic Etamycin against Methicillin-Resistant *Staphylococcus aureus*. *J. Antibiot. (Tokyo)* **63**, 219–224 (2010).
- [184] Como-Sabetti, K. *et al.* Community-Associated Methicillin-Resistant *Staphylococcus aureus*: Trends in Case and Isolate Characteristics from Six Years of Prospective Surveillance. *Public Health Rep.* **124**, 427–435 (2009).
- [185] Patel, M. Community-Associated Meticillin-Resistant *Staphylococcus aureus* Infections: Epidemiology, Recognition and Management. *Drugs* **69**, 693–716 (2009).
- [186] Misiek, M. *et al.* Telomycin, a New Antibiotic. *Antibiot. Ann.* **1957-1958**, 852–855 (1957).
- [187] Gourevitch, A. *et al.* Microbiological Studies on Telomycin. *Antibiot. Ann.* **5**, 856–862 (1957).
- [188] Sheehan, J. C., Mania, D., Nakamura, S., Stock, J. A. & Maeda, K. Structure of Telomycin. *J. Am. Chem. Soc.* **90**, 462–470 (1968).
- [189] Kumar, N. G. & Urry, D. W. Proton Magnetic Resonance Assignments of the Polypeptide Antibiotic Telomycin. *Biochemistry* **12**, 3811–3817 (1973). URL <http://pubs.acs.org/doi/abs/10.1021/bi00744a003>.
- [190] Kumar, N. G., Izumiya, N., Miyoshi, M., Sugano, H. & Urry, D. W. Conformational and Spectral Analysis of the Polypeptide Antibiotic *N*-Methyllleucine Gramicidin S Dihydrochloride by Nuclear Magnetic Resonance. *Biochemistry* **14**, 2197–2207 (1975).
- [191] Lawrence, C. C. *Studies on Bacterial Proline 4-Hydroxylase*. DPhil thesis, Department of Chemistry, University of Oxford (1993).
- [192] Baldwin, J. E., Field, R. A., Lawrence, C. C., Merritt, K. D. & Schofield, C. J. Proline 4-Hydroxylase: Stereochemical Course of the Reaction. *Tetrahedron Lett.* **34**, 7489–7492 (1993).
- [193] Lawrence, C. C., Sobey, W. J., Field, R. A., Baldwin, J. E. & Schofield, C. J. Purification and Initial Characterization of Proline 4-Hydroxylase from *Streptomyces griseoviridis* P8648: a 2-Oxoacid, Ferrous-Dependent Dioxygenase Involved in Etamycin Biosynthesis. *Biochem. J.* **313**, 185–191 (1996).
- [194] Shibasaki, T., Mori, H., Chiba, S. & Ozaki, A. Microbial Proline 4-Hydroxylase Screening and Gene Cloning. *Appl. Environ. Microbiol.* **65**, 4028–4031 (1999).
-

REFERENCES

- [195] Shibasaki, T. *et al.* Substrate Selectivities of Proline Hydroxylases. *Tetrahedron Lett.* **40**, 5227–5230 (1999).
- [196] Shibasaki, T., Hashimoto, S., Mori, H. & Ozaki, A. Construction of a Novel Hydroxyproline-Producing Recombinant *Escherichia coli* by Introducing a Proline 4-Hydroxylase Gene. *J. Biosci. Bioeng.* **90**, 522–525 (2000).
- [197] Shibasaki, T., Mori, H. & Ozaki, A. Enzymatic Production of *trans*-4-Hydroxy-L-Proline by Regio- and Stereospecific Hydroxylation of L-Proline. *Biosci. Biotechnol. Biochem.* **64**, 746–750 (2000).
- [198] Klein, C. & Huttel, W. A Simple Procedure for Selective Hydroxylation of L-Proline and L-Pipecolic Acid with Recombinantly Expressed Proline Hydroxylases. *Adv. Synth. Catal.* **353**, 1375–1383 (2011).
- [199] Michaelis, L. & Menten, M. L. Die Kinetik Der Invertinwirkung (Kinetics of Invertase Action). *Biochem. Z.* **49**, 333–369 (1913).
- [200] Michaelis, L., Menten, M. L., Johnson, K. A. & Goody, R. S. The Original Michaelis Constant: Translation of the 1913 Michaelis-Menten Paper. *Biochemistry* **50**, 8264–8269 (2011).
- [201] Shibasaki, T., Mori, H. & Ozaki, A. Cloning of an Isozyme of Proline 3-Hydroxylase and Its Purification from Recombinant *Escherichia coli*. *Biotechnol. Lett.* **22**, 1967–1973 (2000).
- [202] Simossis, V. A. & Heringa, J. PRALINE: a Multiple Sequence Alignment Toolbox that Integrates Homology-Extended and Secondary Structure Information. *Nucleic Acids Res.* **33**, W289–W294 (2005).
- [203] Simossis, V. A. & Heringa, J. PRALINE (Multiple Sequence Alignment Software) (2014). URL <http://www.ibi.vu.nl/programs/pralinewww/>. Centre for Integrative Bioinformatics VU, vrije Universiteit amsterdam (IBIVU), Amsterdam, Netherlands.
- [204] Johnston, R. M. *et al.* Hydroxylation of L-Proline to *cis*-3-Hydroxy-L-Proline by Recombinant *Escherichia coli* Expressing a Synthetic L-Proline-3-Hydroxylase Gene. *Enzyme Microb. Technol.* **45**, 484–490 (2009).
- [205] Clifton, I. J., Hsueh, L. C., Baldwin, J. E., Harlos, K. & Schofield, C. J. Structure of Proline 3-Hydroxylase. Evolution of the Family of 2-Oxoglutarate Dependent Oxygenases. *Eur. J. Biochem.* **268**, 6625–6636 (2001).
- [206] Bontoux, M. C. & Gelo-Pujic, M. Microbial Screening in Hydroxylation of L-Proline. *Tetrahedron Lett.* **47**, 9073–9076 (2006).
- [207] Hara, R. & Kino, K. Characterization of Novel 2-Oxoglutarate Dependent Dioxygenases Converting L-Proline to *cis*-4-Hydroxy-L-Proline. *Biochem. Biophys. Res. Commun.* **379**, 882–886 (2009).

- [208] Jarvis, B. D. W., Pankhurst, C. E. & Patel, J. J. *Rhizobium loti*, a New Species of Legume Root Nodule Bacteria. *Int. J. Syst. Bacteriol.* **32**, 378–380 (1982).
- [209] Pohlman, G. G. Nitrogen Fixation by *Rhizobium meliloti* and *Rhizobium japonicum*. *J. Am. Soc. Agron.* **23**, 22–27 (1931).
- [210] Metzner, L., Kalbitz, J. & Brandsch, M. Transport of Pharmacologically Active Proline Derivatives by the Human Proton-Coupled Amino Acid Transporter hPAT1. *J. Pharmacol. Exp. Ther.* **309**, 28–35 (2004).
- [211] Petersen, L., Olewinski, R., Salmon, P. & Connors, N. Novel Proline Hydroxylase Activities in the Pneumocandin-Producing Fungus *Glarea lozoyensis* Responsible for the Formation of *trans*-3- and *trans*-4-Hydroxyproline. *Appl. Microbiol. Biotechnol.* **62**, 263–267 (2003).
- [212] Houwaart, S., Youssar, L. & Hüttel, W. Pneumocandin Biosynthesis: Involvement of a *trans*-Selective Proline Hydroxylase. *ChemBioChem* **15**, 2365–2369 (2014).
- [213] SRI Consulting Inc. CEH Marketing Research Report: Ethylene. Tech. Rep., SRI International, Englewood, Colorado, USA (2011).
- [214] Ipatieff, V. N. & Corson, B. B. Gasoline from ethylene by catalytic polymerization. *Ind. Eng. Chem.* **28**, 860–863 (1936).
- [215] Kusmiyati, K. & Saidina Amin, N. A. Production of Gasoline Range Hydrocarbons from Catalytic Reaction of Methane in the Presence of Ethylene over W/HZSM-5. *Catal. Today* **106**, 271–274 (2005).
- [216] Worrell, E., Phylipsen, D., Einstein, D. & Martin, N. Energy Use and Energy Intensity of the US Chemical Industry. Tech. Rep., Ernest Orlando Lawrence Berkeley National Laboratory, University of California, Berkeley, California, USA 94720 (2000).
- [217] Johnson, P. R. & Ecker, J. R. The Ethylene Gas Signal Transduction Pathway: A Molecular Perspective. *Annu. Rev. Genet.* **32**, 227–254 (1998).
- [218] Wang, K. L., Li, H. & Ecker, J. R. Ethylene Biosynthesis and Signaling Networks. *The Plant Cell* **14**, 131–152 (2002).
- [219] Chen, Y. F., Etheridge, N. & Schaller, G. E. Ethylene Signal Transduction. *Ann. Botany* **95**, 901–915 (2005).
- [220] Adams, D. O. & Yang, S. F. Ethylene Biosynthesis: Identification of 1-Aminocyclopropane-1-Carboxylic Acid as an Intermediate in the Conversion of Methionine to Ethylene. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 170–174 (1979).

REFERENCES

- [221] Van der Straeten, D., Van Wiemeersch, L. ., Goodman, H. M. & Van Montagu, M. Cloning and Sequence of Two Different cDNAs Encoding 1-Aminocyclopropane-1-Carboxylate Synthase in Tomato. *Proc. Natl. Acad. Sci. USA* **87**, 4859–4863 (1990).
- [222] Nakajima, N., Mori, H., Yamazaki, M. K. & Imaseki, H. Molecular Cloning and Sequence of a Complementary DNA Encoding 1-Aminocyclopropane-1-Carboxylic Synthase Induced by Tissue Wounding. *Plant Cell Physiol.* **31**, 1016–1021 (1990).
- [223] Yip, W.-K., Dong, J.-G., Kenny, J. W., Thompson, G. A. & Yang, S. F. Characterization and Sequencing of the Active Site of 1-Aminocyclopropane-1-Carboxylate Synthase. *Proc. Natl. Acad. Sci. USA* **87**, 7930–7934 (1990).
- [224] Sato, T., Oeller, P. W. & Theologis, A. The 1-Aminocyclopropane-1-Carboxylate Synthase of *Cucurbita*. *J. Biol. Chem.* **266**, 3752–3759 (1991).
- [225] Dong, J. G. *et al.* Cloning of a cDNA Encoding 1-Aminocyclopropane-1-Carboxylate Synthase and Expression of its mRNA in Ripening Apple Fruit. *Planta* **185**, 38–45 (1991).
- [226] Yang, S. F. & Hoffman, N. E. Ethylene Biosynthesis and its Regulation in Higher Plants. *Annual Review of Plant Physiology* **35**, 155–189 (1984).
- [227] Dong, J. G., Fernández-Maculeit, J. C. & Yang, S. F. Purification and Characterization of 1-Aminocyclopropane-1-Carboxylate Oxidase from Apple Fruit. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 9789–9793 (1992).
- [228] Fernández-Maculeit, J. C., Dong, J. G. & Yang, S. F. Activation of 1-Aminocyclopropane-1-Carboxylate Oxidase by Carbon Dioxide. *Biochem. Biophys. Res. Commun.* **193**, 1168–1173 (1993).
- [229] Poneleit, L. S. & Dilley, D. R. Carbon Dioxide Activation of 1-Aminocyclopropane-1-Carboxylate (ACC) Oxidase in Ethylene Biosynthesis. *Postharvest Biol. Technol.* **3**, 191–199 (1993).
- [230] Smith, J. J. & John, P. Activation of 1-Aminocyclopropane-1-Carboxylate Oxidase by Bicarbonate/Carbon Dioxide. *Phytochemistry* **32**, 1381–1386 (1993).
- [231] Adlington, R. M., Aplin, R. T., Baldwin, J. E., Rawlings, B. J. & Osborne, D. On the Biosynthesis of Ethylene. *J. Chem. Soc., Chem. Commun.* **18**, 1086–1087 (1982).
- [232] Baldwin, J. E. *et al.* On the Biosynthesis of Ethylene: Further Evidence for Stepwise Enzymatic Cyclopropane Ring Cleavage. *J. Chem. Soc., Chem. Commun.* **12**, 775–777 (1988).

-
- [233] Blumenthal, S. G., Hendrickson, H. R., Abrol, Y. P. & Conn, E. E. Cyanide Metabolism in Higher Plants III. The Biosynthesis of β -Cyanoalanine. *J. Biol. Chem.* **243**, 5302–5307 (1968).
- [234] Weingart, H., Ullrich, H., Geider, K. & Völksch, B. The Role of Ethylene Production in Virulence of *Pseudomonas syringae* pvs. *glycinea* and *phaseolicola*. *Phytopathology* **91**, 511–518 (2001).
- [235] Mansouri, S. & Bunch, A. W. Bacterial Ethylene Synthesis from 2-Oxo-4-Thiobutyric Acid and from Methionine. *Journal of General Microbiology* **135**, 2819–2827 (1989).
- [236] Shipston, N. & Bunch, A. W. The Physiology of L-Methionine Catabolism to the Secondary Metabolite Ethylene by *Escherichia coli*. *J. Gen. Microbiol.* **135**, 1489–1497 (1989).
- [237] Biale, J. B. Effect of Emanations from Several Species of Fungi on Respiration and Color Development of Citrus Fruits. *Science* **91**, 458–459 (1940).
- [238] Miller, E. V., Winston, J. R. & Fisher, D. F. Production of Epinasty by Emanations from Normal and Decaying Citrus Fruits and from *Penicillium digitatum*. *J. Agric. Res.* **60**, 269–277 (1940).
- [239] Young, R. E., Pratt, H. K. & Biale, J. B. Identification of Ethylene as a Volatile Product of the Fungus *Penicillium digitatum*. *Plant Physiol* **26**, 304–310 (1951).
- [240] Jacobsen, D. W. & Wang, C. H. The Biogenesis of Ethylene in *Penicillium digitatum*. *Plant Physiol.* **43**, 1959–1966 (1968).
- [241] Nagahama, K. *et al.* Purification and Properties of an Ethylene-Forming Enzyme from *Pseudomonas syringae* pv. *phaseolicola* PK2. *J. Gen. Microbiol.* **137**, 2281–2286 (1991).
- [242] Fukuda, H. *et al.* Molecular Cloning in *Escherichia coli*, Expression, and Nucleotide Sequence of the Gene for the Ethylene-Forming Enzyme of *Pseudomonas syringae* pv. *phaseolicola* PK2. *Biochem. Biophys. Res. Commun.* **188**, 826–832 (1992).
- [243] Ishihara, K. *et al.* Overexpression and *In vitro* Reconstitution of the Ethylene-Forming Enzyme from *Pseudomonas syringae*. *J. Ferment. Bioengineer.* **79**, 205–211 (1995).
- [244] Pirkov, I., Albers, E., Norbeck, J. & Larsson, C. Ethylene Production by Metabolic Engineering of the Yeast *Saccharomyces cerevisiae*. *Metab. Eng.* **10**, 276–280 (2008).
-

REFERENCES

- [245] Ishihara, K., Matsuoka, M., Ogawa, T. & Fukuda, H. Ethylene Production Using a Broad-Host-Range Plasmid in *Pseudomonas syringae* and *Pseudomonas putida*. *J. Ferment. Bioengineer.* **82**, 509–511 (1996).
- [246] Tao, L., Dong, H. J., Chen, X., Chen, S. F. & Wang, T. H. Expression of Ethylene-Forming Enzyme (EFE) of *Pseudomonas syringae* pv. *glycinea* in *Trichoderma viride*. *Appl. Microbiol. Biotechnol.* **80**, 573–578 (2008).
- [247] Chen, X. *et al.* Overexpression of Bacterial Ethylene-Forming Enzyme Gene in *Trichoderma reesei* Enhanced the Production of Ethylene. *Int. J. Biol. Sci.* **6**, 96–106 (2010).
- [248] Araki, S., Matsuoka, M., Tanaka, M. & Ogawa, T. Ethylene Formation and Phenotypic Analysis of Transgenic Tobacco Plants Expressing a Bacterial Ethylene-Forming Enzyme. *Plant Cell Physiol.* **41**, 327–334 (2000).
- [249] Fukuda, H. *et al.* Heterologous Expression of the Gene for the Ethylene-Forming Enzyme from *Pseudomonas syringae* in the Cyanobacterium *Synechococcus*. *Biotechnol. Lett.* **16**, 1–6 (1994).
- [250] Sakai, M., Ogawa, T., Matsuoka, M. & Fukuda, H. Photosynthetic Conversion of Carbon Dioxide to Ethylene by the Recombinant Cyanobacterium, *Synechococcus* sp. PCC 7942, which Harbors a Gene for the Ethylene-Forming Enzyme of *Pseudomonas syringae*. *J. Ferment. Bioeng.* **84**, 434–443 (1997).
- [251] Takahama, K., Matsuoka, M., Nagahama, K. & Ogawa, T. Construction and Analysis of a Recombinant Cyanobacterium Expressing a Chromosomally Inserted Gene for an Ethylene-Forming Enzyme at the *psbAI* Locus. *J. Biosci. Bioeng.* **95**, 302–305 (2003).
- [252] Ungerer, J. *et al.* Sustained Photosynthetic Conversion of CO₂ to Ethylene in Recombinant Cyanobacterium *Synechocystis* 6803. *Energy & Environmental Science* **5**, 8998 (2012).
- [253] Guerrero, F., Carbonell, V., Cossu, M., Correddu, D. & Jones, P. R. Ethylene Synthesis and Regulated Expression of Recombinant Protein in *Synechocystis* sp. PCC 6803. *PLoS ONE* **7** (2012).
- [254] Fukuda, H., Fujii, T. & Ogawa, T. Preparation of a Cell-Free Ethylene-Forming System from *Penicillium digitatum*. *Agric. Biol. Chem. (Tokyo)* **50**, 977–981 (1986).
- [255] Fukuda, H. *et al.* Two Reactions are Simultaneously Catalyzed by a Single Enzyme: the Arginine-Dependent Simultaneous Formation of Two Products, Ethylene and Succinate, from 2-Oxoglutarate by an Enzyme from *Pseudomonas syringae*. *Biochem. Biophys. Res. Commun.* **188**, 483–489 (1992).

- [256] Baldwin, J. E. *et al.* Substrate Specificity of Proline-4-Hydroxylase: Chemical and Enzymatic Synthesis of (2*S*,3*R*,4*S*)-Epoxyproline. *Tetrahedron Lett.* **35**, 4649–4652 (1994).
- [257] Hsueh, L. C. *Studies on Proline Hydroxylases*. DPhil thesis, Schofield Laboratory, Department of Chemistry, University of Oxford (2000).
- [258] Küllchen, F. B. *Studies on Proline 3-Hydroxylase*. DPhil thesis, Schofield Laboratory, The Dyson Perrins Laboratory, Department of Chemistry, University of Oxford (2002).
- [259] Watson, N. A New Revision of the Sequence of Plasmid pBR322. *Gene* **70**, 399–403 (1988).
- [260] Balbás, P. *et al.* Plasmid Vector pBR322 and its Special-Purpose Derivatives—a Review. *Gene* **50**, 3–40 (1986).
- [261] Lessard, J. C. Growth Media for *E. coli*. *Methods Enzymol.* **533**, 181–189 (2013).
- [262] Monod, J. & Cohen-Bazire, G. The Effect of Specific Inhibition in Biosynthesis of Tryptophan-Desmase by *Aerobacter aerogenes*. *C. R. Hebd. Seances. Acad. Sci.* **236**, 530–532 (1953).
- [263] Morse, D. E., Mosteller, R. D. & Yanofsky, C. Dynamics of Synthesis, Translation, and Degradation of *trp* Operon Messenger RNA in *E. coli*. *Cold Spring Harb. Symp. Quant. Biol.* **34**, 725–740 (1969).
- [264] Ozaki, A. *et al.* Process for Producing *cis*-3-Hydroxy-L-Proline (2004). URL <http://www.freepatentsonline.com/y2004/0229337.html>. US Patent App: 20040229337 (Filed 2004); Assignee: Kyowa Hakko Kogyo Co., Ltd., Japan.
- [265] Ozaki, A., Mori, H., Shibasaki, T., Ando, K. & Chiba, S. Useful as a Raw Material for Medicines or as an Additive to Foods (2007). US Patent: US 7238501 B2 (Filed 2003); Assignee: Kyowa Hakko Kogyo Co., Ltd., Japan.
- [266] New England BioLabs Inc. *IMPACTTM-CN: Instruction Manual*, version 2.1 Edn. (2006).
- [267] Chong, S. *et al.* Single-Column Purification of Free Recombinant Proteins Using a Self-Cleavable Affinity Tag Derived from a Protein Splicing Element. *Gene* **192**, 271–281 (1997).
- [268] Watanabe, T. *et al.* The Roles of the C-Terminal Domain and Type III Domains of Chitinase A1 from *Bacillus circulans* WL-12 in Chitin Degradation. *Journal of Bacteriology* **176**, 4465–4472 (1994).
- [269] Chong, S. *et al.* Utilizing the C-Terminal Cleavage Activity of a Protein Splicing Element to Purify Recombinant Proteins in a Single Chromatographic Step. *Nucleic Acids Res.* **26**, 5109–5115 (1998).

REFERENCES

- [270] EMD Chemicals Inc., an affiliate of Merck KGaA, Darmstadt, Germany. *Novagen pET System Manual*, 11th Edn. (2011). (User Protocol TB055 Rev. C 0611JN, p. 1–63).
- [271] Studier, F. W. & Moffatt, B. A. Use of Bacteriophage T7 RNA Polymerase to Direct Selective High-Level Expression of Cloned Genes. *J. Mol. Biol.* **189**, 113–130 (1986).
- [272] Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA Polymerase to Direct Expression of Cloned Genes. *Methods Enzymol.* **185**, 60–89 (1990).
- [273] Rosenberg, A. H. *et al.* Vectors for Selective Expression of Cloned DNAs by T7 RNA Polymerase. *Gene* **56**, 125–135 (1987).
- [274] Gilbert, W. & Müller-Hill, B. Isolation of the *lac* Repressor. *Proc. Natl. Acad. Sci. U. S. A.* **56**, 1891–1898 (1966).
- [275] Jerpseth, M., Jerspeth, B., Breister, L. & Greener, A. High-Efficiency Derivatives of the BL21 Series for Protein Expression. *Strategies* **11**, 3–4 (1998).
- [276] Agilent Technologies, Inc. *BL21-Gold Competent Cells, BL21-Gold (DE3) Competent Cells, and BL21-Gold (DE3) pLysS Competent Cells: Instruction Manual*, revision b Edn. (2010).
- [277] Jerpseth, B., Callahan, M. & Greener, A. New Competent Cells for Highest Transformation Efficiencies. *Strategies* **10**, 37–38 (1997).
- [278] Tartof, K. D. & Hobbs, C. A. Improved Media for Growing Plasmid and Cosmid Clones. *Bethesda Res. Lab. Focus* **9**, 12–17 (1987).
- [279] Le Châtelier, H. L. Sur un énoncé Général des Lois des équilibres Chimiques. *Comptes Rendus Académie des Sciences* **99**, 786–789 (1884).
- [280] Le Châtelier, H. L. Recherches Expérimentales et Théoriques sur les équilibres Chimiques. *Annales des Mines* **13**, 157–382 (1888).
- [281] Burgess, R. R. Protein Precipitation Techniques. *Methods Enzymol.* **463**, 331–342 (2009).
- [282] GE Healthcare Ltd. *Sepharose™ Fast Flow Ion Exchange Media and Prepacked Column Formats*, 18-1177-22 ae Edn. (2014).
- [283] Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
- [284] Monod, J. *Recherches sur la Croissance des Cultures Bactériennes*. DPhil thesis, Université de Paris, Paris (1941).

- [285] Monod, J. The Phenomenon of Enzymatic Adaptation. *Growth* **11**, 223–289 (1947).
- [286] Loomis, W. F., Jr. & Magasanik, B. Glucose-Lactose Diauxie in *Escherichia coli*. *J. Bacteriol.* **93**, 1397–1401 (1967).
- [287] Epstein, W., Naono, S. & Gros, F. Synthesis of Enzymes of the Lactose Operon during Diauxic Growth of *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **24**, 588–592 (1966).
- [288] Jobe, A. & Bourgeois, S. *lac* Repressor-Operator Interaction. VI. The Natural Inducer of the *lac* Operon. *J. Mol. Biol.* **69**, 397–408 (1972).
- [289] GE Healthcare Bio-Sciences AB, Björkgatan 30, 751 84 Uppsala, Sweden. *Instructions 71-5017-88 AF (GE Healthcare): Ion-Exchange Columns (Mono QTM 5/50 GL and Mono STM 5/50 GL)* (2014).
- [290] Chayen, N. E. Turning Protein Crystallisation from an Art into a Science. *Curr. Opin. Struct. Biol.* **14**, 577–583 (2004).
- [291] McPherson, A. Introduction to Protein Crystallization. *Methods* **34**, 254–65 (2004).
- [292] McPherson, A. & Gavira, J. A. Introduction to Protein Crystallization. *Acta Crystallogr. F. Struct. Biol. Commun.* **70**, 2–20 (2014).
- [293] McPherson, A. *Crystallization of Biological Macromolecules* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999).
- [294] Block, H. *et al.* Immobilized-Metal Affinity Chromatography (IMAC): A Review. *Guide to Protein Purification, Second Edition* **463**, 439–473 (2009).
- [295] Klein, C. & Hüttel, W. Tertiary Alcohol Preferred: Hydroxylation of *trans*-3-Methyl-L-Proline with Proline Hydroxylases. *Beilstein J. Org. Chem.* **7**, 1643–1647 (2011).
- [296] Gileadi, O. pNIC-28-Bsa4 Vector Information Sheet (2009). URL http://www.sgc.utoronto.ca/sgc-webpages/vector_pdf/pnic28-bsa4.pdf.
- [297] Savitsky, P. *et al.* High-Throughput Production of Human Proteins for Crystallization: The SGC Experience. *J. Struct. Biol.* **172**, 3–13 (2010).
- [298] Hehre, E. J. Enzymic Synthesis of Polysaccharides: A Biological Type of Polymerization. *Adv. Enzymol. Relat. Subj. Biochem.* **11**, 297–337 (1951).
- [299] Avigad, G., Feingold, D. S. & Hestrin, S. The Mechanism of Polysaccharide Production from Sucrose. 3. Donor-Acceptor Specificity of Levansucrase from *Aerobacter levanicum*. *Biochem. J.* **64**, 340–351 (1956).

REFERENCES

- [300] Hestrin, S. & Avigad, G. Mechanism of Polysaccharide Production from Sucrose. 5. Transfer of Fructose to C-1 of Aldose by Levansucrase. *Biochem. J.* **69**, 388–398 (1958).
- [301] Ebert, K. H. & Stricker, H. Fermentpolymerisation. 6. Darstellung + Eigenschaften von Hochgereinigten Lavansaccharase Präparationen Aus *Aerobacter Levanicum*. *Zeitschrift Fur Naturforschung Part B-Chemie Biochemie Biophysik Biologie Und Verwandten Gebiete B* **19**, 23–28 (1964).
- [302] Reese, E. T. & Avigad, G. Purification of Levansucrase by Precipitation with Levan. *Biochim. Biophys. Acta.* **113**, 79–83 (1966).
- [303] Gay, P., Le Coq, D., Steinmetz, M., Ferrari, E. & Hoch, J. A. Cloning Structural Gene *sacB*, which Codes for Exoenzyme Levansucrase of *Bacillus subtilis*: Expression of the Gene in *Escherichia coli*. *J. Bacteriol.* **153**, 1424–1431 (1983).
- [304] Gay, P., Lecoq, D., Steinmetz, M., Berkelman, T. & Kado, C. I. Positive Selection Procedure for Entrapment of Insertion-Sequence Elements in Gram-Negative Bacteria. *J. Bacteriol.* **164**, 918–921 (1985).
- [305] Mullis, K. B. *et al.* Process For Amplifying, Detecting, And/or-cloning Nucleic Acid Sequences. US Patent Number: US 4683195 A (1987).
- [306] Saiki, R. K. *et al.* Enzymatic Amplification of β -Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sick Cell Anemia. *Science* **230**, 1350–1354 (1985).
- [307] Bartlett, J. M. S. & Stirling, D. *PCR Protocols*, Vol. 226 of *Methods in Molecular Biology*, Chap. A Short History of the Polymerase Chain Reaction, 3–6 (Humana Press, 2003), 2nd edition Edn.
- [308] Henke, W., Herdel, K., Jung, K., Schnorr, D. & Loening, S. A. Betaine Improves the PCR Amplification of GC-Rich DNA Sequences. *Nucl. Acids Res.* **25**, 3957–3958 (1997).
- [309] Sun, Y., Hegamyer, G. & Colburn, N. H. PCR-Direct Sequencing of a GC-Rich Region by Inclusion of 10% DMSO: Application to Mouse c-Jun. *Biotechniques* **15**, 372–374 (1993).
- [310] Huang, W. M. & Lehman, I. R. On the Exonuclease Activity of Phage T4 Deoxyribonucleic Acid Polymerase. *J. Biol. Chem.* **247**, 3139–3146 (1972).
- [311] Qing, G. *et al.* Cold-Shock Induced High-Yield Protein Production in *Escherichia coli*. *Nat. Biotechnol.* **22**, 877–882 (2004).
- [312] Schein, C. H. Solubility as a Function of Protein Structure and Solvent Components. *Nat. Biotechnol.* **8**, 308–317 (1990).

- [313] Shirano, Y. & Shibata, D. Low Temperature Cultivation of *Escherichia coli* Carrying a Rice Lipoxygenase L-2 cDNA Produces a Soluble and Active Enzyme at a High Level. *FEBS Lett.* **271**, 128–130 (1990).
- [314] Sørensen, H. P. & Mortensen, K. K. Soluble Expression of Recombinant Proteins in the Cytoplasm of *Escherichia coli*. *Microbial Cell Factories* **4**, 1–8 (2005).
- [315] Fakruddin, M., Mohammad Mazumdar, R., Bin Mannan, K. S., Chowdhury, A. & Hossain, M. N. Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant *E. coli*. *ISRN Biotechnology* **2013**, 1–7 (2013).
- [316] Etchegaray, J. P. & Inouye, M. Translational Enhancement by an Element Downstream of the Initiation Codon in *Escherichia coli*. *J. Biol. Chem.* **274**, 10079–10085 (1999).
- [317] Jepson, J. B. & Smith, I. ‘Multiple Dipping’ Procedures in Paper Chromatography: a Specific Test for Hydroxy-proline. *Nature* **172**, 1100–1101 (1953).
- [318] Smith, I. Colour Reactions on Paper Chromatograms by a Dipping Technique. *Nature* **171**, 43–44 (1953).
- [319] Schwartz, D. P. Specific Identification of Hydroxyamino Acids on Paper Chromatograms of Protein Hydrolyzates. *Anal. Chem.* **30**, 1855–1856 (1958).
- [320] Kaule, G. & Günzler, V. Assay for 2-Oxoglutarate-Decarboxylating Enzymes Based on the Determination of [1-¹⁴C]-Succinate: Application to Prolyl 4-Hydroxylase. *Anal. Biochem.* **184**, 291–297 (1990).
- [321] Bayer, E., Grom, E., Kaltenegger, B. & Uhmman, R. Separation of Amino Acids by High Performance Liquid Chromatography. *Anal. Chem.* **48**, 1106–1109 (1976).
- [322] Zimmerman, C. L., Appella, E. & Pisano, J. J. Rapid Analysis of Amino Acid Phenylthiohydantoins by High-Performance Liquid Chromatography. *Anal. Biochem.* **77**, 569–573 (1977).
- [323] Lindroth, P. & Mopper, K. High Performance Liquid Chromatographic Determination of Subpicomole Amounts of Amino Acids by Precolumn Fluorescence Derivatization with *o*-Phthaldialdehyde. *Anal. Chem.* **51**, 1667–1674 (1979).
- [324] Larsen, B. R. & West, F. G. A Method for Quantitative Amino Acid Analysis Using Precolumn *o*-Phthalaldehyde Derivatization and High Performance Liquid Chromatography. *J. Chromatogr. Sci.* **19**, 259–265 (1981).

REFERENCES

- [325] Cooper, J. & Turnell, D. Fluorescence Detection of Cystine by *o*-Phthalaldehyde Derivatisation and its Separation Using High-Performance Liquid Chromatography. *J. Chromatogr. B Biomed. Sci. Appl.* **227**, 158–161 (1982).
- [326] Yokotsuka, K. & Kushida, T. An Improved Method for Fluorometric Amino Acid Analysis including Proline and Hydroxyproline: Chemical Studies on Coloring and Flavoring Substances in Japanese Grapes and Wines (XIV). *J. Ferment. Technol.* **61**, 1–6 (1983).
- [327] Cooper, J., Lewis, M. & Turnell, D. Pre-Column *o*-Phthalaldehyde Derivatization of Amino Acids and their Separation Using Reversed-Phase High-Performance Liquid Chromatography : II. Simultaneous Determination of Amino and Imino Acids in Protein hydrolysates. *J. Chromatogr. A* **285**, 490–494 (1984).
- [328] Kunte, H. J., Galinski, E. A. & Trueper, H. G. A Modified FMOC-Method for the Detection of Amino Acid-Type Osmolytes and Tetrahydropyrimidines (Ectoines). *J. Microbiol. Methods* **17**, 129–136 (1993).
- [329] Ozaki, A., Shibasaki, T. & Mori, H. Specific Proline and Hydroxyproline Detection Method by Post-Column Derivatisation for High-Performance Liquid Chromatography. *Biosci. Biotechnol. Biochem.* **59**, 1764–1765 (1995).
- [330] Harada, K.-I., Fujii, K., Mayumi, T., Hibino, Y. & Suzuki, M. A Method Using LC/MS for Determination of Absolute Configuration of Constituent Amino Acids in Peptide: Advanced Marfey's Method. *Tetrahedron Lett.* **36**, 1515–1518 (1995).
- [331] Harada, K.-I. *et al.* Application of D,L-FDLA Derivatisation to Determination of Absolute Configuration of Constituent Amino Acids in Peptide by Advanced Marfey's Method. *Tetrahedron Lett.* **37**, 3001–3004 (1996).
- [332] Fujii, K., Ikai, H., Oka, H., Suzuki, M. & Harada, K.-I. A Nonempirical Method Using LC/MS for Determination of the Absolute Configuration of Constituent Amino Acids in a Peptide: Combination of Marfey's Method with Mass Spectrometry and its Practical Application. *Anal. Chem.* **69**, 5146–5151 (1997).
- [333] Edman, P. Method for Determination of the Amino Acid Sequence in Peptides. *Acta Chem. Scand.* **4**, 283–293 (1950).
- [334] Lindblad, W. J. & Diegelmann, R. F. Quantitation of Hydroxyproline Isomers in Acid Hydrolysates by High-Performance Liquid Chromatography. *Anal. Biochem.* **138**, 390–395 (1984).
- [335] Claridge, T. D. W. *High-Resolution NMR Techniques in Organic Chemistry*, Vol. 27 of *Tetrahedron Organic Chemistry* (Elsevier Science Ltd, Linacre House, Jordan Hill, Oxford, U.K. OX2 8DP, 2008), 2nd Edn.

- [336] Wong, K. C. Review of NMR Spectroscopy: Basic Principles, Concepts and Applications in Chemistry. *J. Chem. Educ.* **91**, 1103–1104 (2014).
- [337] Hopkinson, R. J., Hamed, R. B., Rose, N. R., Claridge, T. D. W. & Schofield, C. J. Monitoring the Activity of 2-Oxoglutarate Dependent Histone Demethylases by NMR Spectroscopy: Direct Observation of Formaldehyde. *Chem-BioChem* **11**, 506–510 (2010).
- [338] Leung, I. K. H. *et al.* Structural and Mechanistic Studies on γ -Butyrobetaine Hydroxylase. *Chem. Biol.* **17**, 1316–1324 (2010).
- [339] Leung, I. K. H. *et al.* Reporter Ligand NMR Screening Method for 2-Oxoglutarate Oxygenase Inhibitors. *J. Med. Chem.* **56**, 547–555 (2012).
- [340] DeRitter, E. Vitamins in Pharmaceutical Formulations. *J. Pharm. Sci.* **71**, 1073–1096 (1982).
- [341] British Pharmacopoeia. Tech. Rep., Her Majesty's Stationery Office, London, UK (2009).
- [342] Sweetman, S. C. *Martindale: The Complete Drug Reference* (Pharmaceutical Press, London, UK, 2009), 36th Edn.
- [343] Ahmad, I., Sheraz, M. A., Shaikh, R. H., Ahmed, S. & Vaid, F. H. M. Photostability of Ascorbic Acid in Aqueous and Organic Solvents. *J. Pharm. Res.* **3**, 1237–1239 (2010).
- [344] Ahmad, I. *et al.* Photostability and Interaction of Ascorbic Acid in Cream Formulations. *AAPS PharmSciTech.* **12**, 917–923 (2011).
- [345] Sheraz, M. A., Ahmed, S., Ahamd, I., Qadeer, K. & Vaid, F. H. M. Photodegradation and Photostabilisation of Ascorbic Acid in Pharmaceutical Preparations. *Int. J. Chem. Anal. Sci.* **1**, 68–70 (2010).
- [346] Sheraz, M. A. *et al.* Formulation and Stability of Ascorbic Acid in Topical Preparations. *Syst. Rev. Pharm.* **2**, 86–90 (2011).
- [347] Walker, N. Paramagnetic Properties of Fe(II) and Fe(III). *J. Chem. Educ.* **54**, 431 (1977).
- [348] Leung, I. K. H. *NMR Studies on 2-OG Oxygenases and Related Enzymes*. DPhil thesis, Department of Chemistry, University of Oxford (2012).
- [349] Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. Electrospray Ionisation for Mass Spectrometry of Large Biomolecules. *Science* **246**, 64–71 (1989).
- [350] Bruins, A. P. Mechanistic Aspects of Electrospray Ionisation. *J. Chromatogr. A* **794**, 345–357 (1998).

REFERENCES

- [351] Ho, C. S. *et al.* Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin. Biochem. Rev.* **24**, 3–12 (2003).
- [352] Korfmacher, W. A. Foundation Review: Principles and Applications of LC-MS in New Discovery. *Drug Discov. Today* **10**, 1357–1367 (2005).
- [353] Pitt, J. J. Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *Clin. Biochem. Rev.* **30**, 19–34 (2009).
- [354] SIELC Technologies Inc., 15 East Palatine Road, Suite 107, Prospect Heights, Illinois, USA. 60070. *PrimesepTM: Columns, Methods, Applications* (2004).
- [355] Cecchi, T. Ion Pairing Chromatography. *Crit. Rev. Anal. Chem.* **38**, 161–213 (2008).
- [356] Batchelar, E. T. *et al.* Thioester Hydrolysis and C—C Bond Formation by Carboxymethylproline Synthase from the Crotonase Superfamily. *Angew. Chem. Int. Ed. Engl.* **47**, 9322–9325 (2008).
- [357] Hamed, R. B., Mecinović, J., Ducho, C., Claridge, T. D. W. & Schofield, C. J. Carboxymethylproline Synthase Catalysed Syntheses of Functionalised *N*-Heterocycles. *Chem. Commun.* **46**, 1413–1415 (2010).
- [358] Hamed, R. B. *et al.* Stereoselective C—C Bond Formation Catalysed by Engineered Carboxymethylproline Syntheses. *Nat. Chem.* **3**, 365–371 (2011).
- [359] Hamed, R. B. *et al.* Crotonase Catalysis Enables Flexible Production of Functionalised Prolines and Carbapenams. *J. Am. Chem. Soc.* **134**, 471–479 (2012).
- [360] Hamed, R. B. *et al.* Stereoselective Preparation of Lipidated Carboxymethylproline/Pipelicolic Acid Derivatives *via* Coupling of Engineered Crotonases with an Alkylmalonyl-CoA Synthase. *Org. Biomol. Chem.* **11**, 8191–8196 (2013).
- [361] Hamed, R. B., Batchelar, E. T., Mecinović, J., Claridge, T. D. W. & Schofield, C. J. Evidence that Thienamycin Biosynthesis Proceeds *via* C-5 Epimerisation: ThnE Catalyses the Formation of (2*S*,5*S*)-*trans*-Carboxymethylproline. *ChemBioChem* **10**, 246–250 (2008).
- [362] McGowan, S. J., Bycroft, B. W. & Salmond, G. P. Bacterial Production of Carbapenems and Clavams: Evolution of β -Lactam Antibiotic Pathway. *Trends Microbiol.* **6**, 203–208 (1998).
- [363] Coulthurst, S. J., Barnard, A. M. L. & Salmond, G. P. C. Regulation and Biosynthesis of Carbapenem Antibiotics in Bacteria. *Nature Rev. Microbiol.* **3**, 295–306 (2005).

-
- [364] Núñez, L. E., Méndez, C., Braña, A. F., Blanco, G. & Salas, J. A. The Biosynthetic Gene Cluster for the β -Lactam Carbapenem Thienamycin in *Streptomyces cattleya*. *Chem. Biol.* **10**, 301–311 (2003).
- [365] Freeman, M. F., Moshos, K. A., Bodner, M. J., Li, R. & Townsend, C. A. Four Enzymes Define the Incorporation of Coenzyme A in Thienamycin Biosynthesis. *Proc. Natl. Acad. Sci. U S A* **105**, 11128–11133 (2008).
- [366] Baldwin, J. E. & Schofield, C. J. *The Chemistry of β -Lactams*, Chap. 1: The Biosynthesis of β -Lactams, 10–26 (Blackie Academic & Professional, London, 1992).
- [367] Pascal, R. J., Oliver, M. A. & Chen, Y. C. Alternate Substrates and Inhibitors of Bacterial 4-Hydroxyphenylpyruvate Dioxygenase. *Biochemistry* **24**, 3158–3165 (1985).
- [368] Hopkinson, R. J. *et al.* Is JmjC Oxygenase Catalysis Limited to Demethylation? *Angew. Chem. Int. Ed. Engl.* **52**, 7709–7713 (2013).
- [369] Thornburg, L. D. & Stubbe, J. Mechanism-Based Inactivation of Thymine Hydroxylase, an α -Ketoglutarate-Dependent Dioxygenase, by 5-Ethynyluracil. *Biochemistry* **32**, 14034–14042 (1993).
- [370] Han, H. & Pascal, R. A. J. α -Ketoisocaproate Dioxygenase: The Stereochemical Course of the Hydroxylation Reaction. *J. Org. Chem.* **55**, 5,173–5,176 (1990).
- [371] Hashimoto, T. *et al.* Hyoscyamine 6 β -Hydroxylase, an Enzyme Involved in Tropane Alkaloid Biosynthesis, Is Localised at the Pericycle of the Root. *J. Biol. Chem.* **266**, 4,648–4,653 (1991).
- [372] Remuzon, P. *Trans*-4-Hydroxy-L-Proline, a Useful and Versatile Chiral Starting Block. *Tetrahedron* **52**, 13803–13835 (1996).
- [373] Cahn, R. S., Ingold, C. K. & Prelog, V. Specification of Molecular Chirality. *Angew. Chem. Int. Ed. Engl.* **5**, 385–415 (1966).
- [374] Prelog, V. & Helmchen, G. Basic Principles of the CIP-System and Proposals for a Revision. *Angew. Chem. Int. Ed. Engl.* **21**, 567–583 (1982).
- [375] Gavras, H. *et al.* Antihypertensive Effect of the Oral Angiotensin-Converting Enzyme Inhibitor SQ 14225 in Man. *N. Engl. J. Med.* **298**, 991–995 (1978).
- [376] Bindra, J. S. & Lednicer, D. (eds.) *Chronicals of Drug Discovery*, Vol. 2 (John Wiley and Sons, 605 Third Avenue, New York, NY 10158, 1983).
- [377] Bryan, J. From Snake Venom to ACE Inhibitor - The Discovery and Rise of Captopril. *Pharm. J.* **282**, 455 (2009).
-

REFERENCES

- [378] Loenarz, C. *et al.* Evidence for a Stereoelectronic Effect in Human Oxygen Sensing. *Angew. Chem. Int. Ed. Engl.* **48**, 1784–1787 (2009).
- [379] Stapon, A., Li, R. & Townsend, C. A. Carbapenem Biosynthesis: Confirmation of Stereochemical Assignments and the Role of CarC in the Ring Stereoconversion Process from L-Proline. *J. Am. Chem. Soc.* **125**, 8486–8493 (2003).
- [380] Hughes, A. J. & Keatinge-Clay, A. Enzymatic Extender Unit Generation for *In Vitro* Polyketide Synthase Reactions: Structural and Functional Showcasing of *Streptomyces coelicolor* MatB. *Chem. Biol.* **18**, 165–176 (2011).
- [381] Koketsu, K. *et al.* Refined Regio- and Stereoselective Hydroxylation of L-Pipecolic Acid by Protein Engineering of L-Proline *cis*-4-Hydroxylase Based on the X-ray Crystal Structure. *ACS Synth. Biol.* **4**, 383–392 (2015).
- [382] Bergfors, T. *Protein Crystallization: Techniques, Strategies and Tips* (International University Line, La Jolla, CA., 1999).
- [383] Ducruix, A. & Giége, R. *Crystallization of Nucleic Acids and Proteins, A Practical Approach*. (IRL Press, Oxford, UK; New York., 1992).
- [384] Chowdhury, R. *et al.* Structural Basis for Binding of Hypoxia-Inducible Factor to the Oxygen-Sensing Prolyl Hydroxylases. *Structure* **17**, 981–989 (2009).
- [385] Elkins, J. M. *et al.* Structure of Factor-inhibiting Hypoxia-inducible Factor (HIF) Reveals Mechanism of Oxidative Modification of HIF-1 α . *J. Biol. Chem.* **278**, 1802–1806 (2003).
- [386] Hewitson, K. S. *et al.* Evidence That Two Enzyme-Derived Histidine Ligands Are Sufficient for Iron Binding and Catalysis by Factor Inhibiting HIF (FIH). *J. Biol. Chem.* **283**, 25971–25978 (2008).
- [387] Hopkinson, A., R. J. an Tumber *et al.* 5-Carboxy-8-Hydroxyquinoline is a Broad Spectrum 2-Oxygenase Inhibitor which Causes Iron Translocation. *Chem. Sci.* **4**, 3110–3117 (2013).
- [388] Kuai, J. & Dilley, D. R. Extraction, Partial Purification, and Characterisation of 1-Aminocyclopropane-1-Carboxylic Acid Oxidase from Apple Fruit. *Postharvest Biol. Technol.* **1**, 203–211 (1992).
- [389] Fernández-Maculet, J. C. & Yang, S. F. Extraction and Partial Characterisation of the Ethylene-Forming Enzyme from Apple Fruit. *Plant Physiol.* **99**, 751–754 (1992).
- [390] Wilson, I. D., Zhu, Y., Burmeister, D. M. & Dilley, D. R. Apple Ripening-Related cDNA Clone pAP4 Confers Ethylene-Forming Ability in Transformed *Saccharomyces cerevisiae*. *Plant Physiol.* **102**, 783–788 (1993).

-
- [391] Spanu, P., Reinhardt, D. & Boller, T. Analysis and Cloning of the Ethylene-Forming Enzyme from Tomato by Functional Expression of its mRNA in *Xenopus laevis* oocytes. *EMBO J.* **10**, 2007–2013 (1991).
- [392] Hamilton, A. J., Bouzayen, M. & Grierson, D. Identification of a Tomato Gene for the Ethylene-Forming Enzyme by Expression in Yeast. *Proc. Natl. Acad. Sci. U S A* **88**, 7434–7437 (1991).
- [393] Zhang, Z., Schofield, C. J., Baldwin, J. E., Thomas, P. & John, P. Expression, Purification, and Characterization of 1-Aminocyclopropane-1-Carboxylate Oxidase from Tomato in *Escherichia coli*. *Biochem. J.* **307**, 77–85 (1995).
- [394] McGarvey, D. J., Yu, H. & Christoffersen, R. E. Nucleotide Sequence of a Ripening-Related cDNA from Avocado Fruit. *Plant Mol. Biol.* **15**, 165–167 (1990).
- [395] McGarvey, D. J. & Christoffersen, R. E. Characterisation and Kinetic Parameters of Ethylene-Forming Enzyme from Avocado Fruit. *J. Biol. Chem.* **267**, 5964–5967 (1992).
- [396] Ververidis, P. & John, P. Complete Recovery *In Vitro* of Ethylene-Forming Enzyme Activity. *Phytochemistry* **30**, 725–727 (1991).
- [397] Smith, J. J., Ververidis, P. & John, P. Characterisation of the Ethylene-Forming Enzyme Partially Purified from Melon. *Phytochemistry* **31**, 1485–1494 (1992).
- [398] Wang, H. & Woodson, W. R. A Flower Senescence-Related mRNA from Carnation Shares Sequence Similarity with Fruit Ripening-Related mRNAs Involved in Ethylene Biosynthesis. *Plant Physiol.* **96**, 1000–1001 (1991).
- [399] Peiser, G. D. *et al.* Formation of Cyanide from Carbon 1 of 1-Aminocyclopropane-1-Carboxylic Acid during its Conversion to Ethylene. *Proc. Natl. Acad. Sci. U S A* **81**, 3059–3063 (1984).
- [400] Primrose, S. B. Ethylene and Agriculture: the Role of the Microbe. *J. Appl. Bacteriol.* **46**, 1–25 (1979).
- [401] Primrose, S. B. & Dilworth, M. J. Ethylene Production by Bacteria. *J. Gen. Microbiol.* **93**, 177–181 (1976).
- [402] Primrose, S. B. Evaluation of the Role of Methional, KMBA and Peroxidase in Ethylene Formation by *Escherichia coli*. *J. Gen. Microbiol.* **98**, 519–528 (1977).
- [403] Ince, J. E. & Knowles, C. J. Ethylene Formation by Cultures of *Escherichia coli*. *Arch. Microbiol.* **141**, 209–213 (1985).
-

REFERENCES

- [404] Kang, B. G. *Physiology of Movements*, Vol. 7 of *Encyclopedia of Plant Physiology*, Chap. Epinasty, 647–667 (Springer-Verlag, Berlin, Germany, 1979).
- [405] Palmer, J. H. *Hormone Regulation of Development III*, Vol. 11 of *Encyclopedia of Plant Physiology*, Chap. Epinasty, Hyponasty, and Related Topics, 139–168 (Springer-Verlag, Berlin, Germany, 1985).
- [406] Krebs, H. A. The Intermediary Stages in the Biological Oxidation of Carbohydrate. *Adv. Enzymol. Relat. Areas Mol.* **3**, 191–252 (1943).
- [407] Nagahama, K., Ogawa, T., Fujii, T., Goto, M. & Fukuda, H. L-Arginine is Essential for the Formation *In Vitro* of Ethylene by an Extract of *Pseudomonas syringae*. *J. Gen. Microbiol.* **137**, 1641–1646 (1991).
- [408] Levine, R. L. Oxidative Modification of Glutamine Synthetase. I. Inactivation is due to Loss of One Histidine Residue. *J. Biol. Chem.* **258**, 11823–11827 (1983).
- [409] Farrant, R. D., Walker, V., Mills, G. A., Mellor, J. M. & Langley, G. J. Pyridoxal Phosphate De-activation by Pyrroline-5-Carboxylic Acid. Increased Risk of Vitamin B6 Deficiency and Seizures in Hyperprolinemia Type II. *J. Biol. Chem.* **276**, 15107–15116 (2001).
- [410] Soman, G., Hurst, M. O. & Graves, D. J. Chemical Modification of Canavanine with *p*-Nitrophenylglyoxal. Factors Influencing the Chemistry and Reactivity of α -Dicarbonyl-Guanidino Reactions. *Int. J. Pept. Protein Res.* **25**, 517–525 (1985).
- [411] Worthen, D. R. & Crooks, P. A. *Herbal Radiomodulators: Applications in Medicine, Homeland Defence and Space*, Chap. Chapter 18: The Radiosensitising Effects of L-Canavanine, 255–269 (CABI, Wallingford, Oxfordshire, UK, 2008).
- [412] Krojer, T. J. *et al.* Crystal Structure of Human Gamma-Butyrobetaine, 2-Oxoglutarate Dioxygenase 1 (BBOX1). *RCSB Protein Data Bank* **3MS5** (2010). URL <http://www.rcsb.org/pdb/explore.do?structureId=3ms5>.
- [413] Hanahan, D. Studies on Transformation of *Escherichia coli* with Plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).