

HALOGENASES FOR BIOSYNTHESIS AND BIOCATALYSIS

A thesis submitted to the University of Manchester for the degree of
Doctor of Philosophy
in the Faculty of Engineering and Physical Sciences.

2013

Chinnan Velmurugan Karthikeyan

**School of Chemistry and
Manchester Institute of Biotechnology**

Table of Contents

Table of Contents	2
List of tables	8
List of Figures	10
List of schemes	16
Abbreviations	17
Abstract	19
Declaration	20
Copyright Statement	20
Acknowledgements	21
The Author	22
Chapter 1 Introduction	23
1.1 Halogenation	23
1.2. Biocatalysis: a general overview	25
1.3. Biocatalyst engineering and the host system	26
1.4. Biohalogenases	27
1.5. Halogenases for biocatalysis	32
1.6. Flavin dependent halogenases	37
1.6.1. Mechanism of flavin dependent halogenases	38
1.6.2. Enzymes characterised so far and their substrate scope	39
1.6.3. Crystallographic studies of flavin dependent halogenases	42
1.6.4. Regioselectivity studies based on crystal structures	46
1.6.5 Structural evidence of halogenation	48
1.6.6. Conclusion	48
1.7. Aims	49
1.7.1. Aim 1	49
1.7.2. Aim 2	50
Chapter 2 Construction of an <i>in-vitro</i> halogenation system for chlorination of Hpg-17: results and discussion	51
2.1. Introduction	51
2.2. Cloning and expression of the small PCP (S-PCP) domain in <i>Escherichia coli</i> (<i>E. coli</i>)	53
2.3. Cloning and expression of the large PCP domain (L-PCP) in <i>E. coli</i>	54
2.4. Synthesis of co-enzyme A-Hpg (CoA-Hpg) and N-acetylcysteamine thioester-Hpg (SNAC-Hpg)	58
2.5. Substrate loading and analysis of PCP and PCP tethered Hpg by MALDI	60
2.6. Ram20- bioinformatics, cloning and expression in <i>E. coli</i>	62

2.7.	Flavin reductases cloning and expression	64
2.7.1.	Cloning and expression of <i>ssuE</i>	65
2.7.2.	Determination of reductase activity of purified His ₍₆₎ -SsuE	66
2.7.3.	Cloning and expression of flavin reductase Fre	67
2.8.	Halogenation assay of Ram20 on Hpg, SNAC-Hpg and thioester Hpg	68
Chapter 3	Flavin dependent Halogenases for biocatalysis: result and discussion	69
3.1.	KtzQ halogenase	69
3.1.1.	Cloning and expression of KtzQ in <i>E. coli</i>	70
3.1.2.	Identification of KtzQ by trypsin digestion: mass spectrometry	74
3.1.3.	Enzymatic assay of KtzQ with L-tryptophan	76
3.2.	PrnA	77
3.2.1.	PrnA mass spectrometry	77
3.2.2.	PrnA enzymatic assay for substrate scope	78
3.2.3.	Protein engineering to widen substrate scope	80
3.2.4.	Enzyme assay of mutant proteins with substrates	82
3.2.5.	Assay of PrnA mutant's with anthranilic acid	86
3.2.6.	Optimisation of protein purification of PrnA for crystallographic studies	86
3.2.7.	Multiangle light scattering (MALS) of PrnA	92
3.2.8.	Crystal structure of PrnA F454K mutant	92
3.3.	KtzR (tryptophan-6-halogenase)	97
3.3.1.	Cloning and expression of KtzR	97
3.3.2.	Re-cloning and expression of KtzR	99
3.3.3.	Halogenase activity of KtzR on tryptophan	100
3.3.4.	Purification of KtzR for obtaining crystal structure	101
3.4.	SttH (tryptophan-6-halogenase)	102
3.4.1.	Cloning and expression of tryptophan-6-halogenase SttH	103
3.4.2.	Substrate specificity of SttH enzyme	104
3.4.3.	Regioselectivity of SttH	105
3.4.3.1.	Enzyme assay of SttH F98A	107
3.4.3.2.	Enzyme assay of SttH loop (SttH LI) inserted mutant for regioselectivity	108
3.4.4.	SttH for crystal trials	109
3.4.4.1.	Purification of SttH for crystal trials	109
3.5.	PyrH (tryptophan-5-halogenase)	116
3.5.1.	Cloning and expression of tryptophan-5-halogenase PyrH	116
3.5.2.	Stability of purified PyrH in DMSO and ethanol	118

3.5.3.	PyrH Substrate specificity	121
3.5.3.1.	Halogenase activity of PyrH with tryptophan	121
3.5.3.2.	Halogenase activity of PyrH with kynurenine	125
3.5.3.3.	Halogenase activity of PyrH with anthranilamide	129
3.5.3.4.	Halogenase activity of PyrH with 2-amino-4-methylbenzamide	131
3.5.3.5.	Overall relative activity of PyrH with the substrates	134
3.6.	Regioselective studies on PrnA and PyrH	137
3.6.1.	Engineering PrnA to regioselectivity chlorinate at the 5-position based on PyrH	137
3.6.2.	Enzymatic assay of PrnA mutants with tryptophan for regioselectivity	140
3.6.3.	Engineering PyrH to chlorinate 7-position based on PrnA	141
3.6.4.	Enzyme assays of single, double and triple mutant for regioselectivity	142
3.6.5.	Enzyme assay of PyrH loop inserted mutant for regioselectivity	146
3.6.6.	Insertion of 8 the amino acid loop in PyrH for regioselectivity	146
3.6.7.	Enzyme assay of PyrH loop inserted mutant for regioselectivity	147
Chapter 4.	Summary and future work	148
Chapter 5.	General materials and methods	152
5.1.	General information	152
5.1.1.	Chemicals, enzymes and kits	152
5.2.	Stock solutions for experiments	152
5.2.1.	Antibiotics solution	152
5.2.2.	Substrates stock solution	153
5.3.	Preparation of competent cells using calcium chloride	154
5.3.1.	Materials required	154
5.3.2.	Procedure	154
5.4.	Preparation of competent cells using Innoue solution	155
5.4.1.	Materials required	155
5.4.2.	Procedure	155
5.5.	Growth of cells	156
5.5.1.	<i>Kutzneria spp.</i> 744	156
5.5.2.	Growth of <i>E. coli</i> BL21 (DE3) cells	156
5.5.3.	Growth of <i>Actinoplanes</i> ATCC30766	156
5.6.	Isolation of genomic DNA	157
5.7.	Plasmid DNA extraction	157
5.8.	Agarose gel Electrophoresis	157

5.9.	PCR amplification of genomic DNA	158
5.9.1.	PCR amplification and conditions	158
5.10.	Gel-extraction of DNA from Agarose Gels	159
5.11.	Digestion, Ligation and Transformation	159
5.12.	Ligation Reaction	160
5.13.	Transformation	160
5.13.1.	Analysis of the transformed colonies	160
5.13.2.	Sequencing	161
5.14.	Site directed mutagenesis	161
5.14.1.	Site directed mutagenesis to insert and delete an 8 amino acid loop region	162
5.15.	Protein Expression and purification	163
5.15.1.	Protein Expression Studies	163
5.15.2.	Protein expression using <i>E. coli</i> BL21 (DE3) cells	164
5.15.3.	Protein expression using <i>E. coli</i> Rosetta2 (DE3) RP cells	164
5.15.4.	Protein expression using <i>E. coli</i> Rosetta2 cells	164
5.15.5.	Protein expression using <i>E. coli</i> ArcticExpress (DE3) RP cells	164
5.15.6.	Isolation, purification and concentration of protein	165
5.15.6.1.	Protein purification by HiTrap™ Chelating HP column using AKTA FPLC system	166
5.15.6.2.	Protein purification by Gravity flow method	168
5.15.6.3.	Removal of chaperones by ATP Buffer method	168
5.15.6.4.	Chromatographic techniques	168
5.15.6.4.1.	Anion exchange chromatography	168
5.15.6.4.2.	Size exclusion chromatography	169
5.15.6.5.	Purification of protein using HRV 3C protease	170
5.15.6.6.	Purification using hydrophobic column	171
5.15.7.	Buffer exchange and concentration	171
5.16.	Sodium-Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)	172
5.17.	Determination of protein concentration by Bradford Assay	173
5.18.	Mass spectrometry analysis by trypsin cleavage	173
5.19.	Western blot analysis	174
5.20.	Substrate loading assay	175
5.21.	FMN reductase assay	175
5.22.	Halogenase assays for chlorination and bromination	175
5.23.	Enzyme stability assays in solvent environment	175
5.24.	High pressure liquid chromatography and method	176
5.25.	Mass spectrometry of the HPLC samples	177
5.26.	Preparation of samples for NMR	177

5.27.	NMR results	178
5.28.	Crystal soaking and crystal trial conditions	179
Chapter 6	Reference	180
	Appendix	191

List of Table

Table No	Title	Page No
1.	BLAST analysis of Ram20 showing identical proteins	62
2.	BLAST analysis of KtzQ showing identity with tryptophan halogen	69
3.	Scores obtained for matching proteins by MASCOT analysis	75
4.	Scores obtained for matching proteins by MASCOT analysis	78
5.	Previous mutations carried out on PrnA for regioselective chlorination	80
6.	PrnA mutant protein concentrations for widening substrates	82
7.	Table of substrates tested against various PrnA mutants	83
8.	PrnA mutant protein concentrations for anthranilic acid.	85
9.	Structural parameters obtained for PrnA F454K mutant	93
10.	Overall relative activity of SttH mutants obtained for regioselective studies	109
11.	Overall relative activity of halogenation by PyrH on various substrates	134
12.	Differences in interaction distance between tryptophan and kynurenine by docking experiments	135
13.	Differences in interaction distance between tryptophan and 2-amino-4-methylbenzamide by docking experiments	136
14.	PrnA mutant protein concentrations obtained for regioselective studies	139
15.	Overall relative activity of the PrnA mutants obtained for regioselective studies	141
16.	Protein concentration of PyrH mutants for switching regioselectivity	142
17.	Relative activity of PyrH mutants with tryptophan for halogenation	143
18.	Antibiotic stock solutions	152
19.	Substrates stock solution	153
20.	Cocktail mixture of PCR reaction	158
21.	PCR cycle used in thermocycler	159
22.	Restriction digestion reactions mixture	159
23.	Ligation reactions mixture	160
24.	PCR reactions mixture of performing colony PCR	161
25.	Thermocycler conditions used for colony PCR	161
26.	PCR reactions mixture of performing site directed mutagenesis	162
27.	Thermocycler conditions used for site directed mutagenesis	162
28.	Phosphate salts required for obtaining 100mM phosphate buffer of desired pH	166

29.	Amount of imidazole required for preparation of desired concentration of buffers	169
30.	Reaction mixture for HRV 3C protease cleavage	170
31.	SDS-PAGE recipe	172
32.	List of primers used for the halogenation studies	191
33.	List of constructs made for the halogenation studies	193

List of Figures

Figure No	Figure title	Page No
1.	First identified and isolated halometabolites	24
2.	Examples of halogenated compounds used in pharmaceutical and agrochemical industries	24
3.	Structures of ramoplaninA1, A2 and A3 (Ornornithine, Chp-chlorinated hydroxyphenyl glycine, Hpg are highlighted in circle and mannose group in red)	29
4.	Peptidoglycan biosynthesis showing initial formation of pentapeptide followed by formation of lipid I, lipid II and transglycosylation and transpeptidation step	30
5.	Structure of U shaped ramoplanin (PDB ID:1DSR) showing residues inside the disc (orange), the fatty acyl chain (red) and other residues lying exterior (green)	31
6.	Structure of the active site of heme chloroperoxidase from <i>Caldariomyces fumago</i> showing the heme complex bound at conserved C29 residue allowing 1,3-cyclopentanedione (CPD) to bind the complex (PDB ID: 2CIX)	33
7.	Structure of the active site of a vanadium chloroperoxidase from <i>Cur. inaequalis</i> (PDB ID:1IDU) showing vanadium (silver) bound to the conserved H496 residue and interaction with R360,R490, S402,G403 and K353 creating overall positive charge in the site	35
8.	Structure of the active site of SyrB2 from <i>P. syringae</i> (PDB ID:2FCT) generated using PYMOL software showing Fe- α -ketoglutarate complex bound to the conserved H116, A118, H235(Orange- α -ketoglutarate, red sphere- Fe and purple- amino acid residues)	37
9.	Crystal structure of overlaid PrnA(PDB ID:2AR8) (green) and RebH(PDB ID:2OA1) (yellow) showing the trigonal shape, loops that adopt different conformations are highlighted in red for RebH	42
10.	Crystal structures showing similar FAD, Cl ⁻ , tryptophan, K79 and E346/E347 binding site in overlaid of RebH (PDB ID:2OA1) and PrnA(2AR8) generated using PYMOL software (orange aa residues-	43

	RebH, green aa residues-PrnA, cross reds- tryptophan in RebH, blue-chlorinated tryptophan by PrnA)	
11.	Active site showing similar interactions in overlaid crystal structures of RebH (PDB ID:2E4G) and PrnA(2AR8) generated using PYMOL software ¹ (orange aa residues-RebH, green aa residues-PrnA, magenta-tryptophan in RebH, blue- chlorinated tryptophan in PrnA)	44
12.	Crystal structure of PyrH (PDB ID: 2WET) generated using PYMOL showing the interactions of tryptophan with residues in the active site pocket	45
13.	Crystal structure of (PrnA PDB ID:2AR8) generated using PYMOL showing residue F103, H101 and W455 forming π stacking (orange) around the tryptophan and the residues(green) providing the hydrogen bond interactions holding 7-Cl-tryptophan (blue)	46
14.	Regioselectivity of PyrH(PDB ID:2WET) generated using PYMOL showing, Y454, P93 and F94(orange) forming sandwich residues around tryptophan, red-residue which would allow space for the different orientation of tryptophan relative to tryptophan-7-halogenases, other interacting residues (green)	47
15.	Crystal structure(PDB ID:2AR8) generated using PYMOL showing Cl ⁻ bound to the face of FAD with 10 Å away from the product 7-Cl-trp(cyan), and the residues forming the tunnel between the FAD and substrate binding sites are shown in red	48
16.	Structure of ramoplanin A3 showing chlorination of Hpg-17 by Ram20	51
17.	A) PCR amplification of 420 bp <i>s-pcp</i> and B) Restriction digestion analysis of <i>s-pcp</i> cloning	53
18.	Analysis of pure S-PCP protein by 20% (v/v) SDS-PAGE	54
19.	A) PCR amplification of 975 bp <i>l-pcp</i> and B) Restriction digestion analysis of <i>l-pcp</i> cloning	54
20.	DNA sequencing chromatogram showing mutations of CGA to CGC in <i>l-pcp</i> . (A) and (B) – <i>l-pcp</i> wild type sequence with CGA codons (A1) & (B1) – <i>l-pcp</i> mutant sequence with CGC codons. The relevant codons are underlined in both the wild type and mutant sequences. Colours (Red- T, Blue- C, Black-G, Green-A)	56
21.	Expression studies of L-PCP protein analysed by 12%	57

	(v/v) SDS-PAGE	
22.	Analysis of pure L-PCP protein by 15% (v/v) SDS-PAGE	58
23.	Analysis of pure Sfp protein by 20% (v/v) SDS-PAGE	60
24.	MALDI of S-PCP	61
25.	MALDI analysis of CoA-Hpg loading on S-PCP (Red-S-PCP, Blue-S-PCP loaded with CoA)	61
26.	<i>ram20</i> gene in the pMK-RQ plasmid	62
27.	Restriction digestion analysis of pET-KK4 with <i>NdeI/XhoI</i>	63
28.	Analysis of pure Ram20 protein by 12% (v/v) SDS-PAGE	64
29.	A) PCR amplification of 576 bp <i>ssuE</i> and B) Restriction digestion analysis of <i>ssuE</i> cloning	65
30.	Analysis of pure SsuE protein by 20% (v/v) SDS-PAGE	66
31.	SsuE activity assay with FAD and NADPH (A). No enzyme control, (B). Enzyme reaction	67
32.	A) PCR amplification of 702 bp <i>fre</i> and B) Restriction digestion analysis of <i>fre</i> cloning	67
33.	Analysis of pure Fre protein by 20% (v/v) SDS-PAGE	68
34.	A) PCR amplification of 1602 bp <i>ktzQ</i> and B) Restriction digestion analysis of <i>ktzQ</i> cloning	70
35.	A) PCR amplification of 1602 <i>ktzQ</i> and B) Restriction digestion analysis of <i>ktzQ</i> cloning	71
36.	DNA sequencing chromatogram showing the mutation of CGA to CGC in <i>ktzQ</i> . Top- wild type <i>ktzQ</i> sequence with the CGA codon, bottom- <i>ktzQ</i> mutant sequence with the CGC codon. The codons are shown in boxes. Colours- Red-T, blue-C, green-A and black-G	72
37.	Expression studies of KtzQ protein using <i>E. coli</i> Rosetta 2 cells analysed by 12% (v/v) SDS-PAGE	73
38.	Analysis of pure KtzQ protein by 12% (v/v) SDS-PAGE	74
39.	Trypsin cleaved proteins obtained by MASCOT analysis (Red-cleaved proteins)	75
40.	Western blot analysis showing the detection of KtzQ protein with anti His-antibody	76
41.	Western blot analysis showing the detection of PrnA protein bound with His-antibody	78
42.	Substrates that showed chlorination at 7-position by PrnA	79
43.	Crystal structure of PrnA (PDB ID:2AR8) generated using PYMOL showing amino acid residues(green)	81

	interacting with chlorinated tryptophan(orange) in the active site	
44.	Analysis of purified mutants Y444W, N459W, E450W and Y443S by 12% (v/v) SDS-PAGE	82
45.	Analysis of E450R mutant protein by 12% (v/v) SDS-PAGE	84
46.	Crystal structure of PrnA(PDB ID:2AR8) generated using PYMOL showing active site showing F454 (red) and other residues (green) interacting with chlorinated tryptophan	85
47.	Analysis of purified PrnA from <i>E. coli</i> BL21 DE3 cells by 12% (v/v) SDS-PAGE	87
48.	Analysis of PrnA treated with ATP by 12% (v/v) SDS-PAGE	89
49.	Analysis of PrnA treated with an ATP wash by 12% (v/v) SDS-PAGE	89
50.	Analysis by 12% (v/v) SDS-PAGE (A). Purified PrnA following the ATP Wash showing removal of the chaperones (B). Purified PrnA following gel-filtration and concentration	90
51.	Analysis of HRV 3C cleaved PrnA by 12% (v/v) SDS-PAGE for removing chaperones	91
52.	(A). Diffraction image, (B). PrnA F454K mutant crystal in tetragonal shape	93
53.	(A).Overlay of PrnA wild type(PDB ID:2AR8) (blue) with PrnA F454K mutant (green) showing similar structures, (B).Similar binding of FAD and Cl ⁻ by PrnA (blue) and the F454K mutant (green), generated using PYMOL	94
54.	Crystal structure of PrnA wild type(PDB ID:2AR8) (blue) showing the α -helix (magenta) and F454 mutant (green) showing the non-aligning extended loop (red, generated using PYMOL	95
55.	The distance differences of the interaction between PrnA F454K mutant (green) residues Y444 and Y443 and the wild type (PDB ID:2AR8) residues (blue) with tryptophan, the K454 mutation is shown in red and chlorinated tryptophan is orange	96
56.	Docking of anthranilic acid in the F454K mutant (green) structure showing interaction with Y444, Y443 and K454 (red), wild type PrnA (PDB ID:2AR8) (blue) used as template, anthranilic acid (magenta)	96
57.	<i>ktzR</i> constructed in pMK-RQ	98

58.	Restriction digestion analysis of pET-KK29 with <i>NdeI/NotI</i>	99
59.	Analysis of pure KtzQ protein from (A). <i>E. coli</i> BL 21(DE3) cells, (B) <i>E. coli</i> ArcticExpress DE3(RP) by 12% SDS-PAGE	100
60.	Analysis of pure KtzR protein by 12%(v/v) SDS-PAGE	102
61.	<i>SttH</i> constructed in pMK-RQ plasmid	103
62.	Analysis of pure SttH by 12%(v/v) SDS-PAGE	104
63.	Substrates that were chlorinated at 6-position by SttH	105
64.	Sequence alignment of tryptophan 5-, 6- and 7-halogenases showing the conserved phenylalanine residue	106
65.	Analysis of pure SttH F98A by 12% (v/v) SDS-PAGE	106
66.	HPLC chromatogram showing low chlorination activity by SttH F98A mutant	108
67.	Analysis of ATP treated SttH protein fractions by 12%(v/v) SDS-PAGE	110
68.	Analysis of fractions purified from a HiTrap™ Chelating HP column by overloading SttH by 12% (v/v) SDS-PAGE	111
69.	Analysis of purified fractions from the HiTrap™ Q HP anion exchange column by 12% (v/v) SDS-PAGE	111
70.	Analysis of pure SttH protein on 12% resolving gel	112
71.	SttH crystals obtained on crystal trials	113
72.	Analysis of purified SttH fractions by nickel binding method by 12%(v/v) SDS-PAGE	114
73.	Analysis of purified SttH fractions from HiTrap Q HP anion exchange column by 12%(v/v) SDS-PAGE	114
74.	Analysis of purified SttH fractions from the MonoQ HP anion exchange column by 12%(v/v) SDS-PAGE	115
75.	Analysis of purified SttH fractions from the Superdex™ 200 10/300 column by 12%(v/v) SDS-PAGE	116
76.	pMK-RQ:: <i>pyrH</i> synthetic gene construct	117
77.	Restriction digestion analysis of sub-cloned construct pETKK32 with <i>NdeI/XhoI</i>	117
78.	Analysis of purified PyrH by 12%(v/v) SDS-PAGE	118
79.	(A). HPLC chromatogram of PyrH chlorination activity in DMSO, (A1). Relative activity of PyrH chlorination activity in DMSO, (B) HPLC chromatogram of PyrH chlorination activity in EtOH, (A1). Relative activity of PyrH chlorination activity in	120

	EtOH	
80.	Substrate library screened with purified flavin dependent halogenases and mutants	122
81.	HPLC chromatogram and mass spectrum of tryptophan chlorinated by PyrH with a mass of 239.0 and 241.0	123
82.	¹ H NMR data showing chlorination of tryptophan at the 5-position by PyrH with the complete ¹ H NMR spectrum shown at the top and an enlarged aromatic region shown at the bottom of the figure	124
83.	COSY spectra of tryptophan (left) and 5-chlorinated-tryptophan (right)	125
84.	HPLC chromatogram and mass spectrum of kynurenine chlorinated by PyrH showing masses of 243.0 and 245.0	126
85.	¹ H NMR data showing chlorination of kynurenine at the 5-position by PyrH with the complete ¹ H NMR spectrum shown at the top and an enlarged aromatic region shown at the bottom of the figure	128
86.	HPLC chromatogram and mass spectrum of anthranilamide chlorinated by PyrH with a mass of 171.0 and 173.0 with a neutral loss of ammonia	129
87.	¹ H NMR data showing chlorination of anthranilamide at the 5-position by PyrH with the complete ¹ H NMR spectrum shown at the top and an enlarged aromatic region shown at the bottom of the figure	130
88.	HPLC chromatogram and mass spectrum of 2-amino-4-methylbenzamide chlorinated by PyrH with mass of 185.0 and 187.0	131
89.	¹ H NMR data showing chlorination of 2-amino-4-methylbenzamide by PyrH with the complete ¹ H NMR spectrum shown at the top and an enlarged aromatic region shown at the bottom of the figure	132
90.	ROESY NMR data showing chlorination of 2-amino-4-methylbenzamide at the 5-position by PyrH	133
91.	Docking experiment showing the distance interaction of kynurenine in the PyrH(PDB ID:2WEU) active site(red line-tryptophan, blue stick-kynurenine)	135
92.	Docking experiment showing the distance interaction of 2-amino-4-methylbenzamide in the PyrH(PDB ID:2WEU) active site(red line-tryptophan, blue stick-2-amino-4-methylbenzamide)	137
93.	Sequence alignment of tryptophan 7-halogenases with	139

	tryptophan 5-halogenases	
94.	The three important residues selected for mutation studies Red-PrnA (tryptophan 7-halogenase, Yellow-PyrH (tryptophan 5-halogenase)	139
95.	HPLC chromatogram of PrnA mutants with tryptophan for regioselective chlorination	140
96.	HPLC chromatogram of PrnA mutants with tryptophan for regioselective bromination	141
97.	HPLC chromatogram of chlorination activity of the PyrH mutants with tryptophan for regioselectivity	143
98.	HPLC chromatogram of bromination activity of the PyrH mutants with tryptophan for regioselectivity	143
99.	Sequence alignment of tryptophan-5-halogenases with tryptophan-7-halogenases (yellow-loop region)	143
100.	The 8 amino acid loop present in PrnA (Blue) is absent in PyrH, Red-7-chloro-tryptophan, yellow- 5-chloro-tryptophan	145
101.	Analysis of the PrnA loop deleted mutant protein by 12% (v/v) SDS-PAGE	145
102.	Analysis of pure PyrH loop inserted mutant by 12% (v/v) SDS-PAGE	146
103.	Site directed mutagenesis strategy to insert/delete an 8 amino acid loop	163
104.	HPLC method used to separate halogenated compounds	177

List of Schemes

Scheme No	Scheme Title	Page No
1.	Halogenation of arene by halogens	23
2.	Example of Sandmeyer reaction for halogenation	23
3.	Mechanism of Non-ribosomal peptide synthesis showing initial activation of amino acid using AMP followed by loading onto PCP domain, followed by elongation and chain termination by internal nucleophile attack by thioesterase domain(AMP-adenosine monophosphate, aa-aminoacyl, ppi-pyrophosphate, R-amino acid side chains)	28
4.	Reaction mechanism of heme dependent haloperoxidase	33
5.	Reaction mechanism of vanadium dependent haloperoxidase	34
6.	Reaction mechanism of α keto-glutarate dependent haloperoxidase.	36
7.	Mechanism of the two-component system of flavin dependent halogenase showing the formation of hypochlorous acid which reacts with conserved K79 residue forming chloramine as halogenating agent followed by electrophilic aromatic substitution using conserved E346 residue	38
8.	Scheme showing biosynthesis of pyrrolnitrin from tryptophan using enzymes PrnA, PrnB, PrnC and PrnD	39
9.	Conserved Glu151 residue of Sfp causes deprotonation of hydroxyl group of the conserved serine residue of PCP domain followed by nucleophilic attack on the β -phosphate of coenzyme A	52
10.	A) Synthesis of CoA-Hpg using N-Boc protected Hpg and thiophenol and B) syntehsis of SNAC-Hpg using SNAC and N-Boc protected Hpg	59

Abbreviations

A domain	Adenylation domain
ATP	Adenosine-5'-triphosphate
α -KG	α -ketogulatarte
BOC	<i>t</i> -butyloxycarbonyl
bp	base pair
C domain	Condensation domain
COSY	Correlation spectroscopy
Cpn	Chaperone
Da	Dalton
DMSO	Dimethylsulfoxide
EAS	Electrophilic aromatic substitution
ECL	enhanced chemiluminescence
<i>E. coli</i>	<i>Escherichia coli</i>
EtOH	Ethanol
FACS	Fluorescens activated cell sorting
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
FPLC	Fast Protein Liquid Chromatography
HPLC	High Pressure Liquid Chromatography
IPTG	Isopropyl β -D-1-galactopyranoside
kDa	Kilo dalton
LC-MS	Liquid Chromatography-Mass Spectrometry
L-PCP	Large Peptidyl Carrier Protein
M9 medium	Minimal medium
MALDI	Matrix-assisted laser desorption/ionization
MALS	Mutiangle light scattering
MRSA	Methicillin resistant <i>Staphylococcus aureus</i>
MOR	Monooxygenase reaction
MS	Mass Spectrometry
NADH	Nicotinamide adenine dinculeotide
NMR	Nuclear magnetic resonance
NRP	Non-Ribosomal Peptide
NRPS	Non-Ribosomal Peptide Synthetase
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCP domain	Peptidyl Carrier protein domain
PCR	Polymerase chain reaction
PDB	Protein Data Bank
Ppant	4'-phosphopantetheine
PrnA LD	PrnA loop deletion
PPTase	Phosphopantetheinyl transferase
PyrH LI	PyrH loop insertion

RA	Relative activity
RaCC	Rare Codon Calculator
ROESY	Rotating-frame Overhauser Effect Spectroscopy
SDS	Sodium dodecylsulfate
SNAC	N-acetylcystamine thioester
S-PCP	Small Peptidyl Carrier protein
SttH LI	SttH loop insert
TE domain	Thioesterase domain
UV	Ultraviolet
VRE	vancomycin-resistant Enterococci
WT	Wild type

Amino acids

Amino acids	Three letters	Single letter
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Non proteinogenic amino acids

Orn	Ornithine
Hpg	Hydroxyphenylglycine
End	Enduracidin
Chp	Chlorinated hydroxyphenyl glycine

Abstract

The University of Manchester
Chinnan Velmurugan Karthikeyan
Halogenases for biosynthesis and biocatalysis
September 2013

Halogenation is an important chemical and biological process in the production of industrially important halogenated compounds. The presence of the halogen enhances the desirable properties of many of these compounds. However, production of halogenated compounds chemically often generates by-products that cause environmental pollution and the reactions often lack regio/stereo-selectivity. Biocatalysis have emerged to overcome these issues providing us with a novel technique to produce much greener regio-selective halogenated compounds.

The first project involved the study of putative halogenase Ram20 from the ramoplanin biosynthetic gene cluster which is predicted to chlorinate the non-proteinogenic amino acid Hydroxyphenylglycine (Hpg) found at position 17 of ramoplanin. We tried to develop an *in vitro* system which would allow us to investigate Ram20's ability to chlorinate Hpg-17. To achieve this we expressed the peptidyl carrier protein (PCP) domain of the ramoplanin non-ribosomal synthetases module to which Hpg-17 is tethered, a phosphopantethenyl transferase Sfp required to load the HPG onto the PCP domain and Ram20. Synthetic compounds CoA-Hpg and CoA-SNAC were synthesised separately. Unfortunately, no evidence for chlorination of Hpg could be obtained.

The second project involved the study of tryptophan halogenases, with the focus on the flavin dependent halogenases that regio-selectively chlorinate small substrates. The flavin dependent halogenases PrnA, KtzQ, SttH, KtzR and PyrH were produced in *E. coli* for the first time. The enzymes were assayed with tryptophan and a wider range of aromatic and phenolic substrates. KtzQ failed to show any promising activity, while KtzR could only chlorinate tryptophan. The other halogenases, PrnA, SttH and PyrH, were able to chlorinate a wider range of substrates including tryptophan, kynurenine, anthranilamide, 2-amino-4-methylbenzamide and anthranilic acid. Mutagenesis was performed on PrnA to widen the substrate scope, wherein a mutant F454K was identified that had higher halogenation activity with anthranilic acid compared to that of the wild type. The crystal structure of this mutant bound with FAD and Cl⁻ was obtained. Work was also carried out to identify important residues that govern regio-selective halogenation of these enzymes. In particular, attempts were made to obtain a crystal structure of SttH, a tryptophan-6-halogenase to study the mechanism of regio-selective halogenation at the 6-position.

Declaration

I declare that no portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

- i. The author of this thesis (including any appendices and/or schedules to this thesis) owns any copyright in it (the “Copyright”) and he has given The University of Manchester the right to use such Copyright for any administrative, promotional, educational and/or teaching purposes.
- ii. Copies of this thesis, either in full or in extracts, may be made online accordance with the regulations of the John Rylands University Library of Manchester. Details of these regulations may be obtained from the Librarian. This page must form part of any such copies made.
- iii. The ownership of any patents, designs, trademarks and any and all other intellectual property rights except for the Copyright (the “Intellectual Property Rights”) and any reproductions of copyright works, for example graphs and tables (“Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property Rights and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property Rights and/or Reproductions.
- iv. Further information on the conditions under which disclosure, publication and exploitation of this thesis, the Copyright and any Intellectual Property Rights and/or Reproductions described in it may take place is available from the Head of School of Chemistry.

ACKNOWLEDGEMENTS

I wish to deeply thank my supervisor **Prof. Jason Micklefield**, for providing me this opportunity to work in his lab, his many ideas, valuable suggestions, constructive criticism, wholehearted help and constant encouragement for the full duration of this study.

I would like to thank Dr. Ming-Cheng Wu for helping me during the start of the PhD and throughout these years with a constant support. I would thank Dr. Matthew Wilding for helping me in all the regards. I would like to thank Dr. Neil Dixon for his kind advice whenever needed. I would like to thank Dr. John Duncan, Dr. Mitchell Jacobs, Dr. Christopher Chesters, Dr. Mark Goodall, Dr. Laura Nunns and Michael Power for making me so much comfortable in lab during the start of my PhD.

My gratitude is due to Sarah Shepherd, Dr. Mark Thompson, Dr. Matthew Styles, Dr. Anna-winona struck and Jonathan Latham who collaborated in this work and helped me obtaining a good knowledge in the field. My sincere thanks to Dr. Helen Vincent, Dr. Mark Thomson, Dr. Jayendra Shankar, Sarah Shepherd, Dr. Anna-winona Struck, Dr. Binuraj Menon and Dr. Venugopal Bhat for a very critical proof reading of my theses. I specially thank lab members Dr. Christopher Robinson, Dr. Lu Shin Wong, Matthew Bennett, Brian Law, Seyed Yasin Tabatabei Dakhili, James Leigh, Ross Lewin, Phillip Lowe, Angelica Jimenez Rosales, Dr. Stewart Carnally, Kate Leigh and Joseph Hosford for their heartly help during the period of research. I would like to thank all past and present lab members for making this four years a memorable one. I would also like to thank Dr. Binto Simon and his family, all my friends for wonderful care.

I sincerely thank Dr. Colin Levy in helping us to obtain crystal structure and Dr. Edward McKenzie for providing guidance on chaperone removal. I personally thank the store members for their help rendered throughout the years.

The love, care, preservance, sacrifices and unstinted support of my parents and my brother are what I am today. What little I do I dedicate it to them. Finally I thank Government of India for providing me the funding to successfully complete my PhD programme.

Chinnan Velmurugan Karthikeyan,
Manchester,
September 2013.

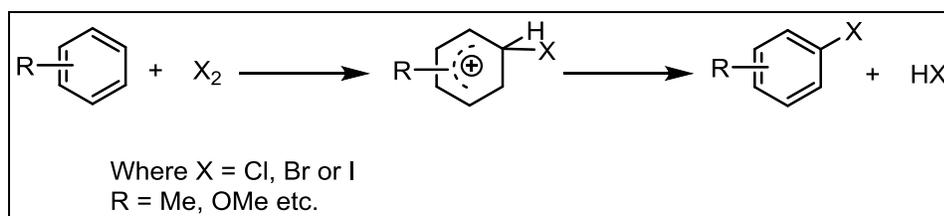
The Author

The author completed his Master degree in Biotechnology from TamilNadu Agricultural University, India in July 2005. After working as a Junior Research Fellow for Prof. Gopaldaswamy, the author undertook a PhD under the supervision of Prof. Jason Micklefield from September 2009 to September 2013.

Chapter 1: Introduction

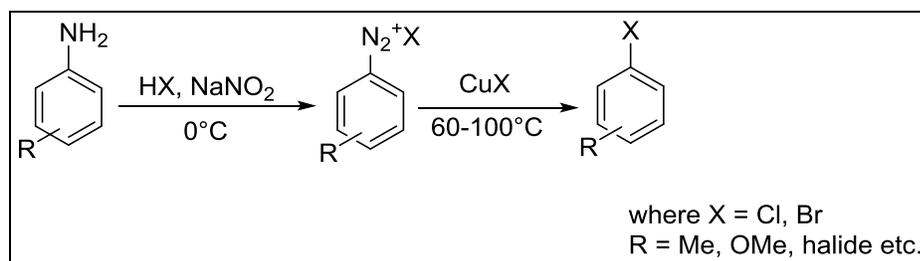
1.1. Halogenation

Halogenated compounds are useful intermediates in synthesis of many important chemical compounds used in industry². They serve as precursors for many organometallic reagents (e.g. Grignard reagents) which have great synthetic utility. Many C-C bond forming reactions using transition metal catalysed chemistry also rely on halogenated compounds². The traditional method of producing halogenated aromatic compounds relies on an electrophilic aromatic substitution reaction promoted by Lewis acids (Scheme 1).



Scheme 1. Halogenation of arenes by halogens

These methods use molecular halogens and therefore large-scale processes based on these reactions pose considerable operational challenges as well as environmental hazards. Other methods of generating aromatic halogenated compounds include the Sandmeyer reaction (Scheme 2) which involves the preparation of a diazonium salt from an amine and its subsequent replacement with a halogen nucleophile. This reaction usually proceeds in the presence of copper as a catalyst.



Scheme 2. Example of Sandmeyer reaction for halogenation

Many halogenated compounds exist in nature that are bioactive with the presence of a halogen in the molecule having considerable influence on its activity. 3,5-diiodotyrosine from the marine organism *Gorgonia cavollinii* was the first identified halometabolite and Diploicin from the lichen *Buellia canescens* was the first isolated halogenated

natural product^{3,4}(Figure 1). Since then, over 5000 halogenated natural products have been isolated.

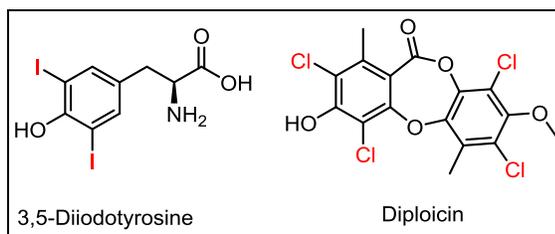


Figure 1. First identified and isolated halometabolites

Chloramphenicol and vancomycin are examples of halogenated antibiotics currently in use where the presence of halogens increases their antibacterial properties⁵⁻⁷. Examinations through crystal structure have provided evidence that the presence of halogen provides selective specificity and substrate binding that occurs with electron rich aminoacids sidechain lewis bases (O,N and S)⁸.

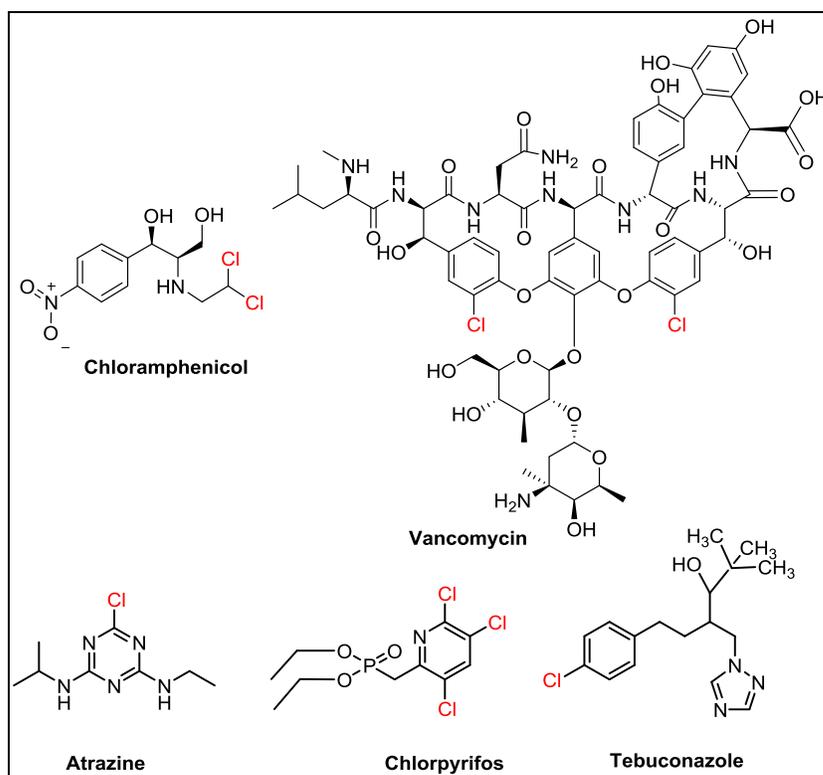


Figure 2. Examples of halogenated compounds used in the pharmaceutical and agrochemical industries

Halogenated compounds are also used in the agrochemical industry where they act as herbicides, pesticides and insecticides. For example, atrazine acts as a herbicide, chlorpyrifos acts as an insecticide and tebuconazole acts as a fungicide⁹ (Figure 2).

Though many novel chemical processes have been developed over the last few years to produce these halogenated compounds, problems persist in scaling up and obtaining regioselective halogenated compounds. The production of halogenated compounds often results in by-products which are tedious to dispose off and some of these by-products cause environmental pollution. However the biggest disadvantage of chemical halogenation is that it often lacks regioselectivity. To overcome the above limitations, biocatalysis has been used as an alternative "environmentally safe" process to produce halogenated compounds.

1.2. Biocatalysis: a general overview

Biocatalysis is a technique, which involves the use of enzymes or whole cells as a catalyst in synthetic chemistry. Enzymes occurring naturally have advantage of much higher kinetic rate¹⁰. Furthermore, using enzyme-based biocatalysts can result in high regioselectivity and or chemoselectivity which is often required for making the desired compounds¹¹. This feature makes them attractive alternative tool as catalyst in the synthesis of complex and high value molecules, especially where chemical routes are difficult to implement^{12,13}. Biocatalysis is extremely significant for the pharmaceutical, chemical and agricultural industries. This is due to the need to move towards green chemistry, using biodegradable materials and mild reaction conditions combined with the need to reduce the cost and manpower involved. Biocatalytic processes generally have a low energy requirement with minimal waste generation.

The history of biocatalysis began centuries ago, when living cells were used for the production of wine and beer. Later, enzymes were used as one of the components for producing compounds¹⁴. The biggest challenge during this period was the poor stability of the enzymes under desired reaction conditions, which still remains as a major concern in most of the laboratories and industries today. The real development of biocatalysis

began with the advancement in DNA sequencing and gene synthesis technologies. Initially only wild-type enzymes that could be isolated from the host organism could be used, but the start of recombinant DNA technology enabled the cloning and expression of almost any enzyme of interest using model organisms. Structure based protein engineering can also be carried out for widening the substrate range allowing the production of a wide variety of synthetic intermediates. In recent times, directed evolution has contributed to the field immensely. This technique involves random amino acid mutations in the protein of interest to form libraries, which can be screened by high throughput methods for mutants with improved activity. Fluorescence activated cell sorting (FACS) or microfluidic devices are common high throughput screening techniques which are currently being used¹⁵⁻¹⁷.

1.3. Biocatalyst engineering and the host system

Biocatalysts perform very efficient regio/stereoselective catalysis to produce compounds. They usually require mild reaction conditions with a minimum energy requirement and produce a very low amount of by-products. The biggest advantage is that they can be prepared on a large scale by fermentation. However, they also have disadvantages. Stability of the proteins has always been the biggest concern because they are inactivated by the high temperatures, extremes of pH and salt that often need to be used in industrial processes. In addition, they are often affected by substrate or product inhibition.

Biocatalyst/protein engineering involves altering the protein by site directed or random mutagenesis to achieve improved function. Protein engineering can provide better substrate scope and activity, higher catalytic rates, decreased product inhibition, desired cofactor use, and reduced substrate competition¹⁸. In natural system, engineering of a single protein is not achievable, hence alternative model expression system are often used. Available expression systems include mammalian cells lines, insect cell lines, yeasts, fungi and bacteria. Among the many systems available, the gram negative bacterium *Escherichia coli* is often the system of choice, due to its rapid growth, well

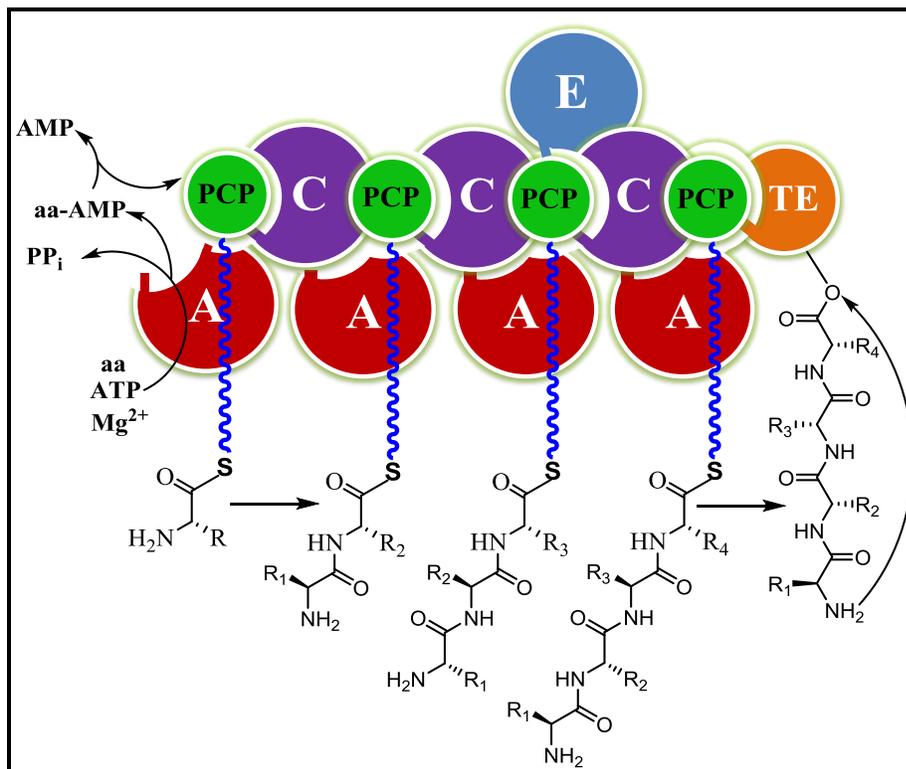
characterised genetics and the availability of a large number of cloning vectors. However, it also has disadvantages. For example, not every foreign gene can be efficiently expressed in *E. coli* due to the low stability of mRNA and slow translational efficiency, the different codon usage between the originating organism of the foreign gene, protein misfolding, the degradation of foreign proteins by *E. coli* proteases, and the toxicity of the expressed protein to *E. coli*. Through various novel techniques, engineered strains of *E. coli* have been developed for the efficient, production of a diverse range of heterologous proteins¹⁹. *Bacillus* and *Pseudomonas* are also used as bacterial host systems to produce proteins. *Pseudomonas fluorescens* and *Pseudomonas putida* are the two major strains that are used as hosts for protein production²⁰. These host strains are amenable to genetic or molecular manipulations and can be cultivated at high cell densities. Their high tolerance to toxic and harsh conditions makes them extremely useful strains for industrial scale up. However, their biggest disadvantage lies in the limited availability of suitable plasmids for producing recombinant proteins. *Pseudomonas* uses a broad range vectors of size 9-10 kb with *tac* and *lacUV5* promoters derived from *E. coli*. The larger size of these plasmids makes them difficult to manipulate for cloning and transformation. Conjugation can be used to transfer the construct into the host strain but this requires additional helper strains. Hence *E. coli* has been very much the preferred strain for use in industry for the production of active biocatalysts.

Production of novel halogenated compounds using halogenating enzymes has been a recent development in the last few years. The following sections will focus on biohalogenases, together with their advantages and drawbacks.

1.4. Biohalogenases

Biohalogenases can be broadly classified into two major groups: halogenases that can accept substrates tethered to the peptidyl carrier protein (PCP) domain of non-ribosomal peptide synthetases (NRPSs) and ones that accept free substrates.

NRPSs are mega-enzymes that are used to produce non-ribosomal peptides (NRPs), linear or cyclic peptides that are synthesised via a nucleic acid independent non-ribosomal mechanism (Scheme 3). The mega-enzymes are arranged as a series of catalytic domains, each responsible for the addition of one amino acid to the peptide. The catalytic domain consists of a peptidyl carrier protein (PCP) or thiolation domain which tethers the peptide to the protein, an adenylation (A) domain which recognises the next specific amino acid to be added and activates it using ATP and Mg^{2+} and the condensation (C) domain which forms the peptide bond between amino acids. The whole module is called the elongation module and thus forms the chain C-A-PCP. A fourth essential domain is a thioesterase (TE) domain, which is involved in the product release by either internal nucleophilic attack or hydrolysis²¹ (Scheme 3).



Scheme 3. Mechanism of Non-ribosomal peptide synthesis showing initial activation of amino acid using AMP followed by loading onto PCP domain, followed by elongation and chain termination by internal nucleophile attack by thioesterase domain (AMP-adenosine monophosphate, aa-aminoacyl, ppi-pyrophosphate, R-amino acid side chains) adapted using chemdraw from ref²²

The advantage of the NRPS system is that non-proteinogenic amino acids can be incorporated into the peptides. Non-proteinogenic amino acids are often key intermediates in the biosynthesis of natural products that are pharmaceutically important. For example, the non-proteinogenic amino acid hydroxyphenylglycine (Hpg) is one of the non-proteinogenic amino acids found in several peptidic natural products including the vancomycin group of antibiotics (e.g. complestatin²³, chloroeremomycin²⁴, vancomycin²⁵ and, ramoplanin) as well as in calcium-dependent antibiotics and plays a structural role in formation of the rigid conformation of the compounds. Hpg is also a precursor in the production of amoxicillin, a commercially successful antibiotic²⁶. It occurs in both the L- and D-isomers but the L-isomer remains the natural substrate of the NRPS.

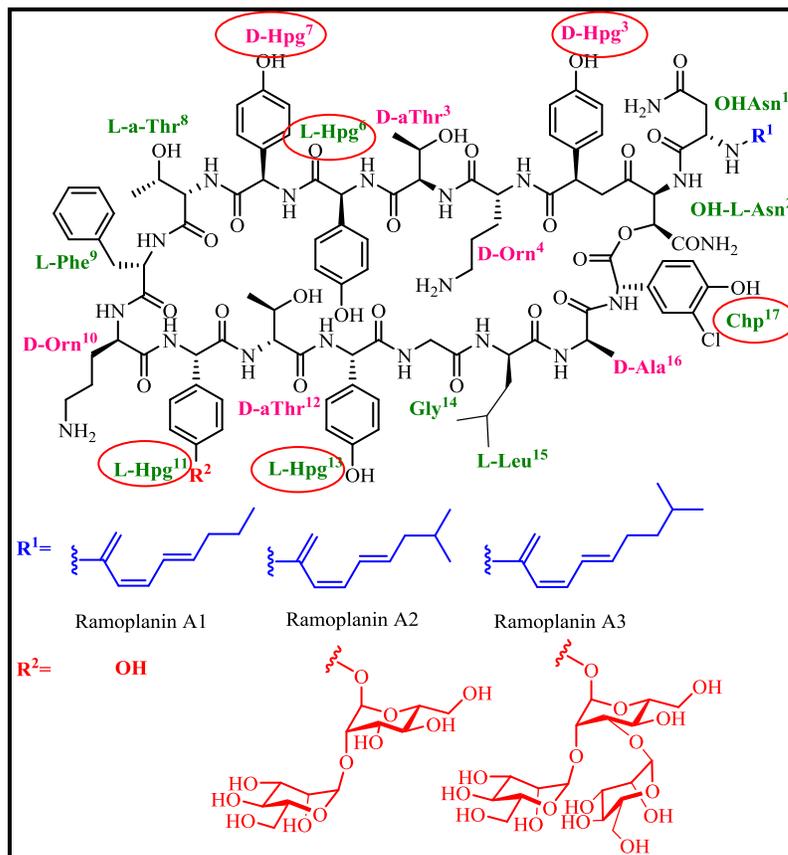


Figure 3. Structures of ramoplaninA1, A2 and A3 (Orn- ornithine, Chp-chlorinated hydroxyphenyl glycine, Hpg are highlighted in circle and mannose group in red)adapted using chemdraw from ref²⁷

The NRP ramoplanin (Figure 3) is a non-ribosomally synthesised peptide first isolated from *Actinoplanes* spp. ATCC 33076^{28,29}. It consists of 17 non-proteinogenic amino acids including ornithine (Orn), hydroxy asparagine (OH Asn), allo-threonine (a-Thr) and abundance of Hpg residues and a chlorinated Hpg (Chp). Chp17 is the last amino acid to be synthesised in the system which undergoes chain termination reaction forming cyclic peptide with OH Asn. Ramoplanin exists in three forms A1, A2 and A3 which differ in the length of the N-acyl Asn chain and sugars at R₂³⁰ (Figure 3). Ramoplanin is an antibiotic highly effective against methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant Enterococci and is currently in phase III clinical trials for the oral treatment of intestinal vancomycin-resistant *Enterococcus faecium* and in phase II clinical trials for the treatment of nasal MRSA. Ramoplanin inhibits bacterial cell-wall biosynthesis; blocking the peptidoglycan synthesis by binding to lipid I and preventing formation of lipid II³¹ (Figure 4).

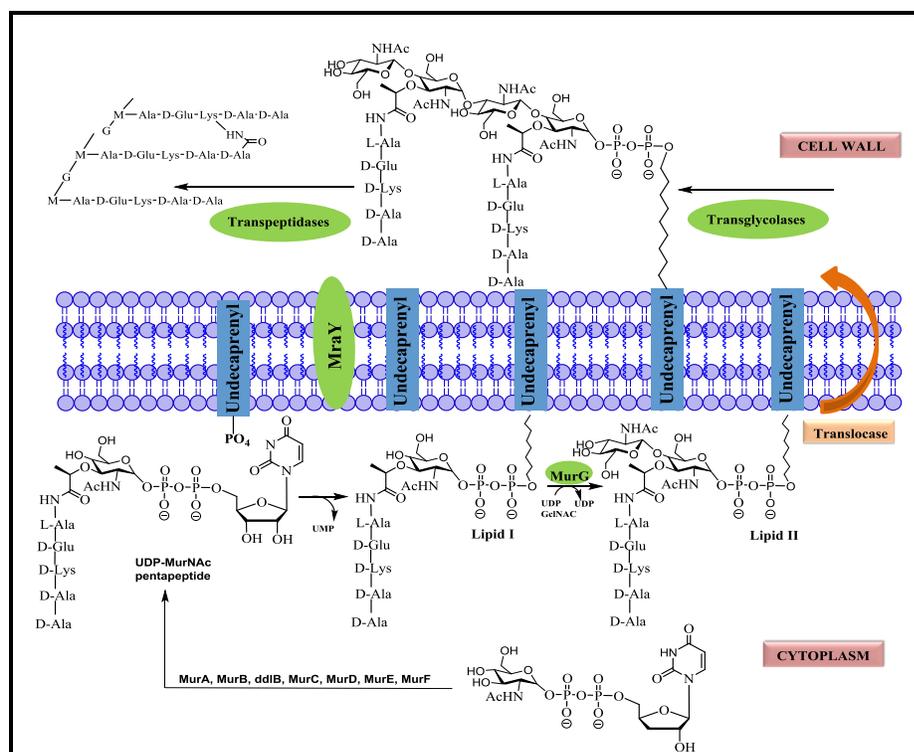


Figure 4. Peptidoglycan biosynthesis showing initial formation of pentapeptide followed by formation of lipid I, lipid II and transglycosylation and transpeptidation step (adapted using chemdraw from ref³¹)

The X-ray crystal structure of ramoplanin has been solved at a resolution of 1.4 Å³². The ramoplanin structure is a dimer with each monomer bent into a U-shape structure. The monomers are joined together by eight carbonyl-amide hydrogen bonds between two extended strands. All of the amino acid side chains are positioned on the exterior of the U-Shaped structure except Hpg-3, Phe-9 and Chp-17, which remain inside the structure (Figure 5). Additionally, the structure is stabilised by hydrogen bonding between the phenolic hydroxyls of the Chp-17 and, Hpg-3 and the carbonyl oxygen's of Hpg-11 and, Phe-9 respectively. The structure shows that the N-acyl chain aids in anchoring the molecule to the membrane and is essential for antibiotic activity. A mutagenesis study of ramoplanin residues confirmed that residues Orn-10, Hpg-3, Hpg-7, and Orn-4 play critical roles in antimicrobial activity.

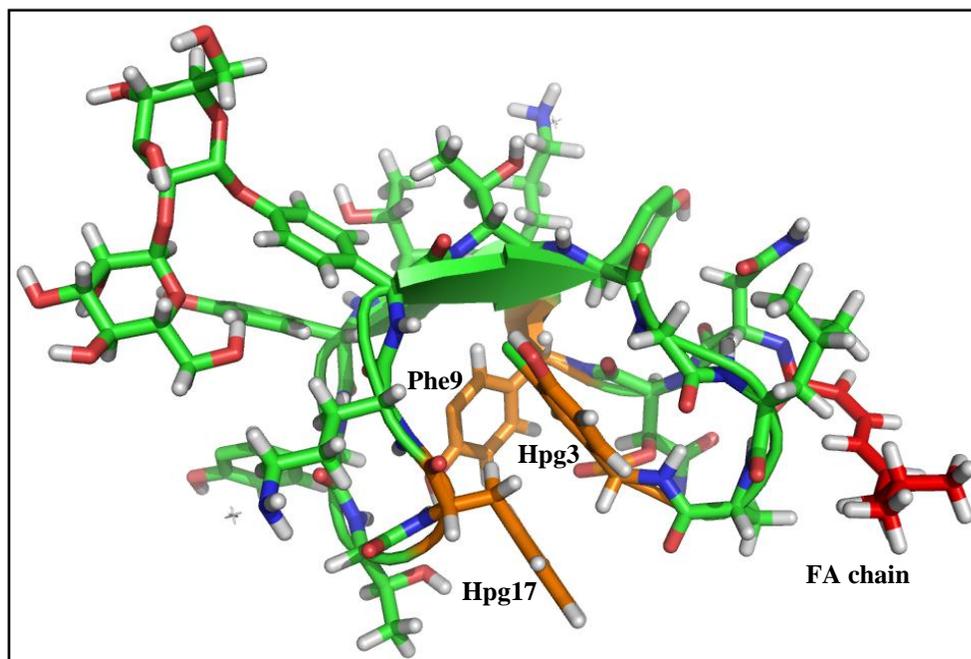


Figure 5. Structure of U shaped ramoplanin (PDB ID:1DSR) showing residues inside the disc (orange), the fatty acyl chain (red) and other residues lying exterior (green) adapted using PYMOL¹ from ref³²

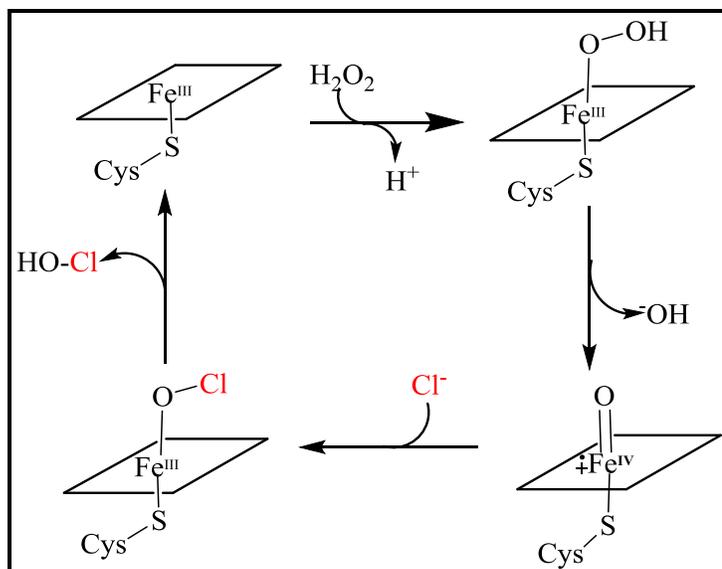
Hpg-17 is the only amino acid found to be chlorinated in the structure of ramoplanin. It is predicted that Orf20/Ram20 could be the halogenase that chlorinates Hpg-17³⁰. An effort was carried out to characterise Ram20 by a previous member of our lab Dr. Ming-

Cheng Wu. Due to the slow growth rate of the ramoplanin producer strain, Ram20 was over-expressed in the faster growing enduracidin producer *Streptomyces fungicidicus*. Enduracidin has a similar structure and, it was expected that Ram20 would chlorinate the Hpg-17 in enduracidin yielding a new enduracidin analogue³³. The result obtained showed negligible amounts of chlorination of the enduracidin. Yin et al., (2010) also over-expressed Ram20 in a knockout enduracidin producing strain Sf Δ end30 and wild-type *Str. fungicidicus* respectively and could observe monochlorinated residues Hpg-13 and Hpg-11 in the mutant, producing monochlorinated and tri-chlorinated products³⁴. However, Ram20 was unable to chlorinate Hpg-17 despite this residue being the natural substrate in ramoplanin. It remains unclear whether Ram20 can chlorinate the Hpg17 residue of ramoplanin or enduracidin.

1.5. Halogenases for biocatalysis

The next group of biohalogenases that we are interested in are the enzymes that accept free/smaller substrates. These enzymes can be further classified as metal dependent and non-metal dependent enzymes.

The first metal dependent halogenases to be discovered were the heme dependent halogenases. This type of enzyme was first detected in the fungus *Caldariomyces fumago* that produces the antibiotic caldariomycin. The enzyme was also found to chlorinate other substrates such as dimedone and tyrosine³⁵. They perform similar reaction to that of cytochrome P450 inserting oxygen into a C-H σ -bonds and C-C π -bonds³⁶. This group of enzymes use H₂O₂ and heme to produce halogenated compounds³⁷. The proposed mechanism involves a Fe^{III} –porphyrin complex which binds H₂O₂. An intermediate Fe^{IV}-oxo species is then generated which subsequently picks up a halide ion to form a Fe^{III} –hypohalite complex. Protonation of this intermediate releases hypophalous acid which halogenates the substrate by an electrophilic substitution mechanism (Scheme 4).



Scheme 4. Reaction mechanism of heme dependent haloperoxidase (adapted using chemdraw from ref³⁸)

The X-ray crystal structure of the heme dependent halogenase from *Cal. fumago* was solved at 1.8 Å with 1,3-cyclopentanedione bound to the active site (PDB ID: 2CIX)³⁹. The crystal structure identified important residues responsible for halogenation. The Fe^{III}-protoporphyrin is bound by C29 in the active site and the hydrogen peroxide interacts with E183 before binding to the heme.

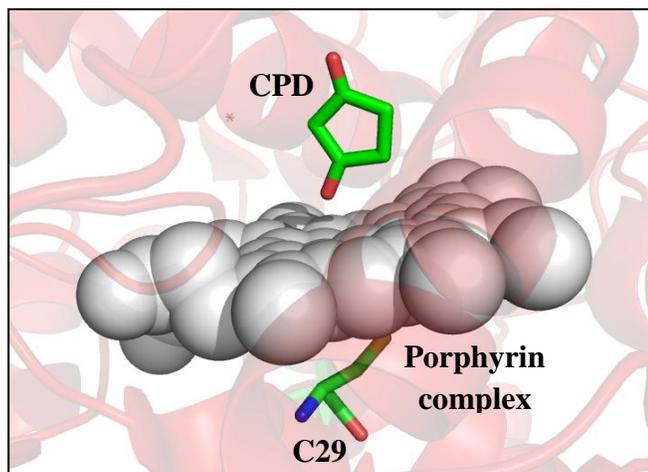
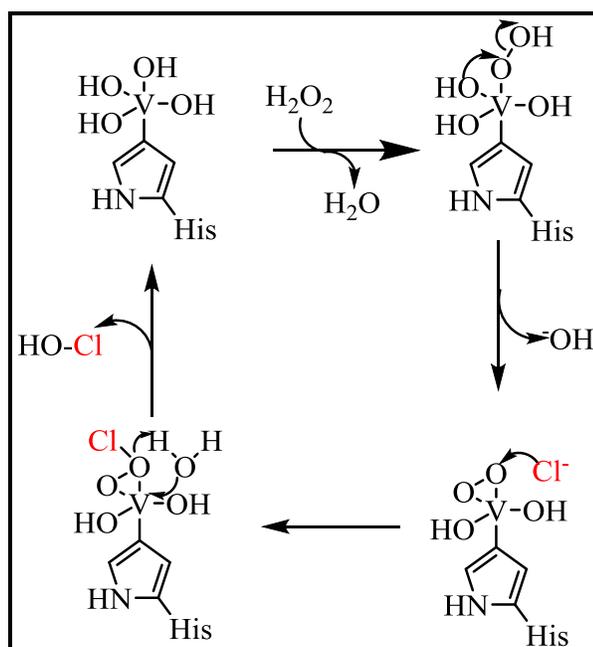


Figure 6. Structure of the active site of heme chloroperoxidase from *Caldariomyces fumago* showing the heme complex bound at conserved C29 residue allowing 1,3-cyclopentanedione (CPD) to bind the complex (PDB ID: 2CIX) generated using PYMOL software¹.

Another group of metal based enzymes use vanadium instead of heme and are termed vanadium dependent haloperoxidases. The first vanadium dependent halogenase was detected in the marine algae *Ascophyllum nodosum* as an iodoperoxidase⁴⁰. Vanadium dependent bromoperoxidase (VBO) are widely distributed in marine sea weeds while the vanadium dependent chloroperoxidases (VCO) have been found in fungi and bacterial species^{41,42}. In vanadium dependent haloperoxidases, the vanadium metal centre performs a similar function to that of the heme-porphyrin in heme dependent haloperoxidase as shown in Scheme 5.



Scheme 5. Reaction mechanism of vanadium dependent chloroperoxidase adapted using chemdraw from ref³⁸.

Overall structural studies of VBO and VCO show a large differences in the overall structure but the active site remains much similar⁴³⁻⁴⁵. The crystal structure of a VBO from *Ascophyllum nodosum* (PDB ID:1QI9) and VCO from *Curvularia inequalis* has been solved (PDB ID: 1IDU)⁴⁶ (Figure 7). The vanadium binding site are much similar and share common conserved motifs. The vanadium at the active site interacts with the side chain of a conserved histidine residue H496, the side chains of two conserved arginine residues R360 and R490 and the backbones of K353, S402 and G403.

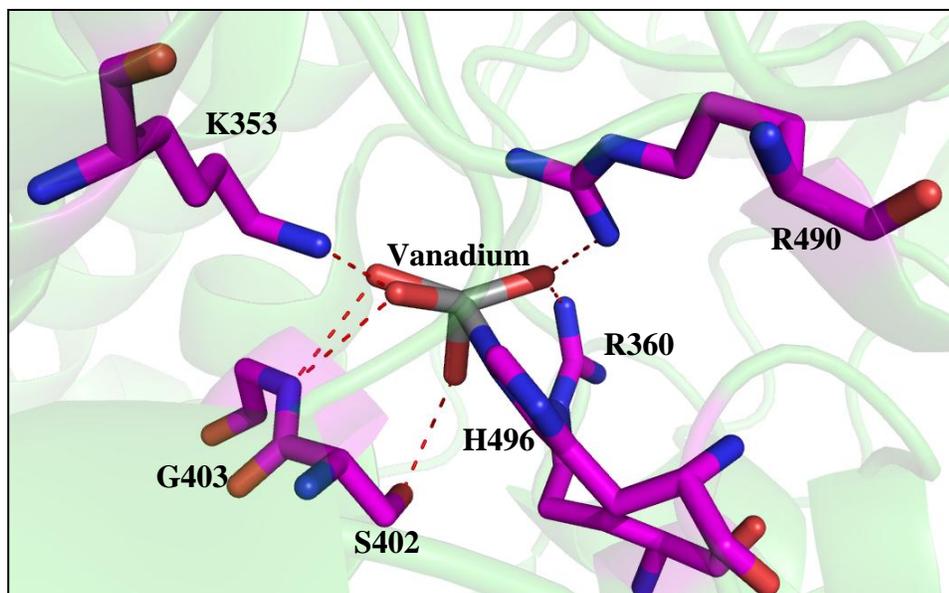
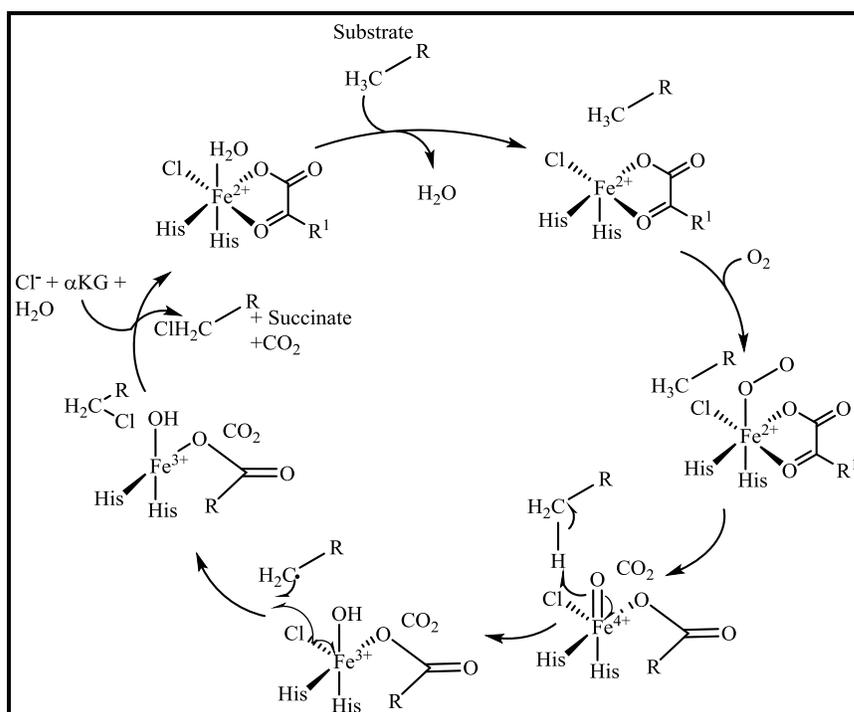


Figure 7. Structure of the active site of a vanadium chloroperoxidase from *Cur. ineaualis* (PDB ID: IIDU) showing vanadium (silver) bound to the conserved H496 residue and interaction with R360, R490, S402, G403 and K353 creating overall positive charge in the site. Generated using PYMOL software¹.

The initial studies on regio/stereoselectivity of V-BPO on oxidation of anisole and other prochiral aromatic compounds failed to show any promise indicating oxidised bromine to be diffusible in the system⁴⁷⁻⁴⁹. However kinetics studies with some of indoles showed that the reaction were not consistent with halogenation by diffusible bromine. This implied that the generated oxidative bromine remains in the active site along with bound substrate, indicating to carry out selective regio/stereoselective bromination⁵⁰. A few examples include the production of diastereospecific bromohydrin formation from *Corallina officinalis*, diastereomers of 2-bromo-1- phenylbutane-1,3-diol from *C. fumago* and regiospecific brominative oxidation to produce 1,3-di-tert-butyl-2-indolinone from *Ascophyllum nodosum* and *C. officinalis*^{51,52}.

The other type of halogenases that use metal are the non heme dependent metal halogenases. These type of halogenases are found associated with the NRPSs systems, wherein the substrate being chlorinated is tethered to the PCP domain of the NRPS. The first type of enzyme was found from cyanobacterium *Lyngbya majuscla* producing

barabamide⁵³. Two enzymes Bar1 and Bar2 are involved in chlorination of methyl group of L-leucine to produce trichloromethyl moiety of barabamide^{54,55}. Other examples of this group includes SyrB2 that chlorinates L-threonine⁵⁶, CmaB chlorinates γ methyl group of L-isoleucine in biosynthesis of cyclopropane⁵⁷ and CytC3 that adds two chlorine to the methyl group of aminobutyric acid in biosynthesis of aremtomycin⁵⁸. The mechanism is a radical mechanism that allows chlorination of unactivated carbon centres. The exact mechanism was obtained after achieving the crystal structure of SyrB2. These enzymes utilise a molecule of α -ketoglutarate, a His-x-Asp/Glu-x-His motif and H₂O octahedrally coordinated Fe^{II}. The binding of substrate displaces the water molecule allowing a molecular oxygen to bind and form an Fe^{III} peroxy compound. This intermediate undergoes decarboxylation to form an Fe^{IV} peroxy compound. A proton is then abstracted from the carbon centre allowing a methyl radical to form. The methyl radical combines with a chloride ion to produce the chlorinated compound. Succinate and CO₂ are released from the enzyme's active site completing the catalytic cycle⁵⁹.



Scheme 6. Reaction mechanism of α -keto-glutarate dependent haloperoxidases (adapted from ref⁶⁰)

The crystal structure of SyrB2 from *Pseudomonas syringae*, an α -ketoglutarate dependent halogenase has been solved (PDB ID:2FCT)⁶¹. The α -ketoglutarate-Fe complex is bound to the enzyme at the active site by the two conserved histidine residues H116 and H235 and an alanine residue A118; the Asp/Glu residues of the His-x-Asp/Glu-x-His motif being replaced by an Ala in SyrB2 where the halide binds (Figure 8). The carboxylate of the α -ketoglutarate substrate hydrogen bonds to the T113, R248 and W145 residues of the protein. These residues are conserved in all Fe α -ketoglutarate family enzymes.

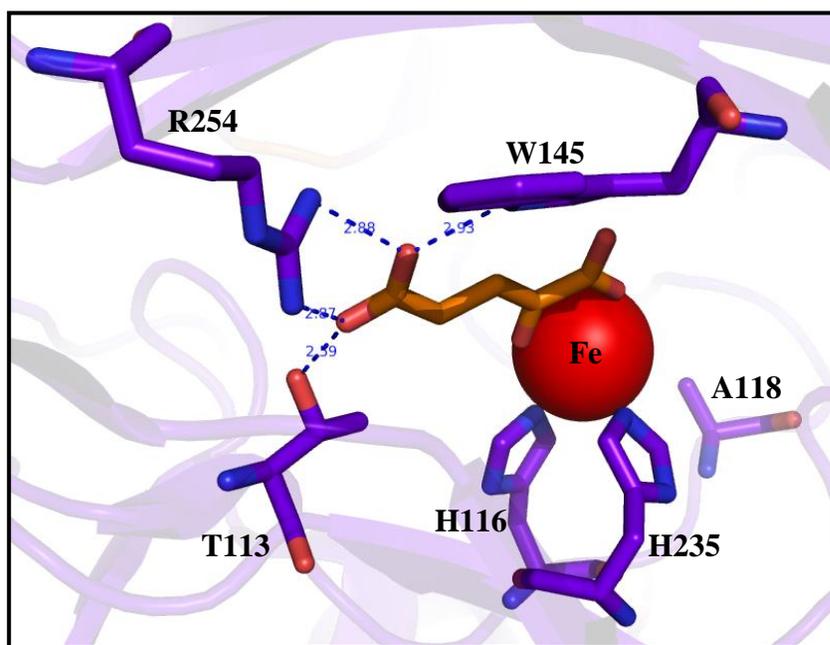


Figure 8. Structure of the active site of SyrB2 from *P. syringae* (PDB ID:2FCT) generated using PYMOL software¹ showing Fe- α -ketoglutarate complex bound to the conserved H116, A118, H235 (Orange- α -ketoglutarate, red sphere- Fe and purple-amino acid residues)

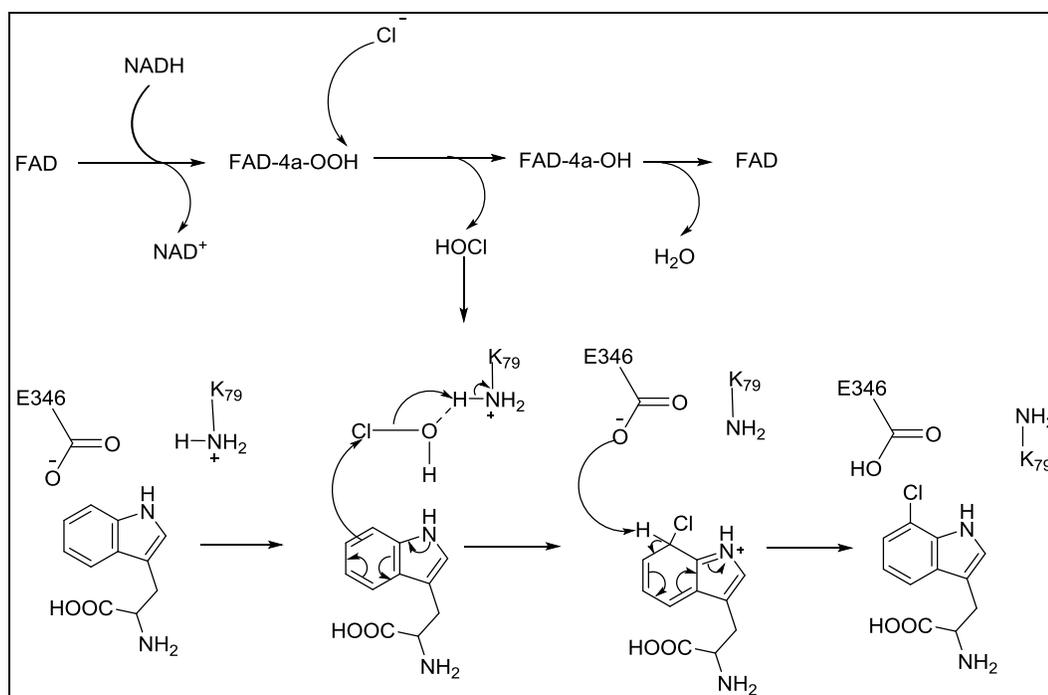
1.6. Flavin dependent halogenases

Flavin dependent halogenases are examples of metal-independent halogenases that do not require any metals for halogenating activity. The study of flavin dependent halogenases began when halogenase enzymes were discovered to be involved in pyrrolnitrin and chlorotetracyclin biosynthesis^{62,63}. These enzymes halogenate large aromatic compounds and aliphatic carbon centres. Most of the halogenated products are

chlorinated or brominated. Iodination is usually not possible for these enzymes due to the larger size of the iodine which will result in incorrect binding. Fluorination is also very difficult due to the enzyme not having enough redox potential to oxidise the electronegative fluoride. Flavin dependent halogenases require a reduced flavin and a molecular oxygen to carry out halogenation of the substrate.

1.6.1. Mechanism of flavin dependent halogenases

The mechanism of flavin dependent halogenases (Scheme 7) was elucidated in 2005 when Dong *et al.* obtained the crystal structure of a tryptophan 7-halogenase PrnA with tryptophan, Cl^- and FAD bound⁶⁴.



Scheme 7. Mechanism of the two-component system of flavin dependent halogenase showing the formation of hypochlorous acid which reacts with conserved K79 residue forming chloramine as halogenating agent followed by electrophilic aromatic substitution using conserved E346 residue (adapted using chemdraw from ref⁶⁵)

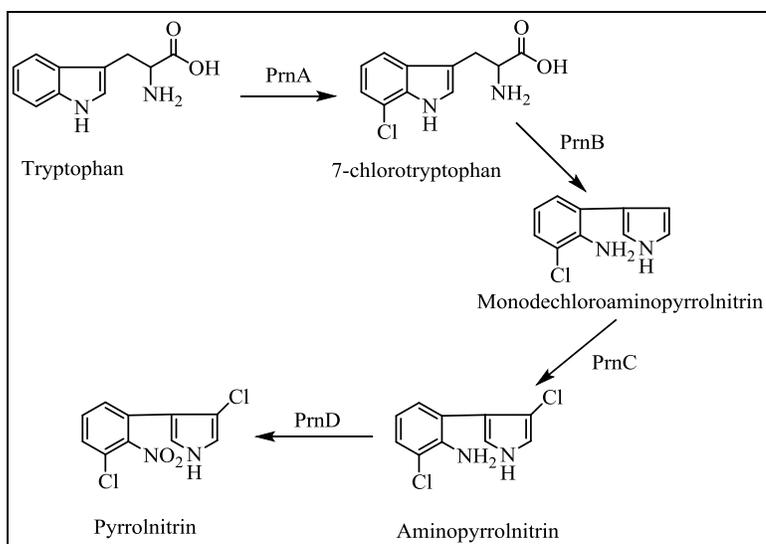
The mechanism first involves the reduction of FAD to FADH_2 by a flavin reductase with the aid of NADH. Examples of flavin reductases that form part of two-component systems in biosynthetic gene clusters include PrnF from *P. fluorescens* PF-5⁶⁶, RebH

from *Lechevalieria aerocolongenes*⁶⁷ and KtzS from *Kutzneria spp.*⁷⁴⁴⁶⁸. Alternatively, FAD reductase Fre and FMN reductase SsuE, both from *E. coli*, can also be used as a source of flavin reductase enzyme^{69,70}. The halogenase bound to the FADH₂ reacts with molecular oxygen to form a hydroperoxide linked to the isoalloxazine ring of FAD. The chlorine bound to FAD reacts with hydroperoxide to form hypochlorous acid. A lysine residue K79 reacts with the hypochlorous acid to form a chloramine as a halogenating intermediate. The glutamate residue E346 causes deprotonation of the intermediate and aids in regioselective incorporation of the chlorine in the indole ring of tryptophan.

1.6.2. Enzymes characterised so far and their substrate scope

Many flavin dependent halogenases have been identified but the enzymes that we are most interested in are the ones that chlorinate regioselectively at the 5-, 6- and 7-positions of tryptophan^{64,71,72}. All these enzymes share a common mechanism (see section 1.6.1) and contain the same conserved motifs. A GXGXXG conserved motif, usually found in the N-terminal region of the protein, is responsible for FAD binding. A conserved WxWxIP motif is found at the centre of the protein, and its role is to prevent FAD from binding too close to the substrate thus preventing a monooxygenase reaction (MOR). MOR reaction follows Bayer villiger reaction that inserts a oxygen to the α -carbon of carbonyl carbon. For example cyclohexanone monooxygenase enzyme follows the reaction to produce caprolactone⁷³. Finally, the lysine residue that forms the chloramine and the glutamate residue that causes deprotonation of the intermediate are also conserved in all flavin dependent halogenases. Most flavin dependent halogenases are identified based on the presence of the above mentioned conserved motifs.

The first flavin dependent halogenase PrnA was discovered in 1997 while studying pyrrolnitrin biosynthesis in *P. fluorescens* BL915⁶². Pyrrolnitrin is an antibiotic with a broad spectrum of antifungal activity and is produced by the *prnABCD* biosynthetic genes (Scheme 8). PrnA catalyses the chlorination of L-tryptophan to form 7-chlorotryptophan. Insights into the mechanism of regioselective chlorination by this enzyme were gained by solving the crystal structure (See section 1.6.3 for more details).



Scheme 8. Scheme showing biosynthesis of pyrrolnitrin from tryptophan using enzymes PrnA, PrnB, PrnC and PrnD (adapted using ChemDraw from ref⁷⁴)

Another tryptophan-7-halogenase RebH was identified from the rebeccamycin biosynthesis gene cluster. Rebeccamycin is an indocarbazole derived natural product produced by *Lechevalieria aerocolonigenes* ATCC 39243 that acts as a topoisomerase I inhibitor and has potential activity as a result of protein kinase inhibition⁷⁵. The protein halogenates tryptophan to produce 7-chloro-L-tryptophan. RebH has 56% identity to PrnA. A detailed mechanistic view of the enzyme was obtained by solving the crystal structure and will be discussed in section 1.6.3. Substrate specificity studies of RebH have been carried out. Tryptamine, a direct precursor to many alkaloid natural products was tested for chlorination using RebH by Sarah O'Connor's group⁷⁶. Attempts were made to bring about regioselective chlorination of tryptamine by site directed mutagenesis of RebH. The RebH mutants were then introduced into *Caranthus roseus* and led to the successful production of halogenated alkaloids. Recently Payne *et al.*, 2013 showed a much wider substrate range for RebH showing chlorination of N-methyltryptamine, tryptophol, 2-methyltryptamine, tryptoline and substituted naphthalenes⁷⁷.

KtzQ is a third tryptophan-7-halogenase which regiospecifically chlorinates at the 7-position of L-tryptophan. KtzQ was identified as part of the kutzneride

biosynthetic pathway⁷⁸. Kutzneride is a cyclic depsipeptide isolated from an actinomycete *Kutzneria* sp.744. KtzQ has 40% identity to PrnA and 60% identity to RebH. The enzyme has not been characterised for its substrate scope or its regioselectivity.

Relative to the tryptophan-7-halogenases, less is known about tryptophan-6-halogenases which chlorinate tryptophan at the 6-position to produce 6-chloro-tryptophan. KtzR is a tryptophan-6-halogenase regiospecifically dichlorinates L-7-tryptophan to produce 6,-7-dichloro-L-tryptophan⁷⁸. SttH is a putative tryptophan-6-halogenase found within the non-ribosomal peptide synthetase gene cluster of *Streptomyces toxyticini* NRRL15443⁷⁹. It has 75% similarity to KtzR. The substrate scope of both of these enzymes and the exact regioselective halogenation mechanism have yet to be elucidated. ThaI is another flavin dependent halogenase showing 58% identity to PrnA, found in biosynthetic gene cluster of thiendolin from *Streptomyces albogrieslus* that chlorinates tryptophan at the 6-position⁸⁰.

PyrH has been identified as a tryptophan-5-halogenase that chlorinates tryptophan to produce 5-chloro-tryptophan and was first identified from the pyrroindomycin biosynthetic gene cluster of *Streptomyces rugosporus* LL-42D005⁷². Pyrroindomycin has excellent antibacterial activity against gram-positive bacteria such as MRSA. PyrH has 56% sequence identity to PrnA, 52% to RebH and 54% identity to KtzR. Insight into the mechanism of regioselective halogenation by PyrH has been obtained from a crystal structure (see section 1.6.4).

CmdE is another tryptophan halogenase from chlondramide biosynthesis in *Chondromyces crocatus* Cm c5 that chlorinates at the 2-position⁸¹. There are other examples of flavin dependent halogenases that have not yet been fully characterised and are able to chlorinate a number of differential group. CndH from chondrochlorine biosynthesis halogenates the PCP bound tyrosine at the position 2⁸². Similarly, Sgc3 also chlorinates PCP tethered tyrosine to produce antitumor compound C-1027 in enediyne biosynthesis⁸³ while PltA chlorinates pyrole bound to PCP to produce 4,5-dichloropyrrolyl in pyoluterin biosynthesis⁸⁴. Rdc2, a fungal flavin dependent

halogenase from *Pochonia chlamydosporia* was found to have substrate specificity towards molecules such as dihydroresocrylide with both mono and dihalogenation⁸⁵. CmlS is another such example that performs a dichlorination of its substrate leading to the formation of chloramphenicol⁸⁶.

1.6.3. Crystallographic studies of flavin dependent halogenases

To date, only three crystal structures of PrnA (PDB ID:2AR8)⁶⁴, RebH(PDB ID:2OAI)⁸⁷ (both tryptophan-7-halogenases) and PyrH (PDB ID: 2WET)⁸⁸ (a tryptophan-5-halogenase) have been solved. The crystal structures of PrnA and RebH were solved at 1.95 Å and 2.15 Å resolution respectively. Both structures were obtained with tryptophan, FAD and Cl⁻ bound to the enzyme. The structures are similar globally. They are dimers forming a pyramidal trigonal structure. However, there are two differences in the structure of RebH compared to that of PrnA. The loop regions between residues 40-48 and residues 86-105 adopt different conformations in RebH compared to PrnA (Figure 9).

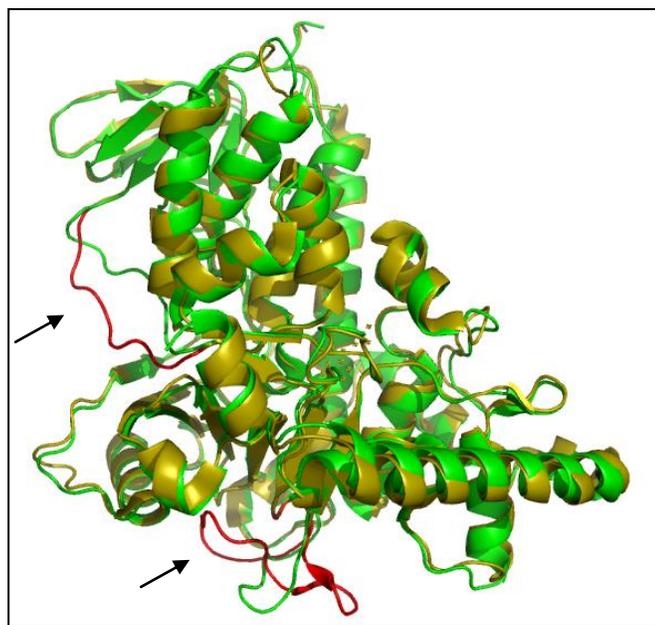


Figure 9. Crystal structure of overlaid PrnA(PDB ID:2AR8) (green) and RebH(PDB ID:2OAI) (yellow) showing the trigonal shape, loops that adopt different conformations are highlighted in red for RebH (generated using PYMOL¹).

Comparison of the active sites of PrnA and RebH shows that the FAD, Cl⁻ and tryptophan are bound in the same orientation in both of these enzymes (Figure 10). The conserved residue K79 is in the same position in both enzymes. The other conserved residue E346 in PrnA and E357 in RebH is also aligned near the active site in both structures.

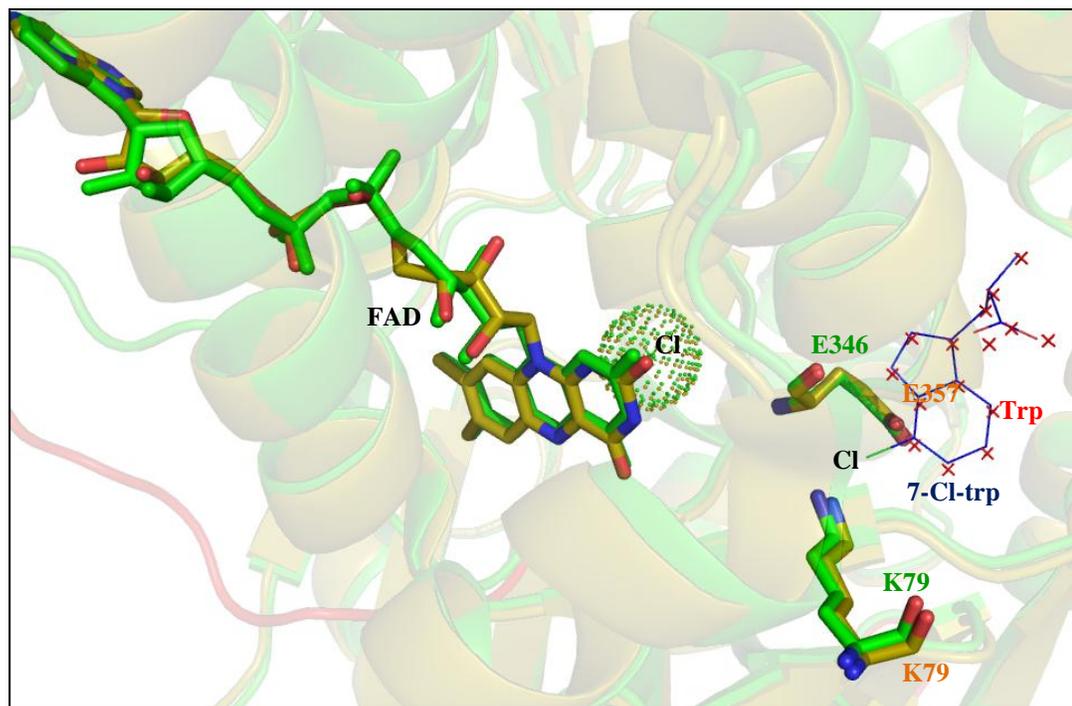


Figure 10. Crystal structures showing similar FAD, Cl⁻, tryptophan, K79 and E346/E347 binding site in overlayed of RebH (PDB ID:2OA1) and PrnA(2AR8) generated using PYMOL software¹ (orange aa residues-RebH, green aa residues-PrnA, cross reds- tryptophan in RebH, blue-chlorinated tryptophan by PrnA)

Other interactions in the active site are also conserved between PrnA and RebH (Figure 11). Y443 in PrnA is equivalent to Y454 in RebH and its hydroxyl group interacts with the amino group of the tryptophan. Similarly, Y444 in PrnA is equivalent to Y455 in RebH and the hydroxyl group of this residue interacts with the carboxyl group of tryptophan. E450 of PrnA is equivalent to E460 in RebH with the oxygen atom interacting with the amino group of the substrate. Finally, F454 in PrnA is equivalent to F464 in RebH with the carboxyl group interacting with the amino group of the substrate.

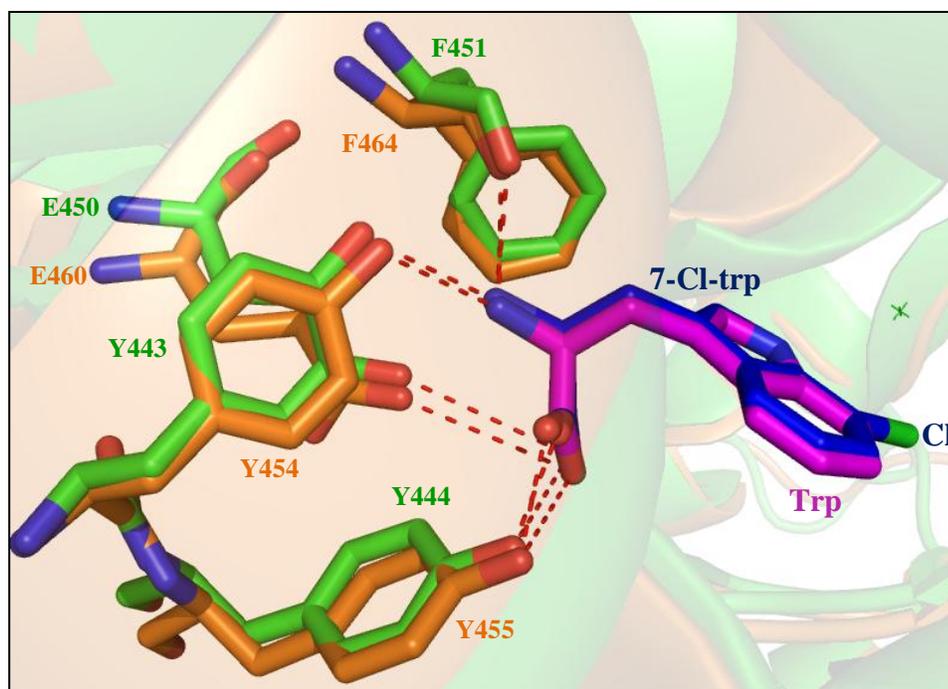


Figure 11. Active site showing similar interactions in overlaid crystal structures of RebH (PDB ID: 2E4G) and PrnA(2AR8) generated using PYMOL software¹ (orange aa residues-RebH, green aa residues-PrnA, magenta-tryptophan in RebH, blue-chlorinated tryptophan in PrnA)

From the above crystal structures it is clear that the tryptophan-7-halogenases adopt similar global structures and have the same active site for regioselective chlorination. However, when the crystal structure of a tryptophan-5-halogenase was solved the active site was found to be different.

The crystal structure of PyrH (tryptophan-5-halogenase) was solved at 2.4 Å (PDB ID: 2WET)⁸⁸. The enzyme was found to consist of 2 major domains, each existing as dimer joined together to form 4 domains. The overall conformation of the enzyme is similar to that of tryptophan-7-halogenases. Sequence alignment showed that PyrH shares only 56% identity to PrnA which could be observed in the overlaid crystal structure with PrnA. The structure revealed that the FAD and chloride ion binding sites are the same as to PrnA. However, the substrate tryptophan is bound in a different orientation in the active site relative to that in PrnA. Some of the interactions are conserved between the two classes of halogenase while others differ. The conserved

lysine (K75 in PyrH and K79 in PrnA) and glutamate (E354 in PyrH and E346 in PrnA) are aligned in the PyrH and PrnA structures. However the tryptophan amino group interacts with the carboxyl group of the F451 residue and the carboxyl group of the tryptophan is held by interactions with the amino group of Q163 and both the carboxyl and amino groups of the S50 residue (Figure 12).

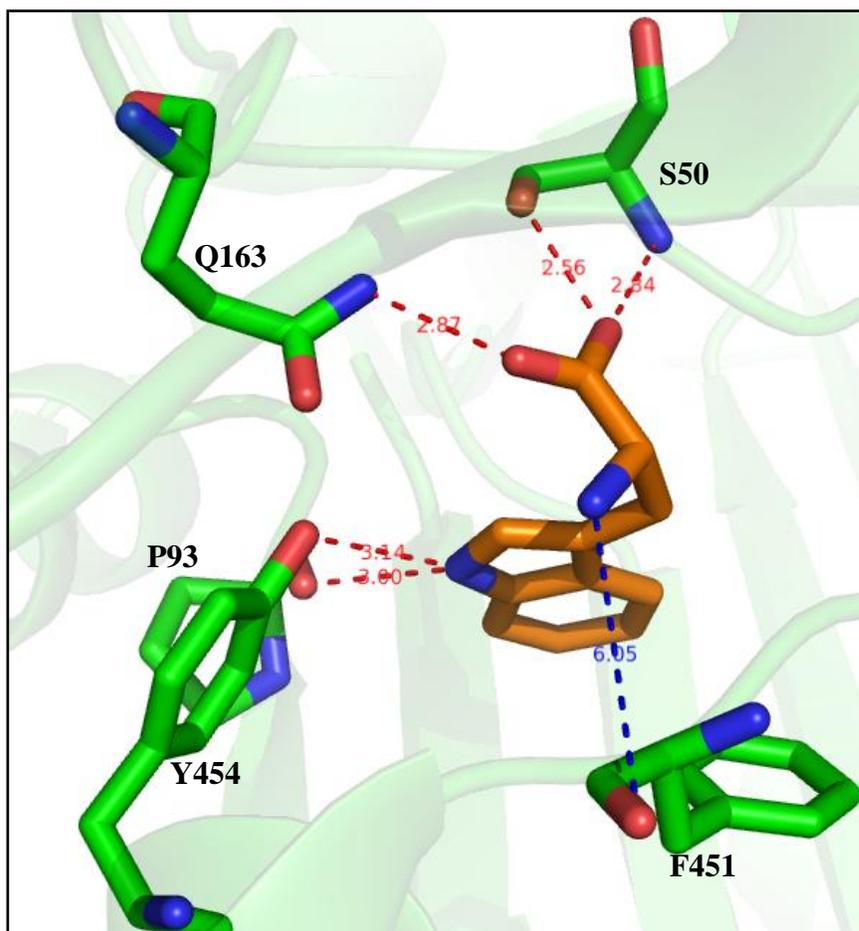


Figure 12. Crystal structure of PyrH (PDB ID: 2WET) generated using PYMOL¹ showing the interactions of tryptophan with residues in the active site pocket

The amino acids that interact with the tryptophan in PrnA are absent from PyrH due to the absence of a loop region holding these amino acids.

1.6.4. Regioselectivity studies based on crystal structures

The biggest advantage of the flavin dependent halogenases is that they can regioselectively halogenate tryptophan which is an important goal of industry when developing them as novel biocatalysis tools. Recently Karl-Heinz van Pee's group have carried out studies to identify important residues governing the regioselectivity of these enzymes. They found that the regioselectivity of the enzymes depends upon the binding of the substrate through hydrogen bonds and the residues that governing π stacking around the active site. In PrnA, residues F103 and H101 forms the π stacking sandwich around tryptophan, while the substrate is held by the hydrogen bond interactions with residues E450, Y443, Y444, F454 and E346 (Figure 13). The indole of tryptophan is shielded by the F103 and W455 residues⁸⁹.

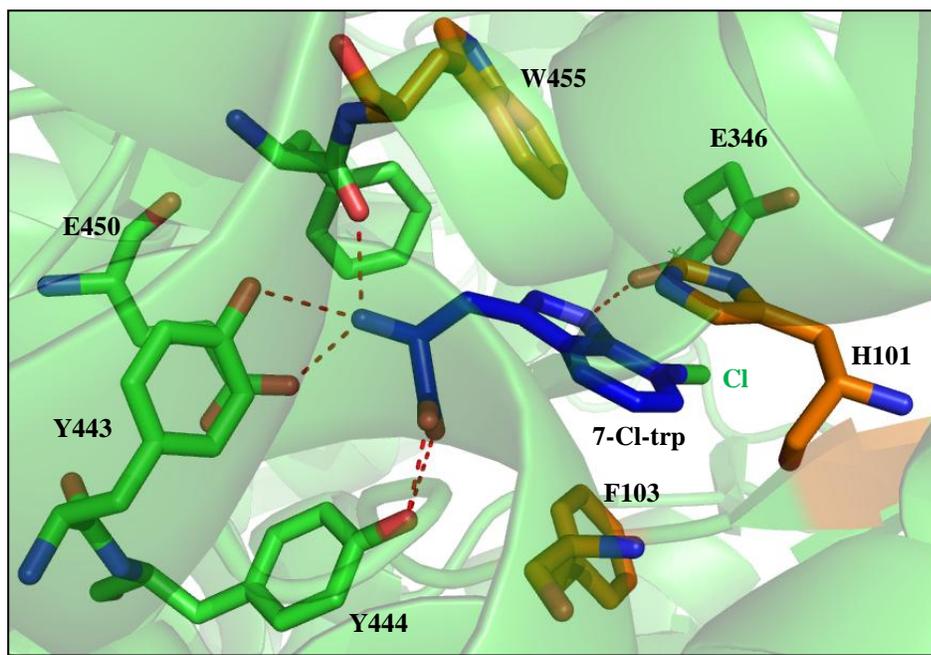


Figure 13. Crystal structure of (PrnA PDB ID:2AR8) generated using PYMOL¹ showing residue F103, H101 and W455 forming π stacking (orange) around the tryptophan and the residues (green) providing the hydrogen bond interactions holding 7-Cl-tryptophan (blue)

A Lang et al., 2011 mutated the residues W455, F103 and H101 to alanine by site directed mutagenesis. They found that the F103A mutant could produce both 7-chloro-

tryptophan and low amount of 5-chlorotryptophan. When the same mutant was analysed for bromination they observed higher production of 5-brominated-tryptophan. The other mutants showed slower kinetics indicating that they are important for substrate binding.

In PyrH the tryptophan binding is very different to that of PrnA. The indole ring of the tryptophan interacts with residues P93 and Y454 which position the substrate for chlorination at the 5-position. The hydroxyl group of Y454 and the carbonyl group of P93 hydrogen bond with the indole nitrogen of the tryptophan. Residues S50, Q163 and F451 also interact with tryptophan through hydrogen bonding. Residues F94 and F49 sandwiches the tryptophan through π stacking interactions⁸⁸. It was observed that residue F49 would allow more space for binding for tryptophan while Y454 would change the orientation (Figure 14).

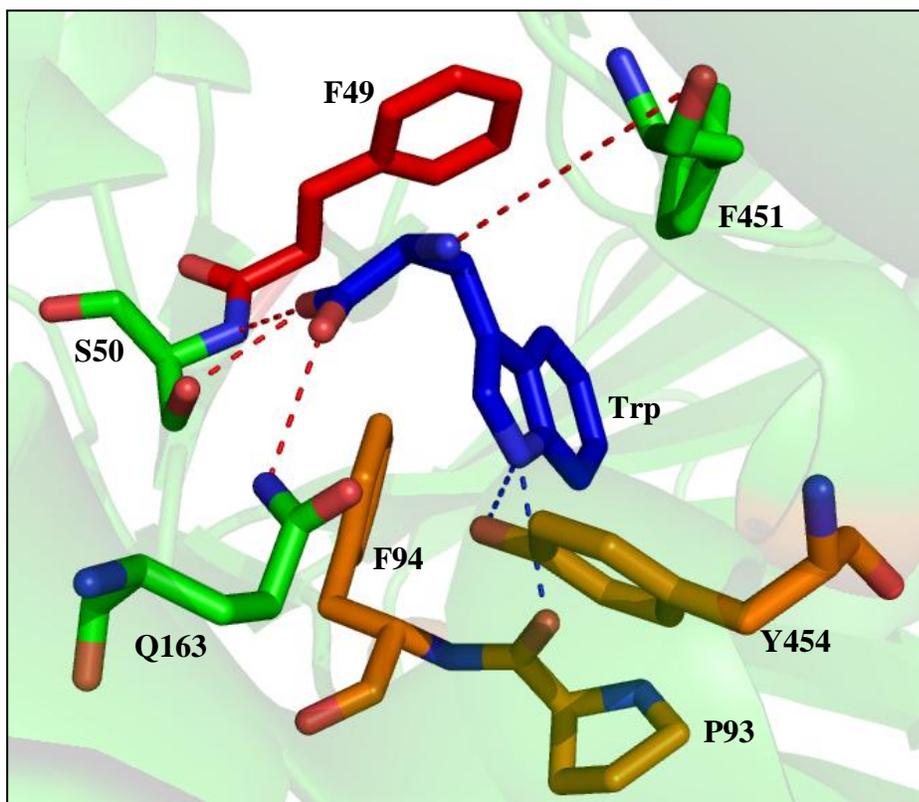


Figure 14. Regioselectivity of PyrH(PDB ID:2WET) generated using PYMOL¹ showing, Y454, P93 and F94(orange) forming sandwich residues around tryptophan, red- residue which would allow space for the different orientation of tryptophan relative to tryptophan-7-halogenases, other interacting residues (green)

Zhu et al, 2009 mutated the F49 residue to an A residue and Y454 to an F residue. However they did not observe any change in the regioselectivity of the halogenation⁸⁸.

1.6.5. Structural evidence of halogenation

The FAD binds to the GxGxxG motif in the active site and the Cl⁻ ions is found to bind between the isoalloxazine ring of the FAD and residues T348 and G349. The chloride ion bound to FAD reacts with hydroperoxide to form hypochlorous acid and is separated from the substrate by a distance of 10 Å. This hypochlorous acid is guided through a tunnel which is made up of amino acids I52, K79, I82, S347, and T348. The K79 reacts with hypochlorous acid forming the chloramine halogenation source which undergoes a further electrophilic aromatic substitution.

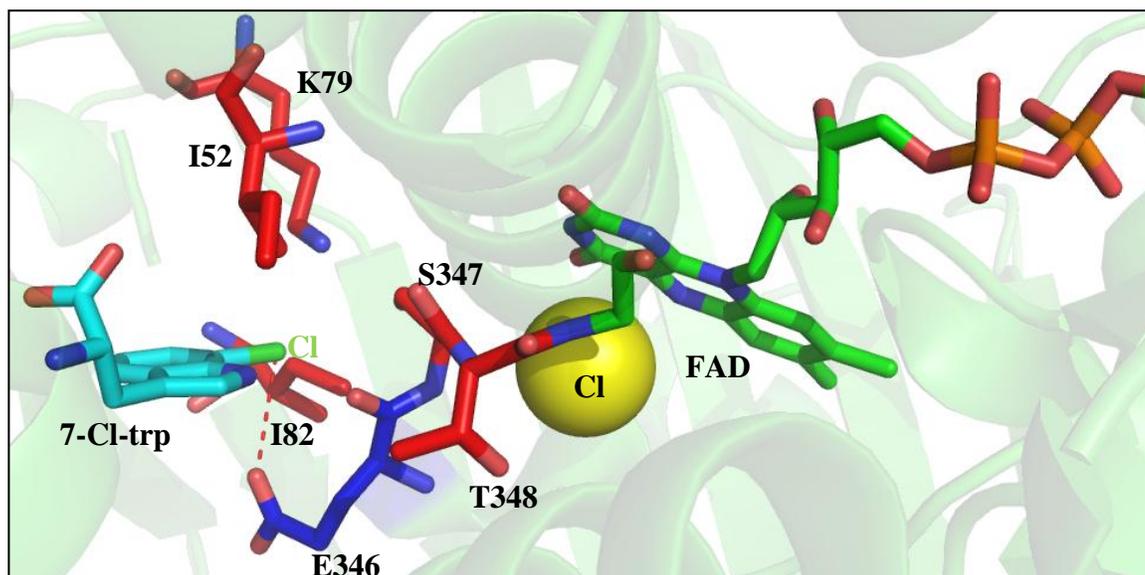


Figure 15. Crystal structure(PDB ID:2AR8) generated using PYMOL¹ showing Cl⁻ bound to the face of FAD with 10 Å away from the product 7-Cl-trp(cyan), and the residues forming the tunnel between the FAD and substrate binding sites are shown in red.

1.6.6. Conclusion

The above sections of this chapter have reviewed the current literature in the field of halogenases. The review focuses especially on Ram20 which chlorinates the Hpg non-proteinogenic amino acid and on flavin dependent halogenases in general. The following

general aims have been formulated to encompass these two broad categories of halogenase:

1.7. Aims

1.7.1. Aim 1

The non-proteinogenic amino acid Hpg has high pharmaceutical value. Ramoplanin has an abundance of both D and L-forms of Hpg and contains a chlorinated Hpg at the chain terminus. It is predicted that the protein Ram20 halogenates the Hpg-17. Previous studies have shown that Ram20 is a halogenase that catalyses chlorination. However, no evidence has yet been obtained for Ram20 chlorinating the Hpg-17 residue of ramoplanin. The aim of the project is to investigate Ram20-mediated chlorination of Hpg-17 by constructing an *in-vitro* system. Due to the fact that the Ram20 accepts only substrates that are tethered to the PCP domain of NRPSs, the PCP from the 8th module of the ramoplanin NRPS will be cloned and expressed in *E. coli*. The PCP of the 8th module will be expressed with the conserved serine residue which is usually the site for attachment of the Co-enzyme A (CoA) synthesised substrate. Simultaneously Co-enzyme A-Hpg (CoA-Hpg) and N-acetylcystamine thioester-Hpg (SNAC-Hpg) will be synthesised in the lab. In order to load the synthesised substrate onto the conserved serine residue of the PCP module the phosphopanthenyl transferase (PPTase) Sfp will be expressed and purified in *E. coli*. Sfp recognises the apo-PCP and posttranslationally modifies the conserved serine residue to form an active holo-PCP. The loading of the CoA-Hpg and SNAC-Hpg by Sfp will be analysed by mass difference either by LC-MS or MALDI. Ram 20 will be cloned and expressed in *E. coli*. Since Ram20 has been predicted to be a flavin dependent halogenase, flavin reductases will also be required. The flavin reductases SsuE and Fre will be cloned and expressed in *E. coli*. Halogenation of Hpg-17 of the loaded substrate with Ram20 and the associated flavin reductases will then be tested. The successful generation of this system could be used as a novel tool to introduce chlorinate Hpg residues in alternative biosynthetic systems for the production of novel antibiotics.

1.7.2. Aim 2

Recent identification of flavin dependent halogenases has opened a new gateway for producing much cleaner and safer halogenated compounds. In addition there is potential to use these enzymes as novel biocatalysts for the production of chemo/regioselective halogenated compounds. Insight into the regioselective mechanism of halogenation of these enzymes have been obtained from crystallographic studies. However, few studies have been carried out on these enzymes with regards to the scope of the substrates they can halogenate and altering their regioselectivity. For example, only RebH has been tested with aromatic compounds other than the natural substrate tryptophan. In terms of regioselectivity, slight production of both 5- and 7-chloro-tryptophan has been observed for a PrnA F103A mutant. In this project we will focus on the PrnA and KtzQ tryptophan-7-halogenases, the KtzR and SttH tryptophan-6-halogenases and PyrH tryptophan-5-halogenase. All of these enzymes will be expressed in *E. coli* for the first time. Flavin reductases SsuE and Fre will also be cloned and expressed in *E. coli* for performing *in vitro* halogenation assays. The purified halogenases will be assayed for halogenating activity with the natural substrate tryptophan and industrially important aromatic and phenol-based compounds. The halogenated compounds produced will be characterised by mass spectrometry and NMR analysis. Rational mutants will be designed using the available crystal structures to improve the function of these enzymes in terms of substrate scope and will be generated by site directed mutagenesis. Studies will also be carried out to identify residues that can switch the regioselectivity of chlorination.

Efforts will also be focussed on obtaining the crystal structure of a tryptophan-6-halogenase, which would provide insight into regioselective halogenation at the 6-position. This crystal structure will be used to guide further mutagenesis studies to widen substrate scope. Finally, structural studies of interesting mutants with target compounds will also be pursued.

Chapter 2. Construction of an *in-vitro* halogenation system for chlorination of Hpg-17

17: results and discussion

2.1. Introduction

Ramoplanin is a non-ribosomal peptide (NRP) natural product that was discovered in 1984 by Biosearch Italia and exhibits antibiotic activity against gram positive bacteria. This NRP contains an abundance of non-proteinogenic Hpg amino acids and chlorination of Hpg-17 located at the C-terminus of the NRP is associated with the antibacterial properties of the NRP. Ram20 is predicted to chlorinate the Hpg-17 residue but there is currently no direct evidence for this (Figure 16). We are interested in investigating Ram20's ability to halogenate the Hpg-17 non-proteinogenic amino acid using an *in vitro* system following ref⁹⁰.

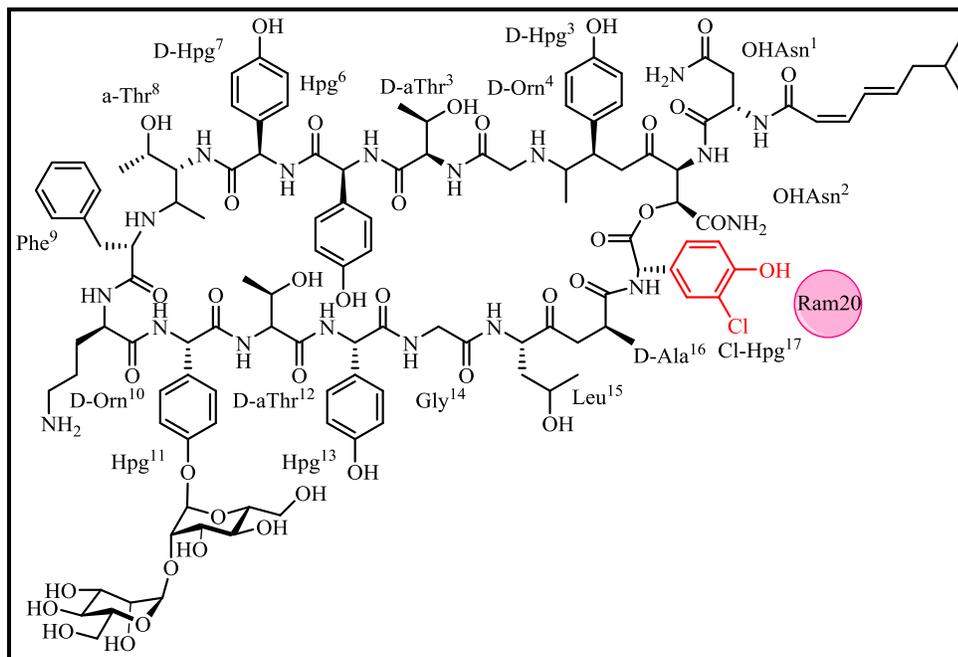
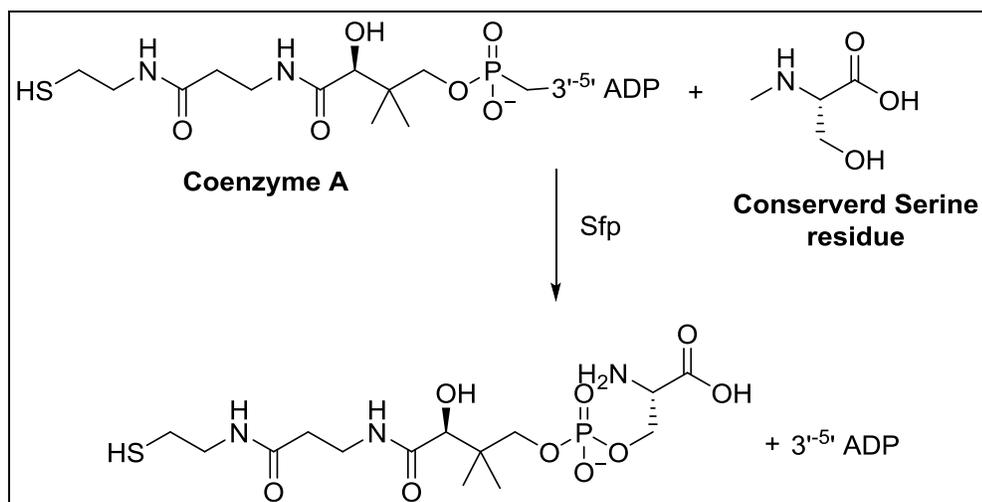


Figure 16. Structure of ramoplanin A3 showing chlorination of Hpg-17 by Ram20 (adapted using chemdraw from ref⁹⁰)

Most of the enzymes responsible for halogenating amino acids in NRPs only accept substrate that are tethered to the peptidyl carried protein (PCP) domain of the catalytic domain of the non-ribosomal peptide synthetase (NRPS) complex. During the biosynthesis of ramoplanin, Hpg-17 is tethered to the PCP domain of 8th catalytic

module of the NRPS. We are interested in determining which, if any form of Hpg is a substrate for Ram20: free Hpg or Hpg tethered to a PCP domain. The initial aim of the project was to produce a model substrate for Ram20 consisting of Hpg tethered to a minimal PCP domain. Two fragments of PCP domain (S-PCP: 140 amino acids and L-PCP: 325 amino acids) from the 8th module of ramoplanin NRPS both containing a conserved serine residue required for Hpg attachment were, cloned and expressed in *E. coli*. CoA-Hpg and SNAC-Hpg were chemically synthesised by Matthew Styles from the laboratory of Prof. Jason Micklefield. *In vivo*, Hpg is loaded onto the PCP domain by phosphopantetheinyl transferases. This enzyme catalyses the formation of phosphodiester bond between the phosphapantetheinyl moiety of coenzyme A (CoA) and the hydroxymethyl side chain of a conserved residue in the PCP domains⁹¹ (Scheme 9). Sfp from *Bacillus subtilis* is an example of such an enzyme and will be cloned and expressed in *E. coli* for the use in Hpg loading. The successful loading of Hpg onto the PCP domain will be monitored by MALDI.



Scheme 9. Conserved Glu151 residue of Sfp causes deprotonation of hydroxyl group of the conserved serine residue of PCP domain followed by nucleophilic attack on the β -phosphate of coenzyme A (adapted using chemdraw using ref⁹¹)

Ram20 and flavin reductases, required for production of FADH₂, were cloned and expressed in *E. coli*. Purified enzymes were used in an *in vitro* halogenation assay with the PCP-tethered Hpg as a substrate to study the chlorination of Hpg by Ram20.

2.2. Cloning and expression of the small PCP (S-PCP) domain in *E. coli*

Genomic DNA (gDNA) isolated from *Actinoplanes* ATCC 30766 mycelium was used as the template to amplify *s-pcp* 420 bp; (appendix) using primers P1 and P2 (Table 32) and *s-pcp* was cloned into *NdeI/NotI* restricted pET-28a(+) for expression of an N-terminal His₍₆₎tagged S-PCP(pET-KK1). The clones were verified by restriction digest of the plasmid (Figure 17) and sequencing (GATC, Germany).

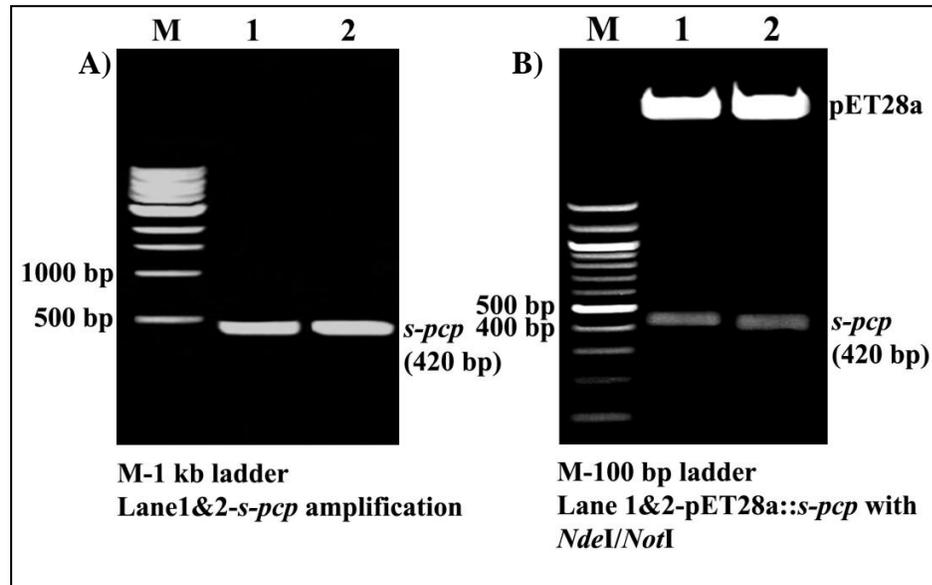


Figure 17. A) PCR amplification of 420 bp *s-pcp* and B) Restriction digestion analysis of *s-pcp* cloning

In order to express S-PCP pET-KK21 was transformed into *E. coli* BL21 (DE3) cells (section 5.13). Expression studies were initially carried out on a small scale at various temperatures (Section 5.15.1). Expression was observed at all of the temperatures, but was highest at 30 °C (data not shown) and this temperature was used for large scale expression for purification. S-PCP was purified by Ni²⁺ affinity (Section 5.15). Pure S-PCP was eluted with 500 mM imidazole buffer. However, some S-PCP eluted along with contaminating proteins at lower imidazole concentrations. Problems also occurred with the storage of the protein in phosphate buffer since degradation occurred. Due to this, only 0.75 ml S-PCP at a low concentration of 1.2 mg ml⁻¹ was obtained. 20 µl of pure S-PCP was analysed by 20% (v/v) SDS-PAGE. A band was observed just above the 15 kDa marker consistent with the expected mass (16.9 kDa) of S-PCP (Figure 18).

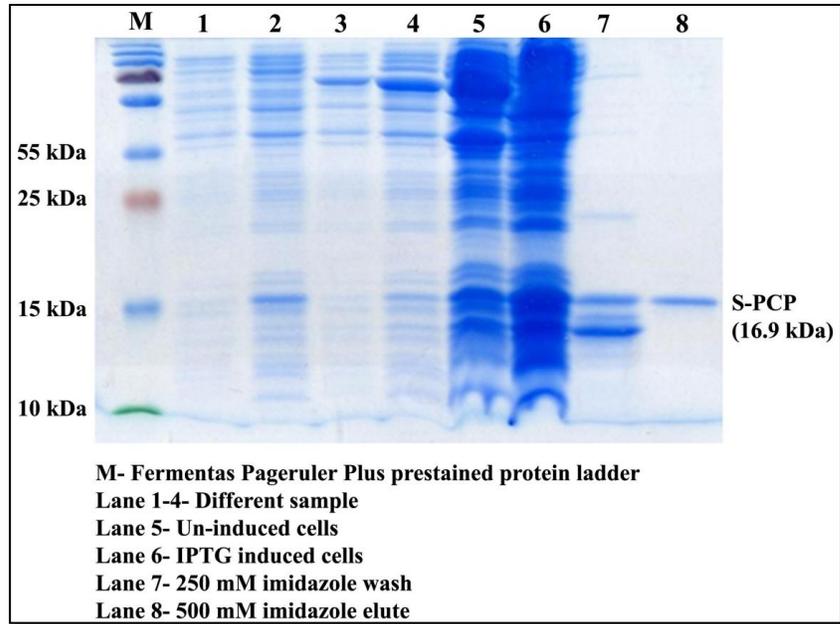


Figure 18. Analysis of pure S-PCP protein by 20% (v/v) SDS-PAGE.

2.3. Cloning and expression of the large PCP domain (L-PCP) in *E. coli*

Cloning of *l-pcp* 975 bp; (appendix) into pET-28a+ was carried out as for *s-pcp* (Section 2.2) using primers P3 and P4 to amplify the relevant DNA sequence for expression as an N-terminal His₍₆₎-tagged L-PCP (pET-KK2). Clones were confirmed by restriction digestion analysis and sequencing (Figure 19).

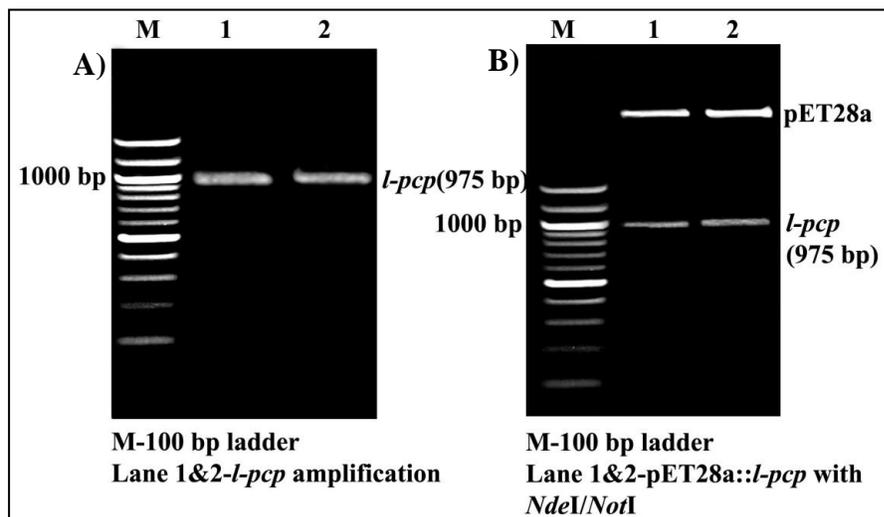
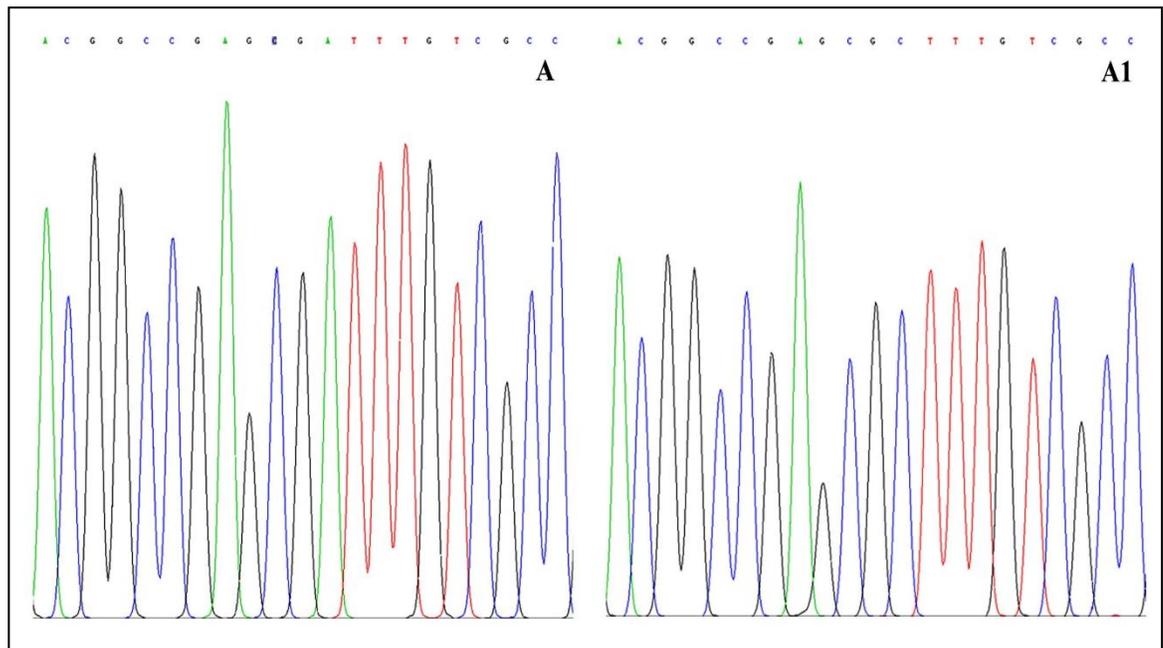


Figure 19. A) PCR amplification of 975 bp *l-pcp* and B) Restriction digestion analysis of *l-pcp* cloning

For expression studies, pET-KK2 was transformed into *E. coli* BL21 (DE3) and BL21 Rosetta 2 (DE3) cells (section 5.13). The BL21 Rosetta 2(DE3) was used because this strain has been engineered to express tRNAs for codons utilised rarely in *E. coli*. Analysis of the *l-pcp* DNA sequence by Rare Codon Calculator (RaCC) software (NIH, MBI laboratory for Structural Genomics and Proteomics, University of California, USA) revealed the presence of two CGA (Arg) that was used rarely (0.3%) in *E. coli*. Expression was tested in small scale (section 5.15.1) in both strains. The results showed no expression in either strains under any of the conditions tested. Consequently, site directed mutagenesis was carried out (section 5.14) using primers P5 and P6 (Table 32) to change the two rare codon CGA to CGC (2.2%). Clones were verified by DNA sequencing (pET-KK3) (GATC, Germany) (Figure 20).



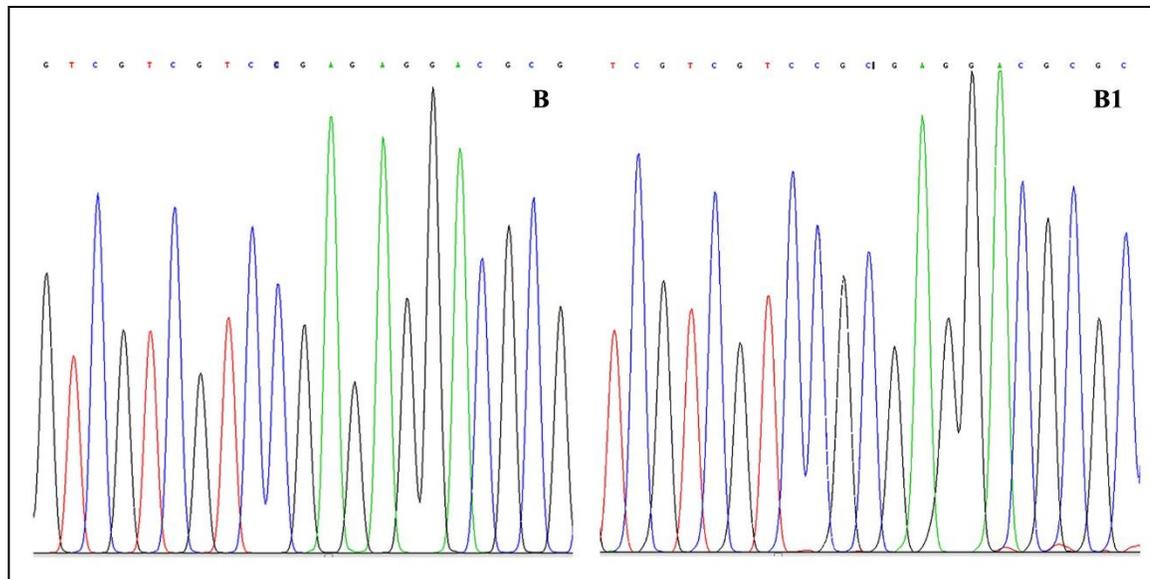


Figure 20. DNA sequencing chromatogram showing mutations of CGA to CGC in *l-pcp*. (A) & (B) – *l-pcp* wild type sequence with CGA codons (A1) & (B1) – *l-pcp* mutant sequence with CGC codons. The relevant codons are underlined in both the wild type and mutant sequences. Colours (Red- T, Blue- C, Black-G, Green-A) Mutant clones were verified for mutations of CGA to CGC by DNA sequencing (pET-KK3) (GATC, Germany) (Figure 20).

The construct pET-KK3 was transformed into *E. coli* Rosetta 2 (DE3) cells and *E. coli* ArcticExpress (DE3). *E. coli* ArcticExpress (DE3) cells have been engineered to express chaperons at low temperature to assist with protein folding. Expressions were tested as described in section 5.15.1. The results showed expression of a soluble protein that runs at a position consistent with that expected for L-PCP (36.9 kDa) in both Rosetta 2 (DE3) and ArcticExpress (DE3). However, better expression was observed with Rosetta 2 (DE3) cells (Figure 21) and this strain was used for large scale expression for purification.

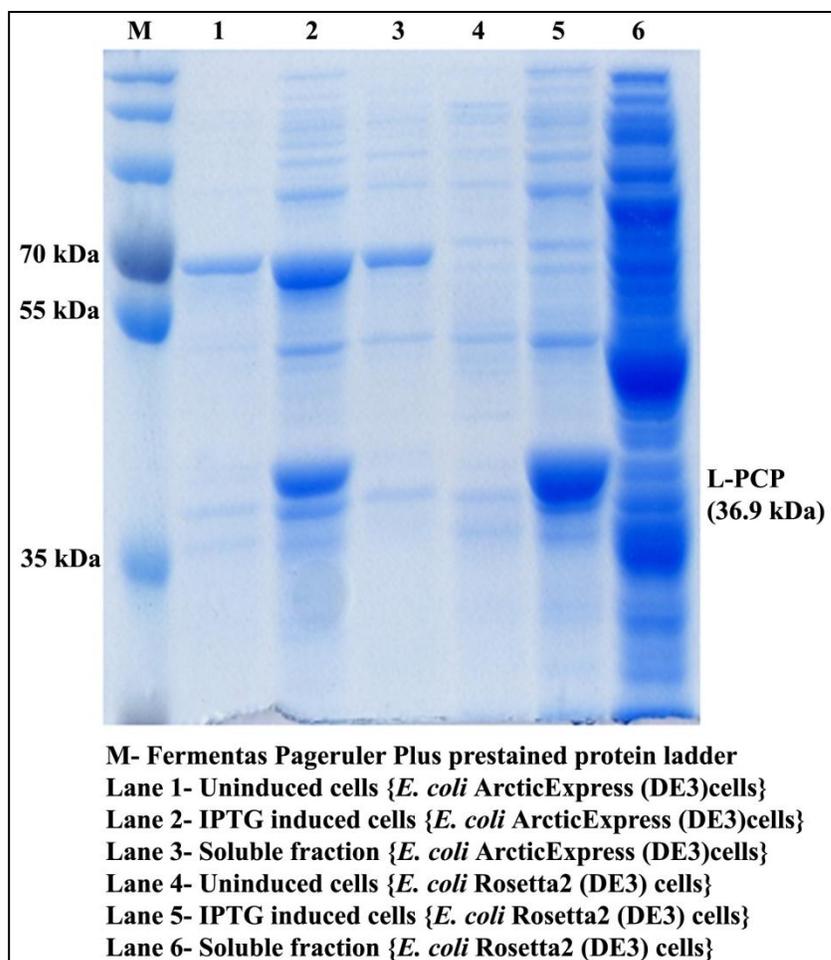


Figure 21. Expression studies of L-PCP protein analysed by 12% (v/v) SDS-PAGE.

L-PCP was expressed on a large scale (800 ml) in *E. coli* Rosetta 2 (DE3) cells. The protein was purified by Ni²⁺ affinity chromatography and concentrated as described in section 5.15. A band consistent with the expected size of L-PCP (36.9 kDa) was observed by 15% (v/v) SDS-PAGE (Figure 22). 1 ml of L-PCP at a concentration of 0.8 mg ml⁻¹ was obtained.

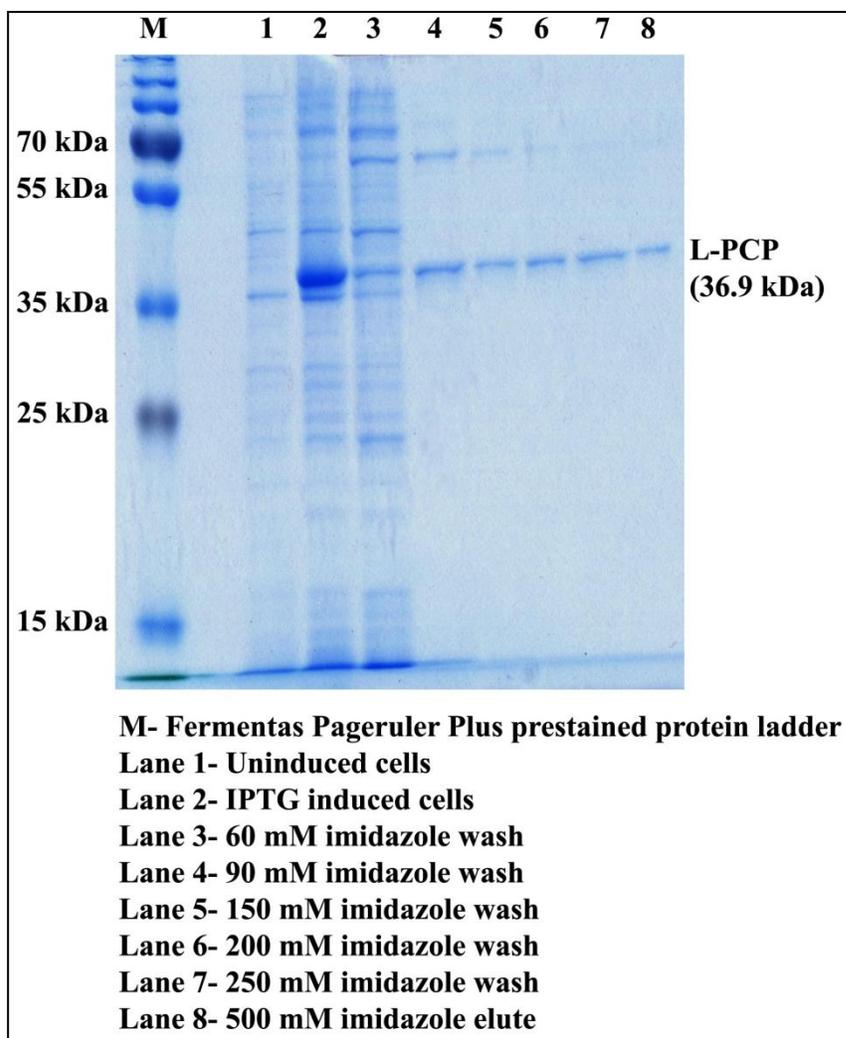
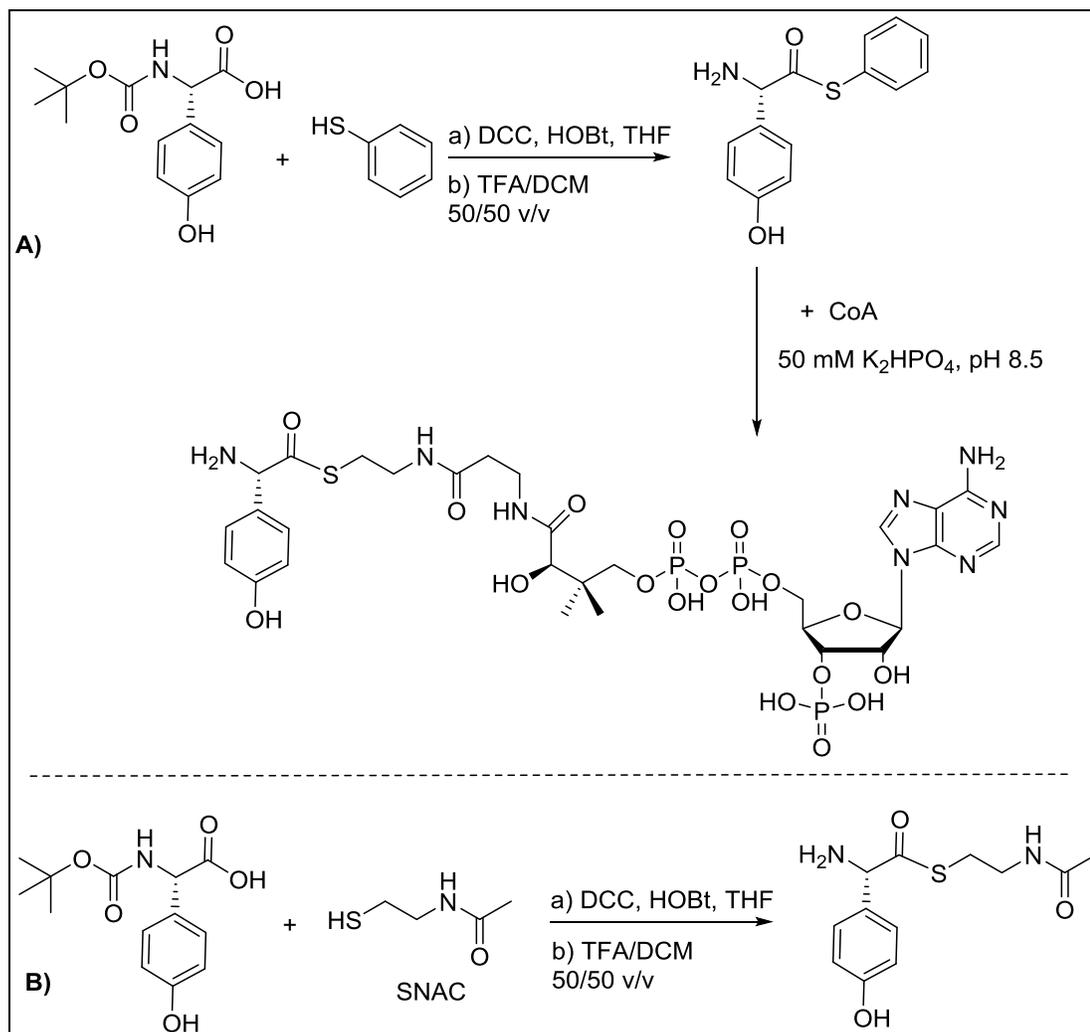


Figure 22. Analysis of pure L-PCP protein by 15% (v/v) SDS-PAGE.

2.4. Synthesis of co-enzyme A-Hpg (CoA-Hpg) and N-acetylcysteamine thioester-Hpg (SNAC-Hpg)

The synthesis of both the Hpg compounds to be loaded onto the PCP domain: CoA-Hpg and SNAC-Hpg was carried out by Matthew Styles in our laboratory as detailed in Scheme 10. The *N*-Boc protected hydroxyphenylglycine was coupled with thiophenol using coupling-deprotection strategy to form thioester intermediate which is further reacted with CoA under standard conditions to give the corresponding CoA-Hpg compound. (Scheme 10, A) Similarly, *N*-Boc-protected hydroxyphenylglycine was

coupled with SNAC using standard coupling strategy followed by deprotection to obtain desired SNAC-Hpg compound. (Scheme 10, B)



Scheme 10. A) Synthesis of CoA-Hpg using N-Boc protected Hpg and thiophenol and B) synthesis of SNAC-Hpg using SNAC and N-Boc protected Hpg

A phosphopantetheinyl transferase (PPTase) enzyme is required to load Hpg onto the PCP domain. The *sfp* gene (675 bp; appendix) encoding the Sfp PPTase was amplified from *B. subtilis* gDNA and was cloned into *NdeI/XhoI* restricted pET-28a(+) to express an N-terminal His(6)-tagged protein by Dr. Lu Shin Wong from Jason Micklefield laboratory. The clone was verified by sequencing (GATC, Germany). The construct pET-28a(+>::*sfp* was transformed into *E. coli* BL21(DE3) cells following section 5.13.

The expression, purification by Ni²⁺ affinity chromatography and concentration was carried out as mentioned in section 5.15. A band of the expected size for pure Sfp (27.8 kDa) was observed by 20% (v/v) SDS-PAGE (Figure 23). 1.2 ml at a protein concentration of 17 mg ml⁻¹ was obtained.

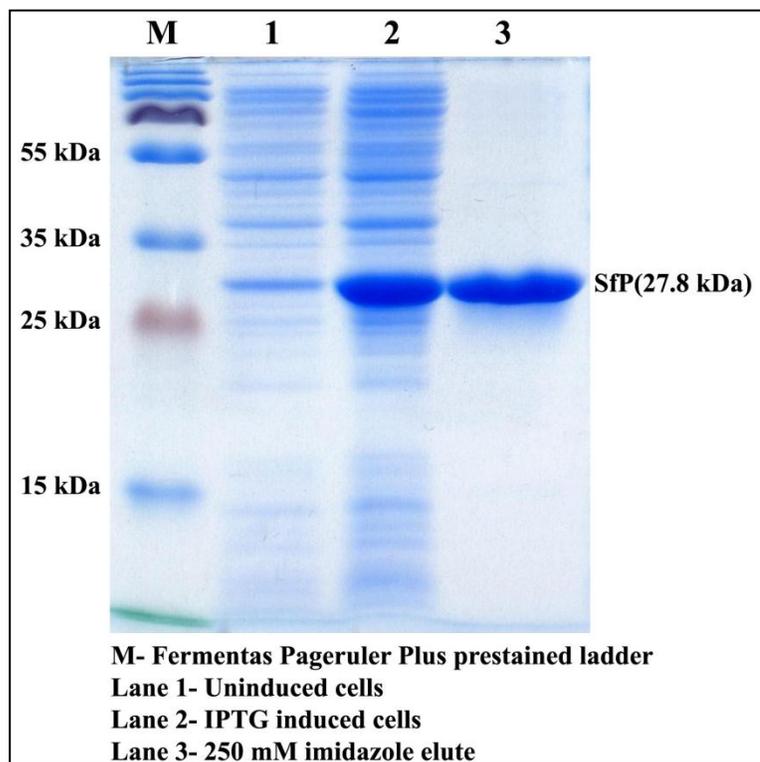


Figure 23. Analysis of pure Sfp protein by 20% (v/v) SDS-PAGE.

2.5. Substrate loading and analysis of PCP and PCP tethered Hpg by MALDI

Assays for loading the Hpg compounds onto PCP were carried out as per section 5.20. All the MALDI experiments were performed by Matthew Styles. Initially S-PCP and L-PCP alone were analysed by MALDI. S-PCP could be detected at a m/z of 16,993.5 (theoretical mass: 16,993 daltons) (Figure 24). But L-PCP could not be detected.

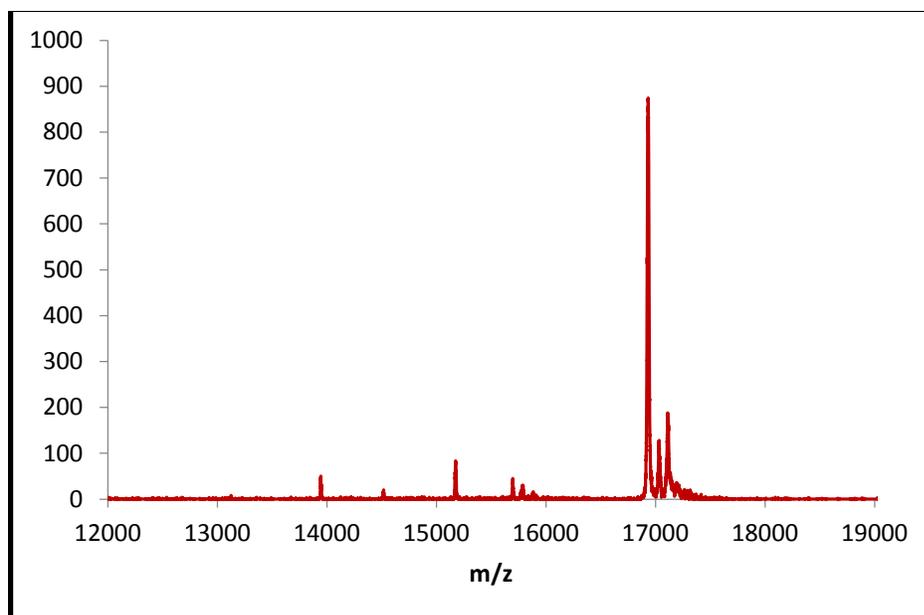


Figure 24. MALDI of S-PCP

Since L-PCP could not be detected by MALDI, only S-PCP was used in the loading assays. It was observed by MALDI, that Sfp could efficiently load CoA (signal at m/z: 17.351.5), but not the Hpg (theoretical mass: 167.16 daltons) onto the PCP domain (Figure 25). This suggests that the CoA-Hpg is hydrolysed in solution after loading onto S-PCP by Sfp.

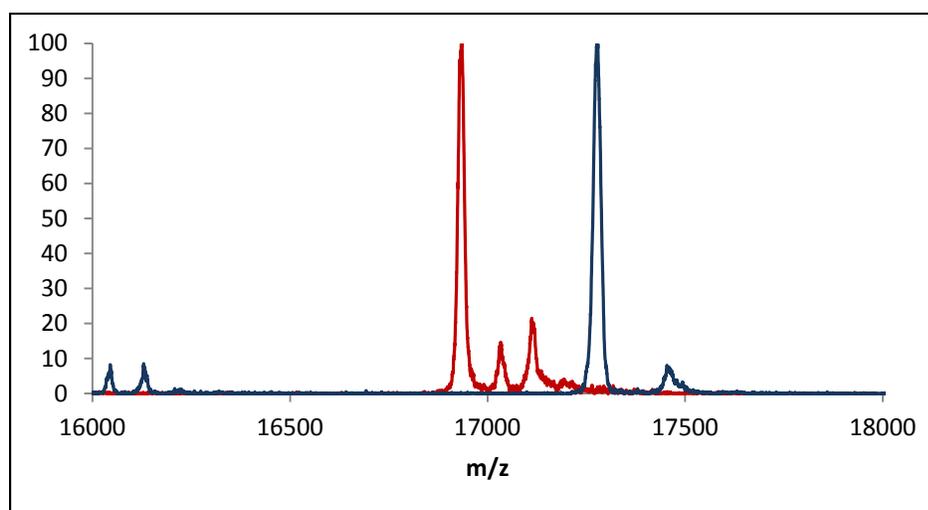


Figure 25. MALDI analysis of CoA-Hpg loading on S-PCP (Red-S-PCP, Blue-S-PCP loaded with CoA).

2.6. Ram20- bioinformatics, cloning and expression in *E. coli*

BLAST⁹² analysis of Ram20 suggests that the enzyme belongs to the tryptophan halogenase family (Table 1). The sequence analysis also revealed that the enzyme is likely to be a flavin dependent halogenase due to the presence of the conserved FAD-binding motif GxGxxG and the associated WxWxIP motif.

Table 1. BLAST⁹² analysis of Ram20 suggesting identical proteins

Enzyme	Identity (%)	Accession Number	Source
Tryptophan halogenase	66	WP 01935591	<i>Streptomyces</i> sp. AA1529
ComH	65	AAK81830	Complestatin biosynthetic gene cluster in <i>Streptomyces lavendulae</i>
Tcp21	65	CAG15020	Teicoplanin biosynthetic gene cluster in <i>Actinoplanes teichomyceticus</i>
StaI	64	AAM80532	<i>Streptomyces toyocaensis</i>
End30	64	ABD65950	Enduracidin biosynthetic gene cluster in <i>S. fungicidicus</i>
BhaA	63	CAA76550 63	Balhimycin biosynthetic gene cluster in <i>Amycolatopsis balhimycina</i>

A codon optimised *ram20* gene with desired flanking regions was designed and obtained in plasmid pMK-RQ (Geneart, Life technologies, Invitrogen, UK) (Figure 26).

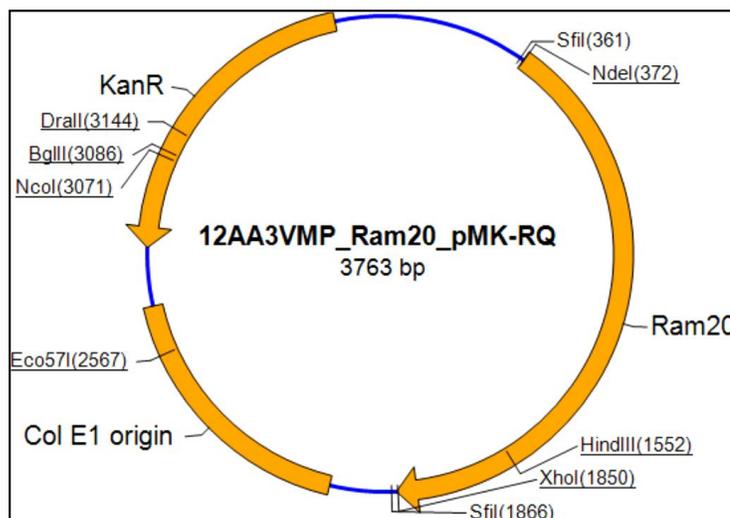


Figure 26. *ram20* gene in the pMK-RQ plasmid (construct map provided by GENEART)

The 1476 bp *ram20* gene (appendix) was subcloned into *NdeI/XhoI* restricted pET-28a(+) (pET-KK4) as described in sections 5.11-5.13. Clones were confirmed by restriction digestion (Figure 27) and verified by sequencing (GATC, Germany).

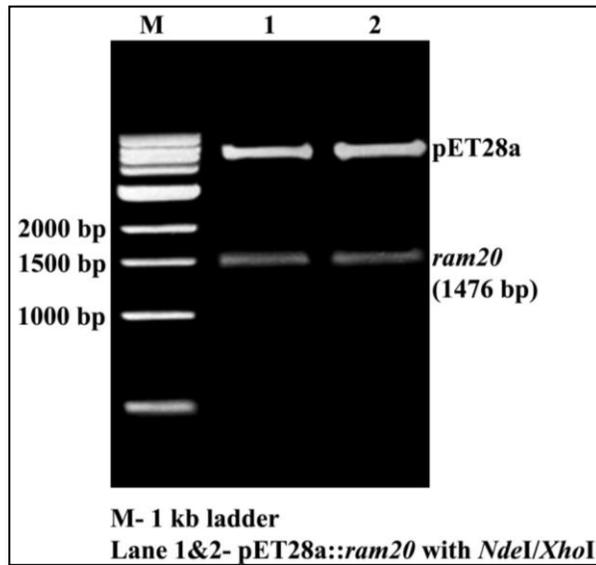


Figure 27. Restriction digestion analysis of pET-KK4 with *NdeI/XhoI*

The construct pET-KK4 was transformed into *E. coli* BL21 (DE3) cells following section 5.13. The protein was expressed on a large scale and purification was carried out using Ni^{2+} affinity chromatography following section 5.15. A band at the expected position for pure Ram20 (55.7 kDa) was observed by 12% (v/v) SDS-PAGE (Figure 28). 800 μl of Ram20 at 3.8 mg ml^{-1} was obtained.

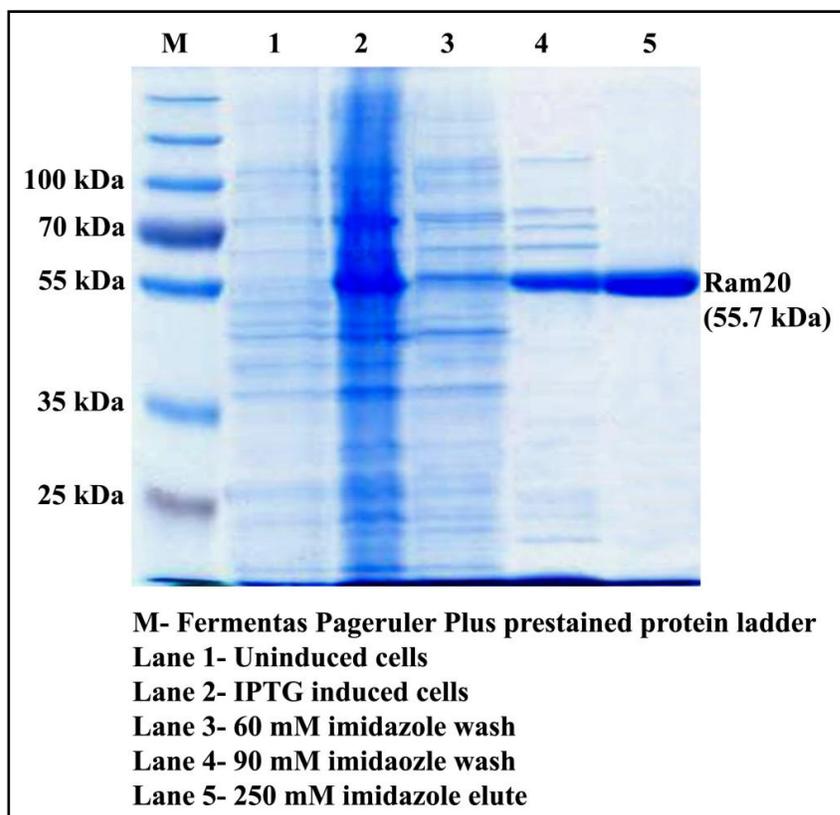


Figure 28. Analysis of pure Ram20 protein by 12% (v/v) SDS-PAGE.

2.7. Flavin reductases cloning and expression

The mechanism of flavin dependent halogenase was previously studied by the Karl-Heinz van Pee group⁹³. In the flavin dependent halogenases, reduced flavin reacts with molecular oxygen to form the flavin hydroperoxide. The enzyme bound flavin hydroperoxide reacts with a chloride ion to form hypochlorous acid. This occurs in the flavin binding pocket. The hypochlorous acid is passed through a tunnel where a conserved lysine residue interacts to form chloramine. Chloramine then reacts with the tryptophan residue, and a conserved glutamate residue deprotonates the intermediate leading to rearomatisation. Flavin reductases that catalyse the reduction of FAD to FADH₂ are usually associated with the biosynthetic gene cluster of the organism from which the halogenase originates from (eg. KtzS in *Kutzneria spp744*)⁷¹. Reduced flavin can also be provided by an alternative sources e.g. SsuE and Fre (flavin reductases) from

E. coli were cloned and expressed to be used in enzyme assays along with the halogenases described in this chapter.

2.7.1. Cloning and expression of *ssuE*

The *ssuE* gene sequence (576 bp) was amplified with primers P7 and P8 (Table 32) from *E. coli* BL21 DE3 gDNA and cloned into *NdeI/XhoI* restricted pET-28a(+) vector to express an N-terminal His₍₆₎-tagged SsuE (pET-KK5). The clone was verified by restriction digest of the plasmid (Figure 29) and sequencing (GATC, Germany).

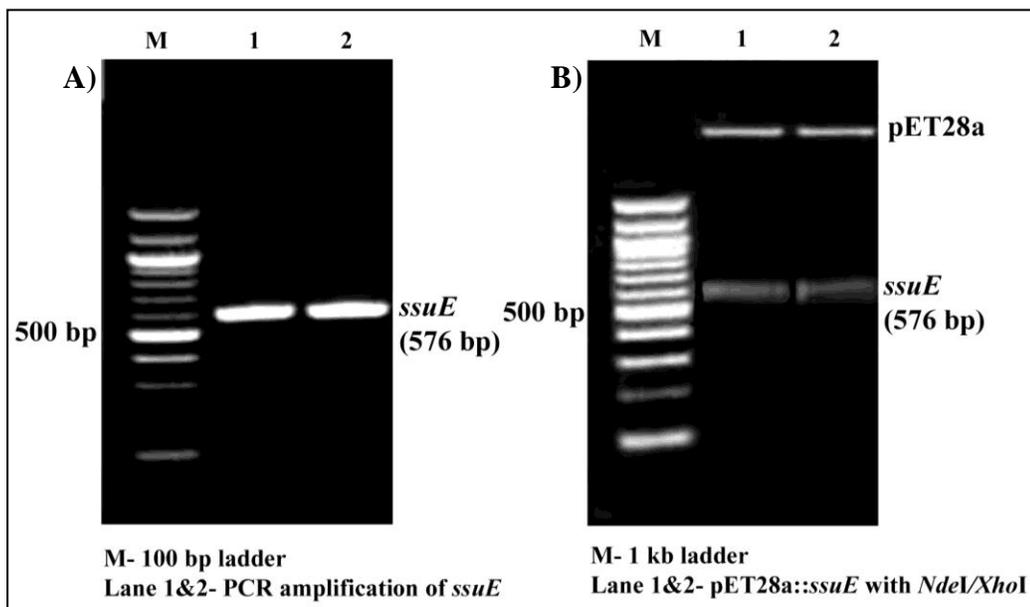


Figure 29. A) PCR amplification of 576 bp *ssuE* and B) Restriction digestion analysis of *ssuE* cloning

The construct pET-KK5 was transformed into *E. coli* BL21 (DE3) cells following (section 5.13). Expression tests were carried out on a small scale (section 5.15.1). Protein expression was observed at all temperatures tested. The highest level of expression was observed at 18 °C (data not shown). A large scale protein expression was carried at 18°C, and protein was purified using Ni²⁺ affinity chromatography as per section 5.15. A band at less than 25 kDa was observed consistent with the expected mass (23.6 kDa) of SsuE (Figure 30). 900 µl of pure SsuE at 2.2 mg ml⁻¹ was obtained.

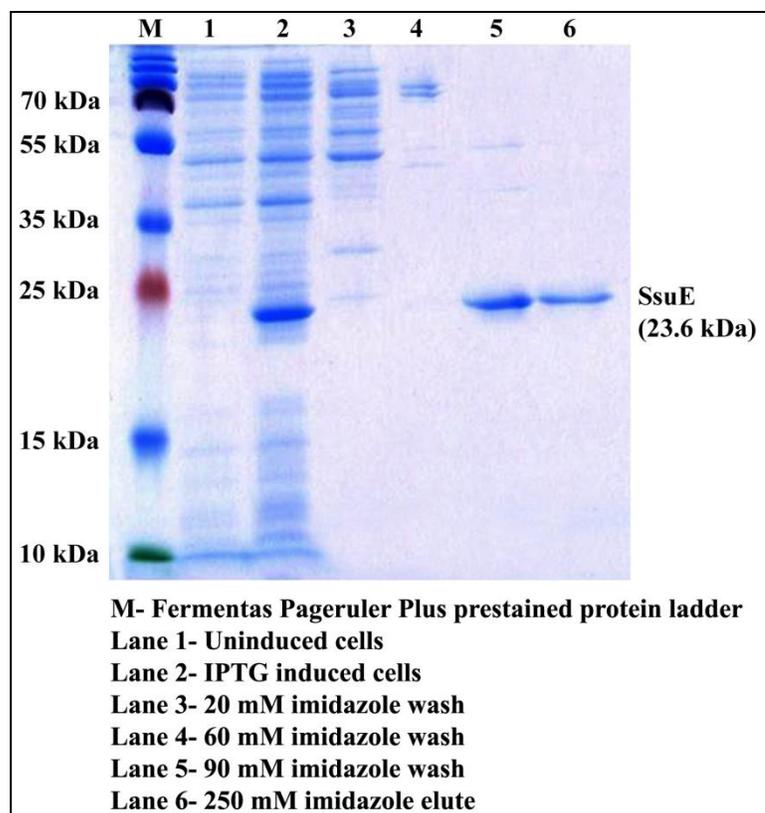


Figure 30. Analysis of pure SsuE protein by 20% (v/v) SDS-PAGE.

2.7.2. Determination of reductase activity of purified His₍₆₎-SsuE

SsuE catalyses the reduction of flavin mono-nucleotide (FMN) to FMNH₂ in the presence of the cofactor NADPH as per the reaction:



An enzymatic assay was carried out as per Gao and Ellis, 2005⁹⁴ (Section 5.21). The reduction of FMN to FMNH₂ was monitored by observing the decreasing absorbance of the reaction at 340 nm. A no enzyme control was also tested. The reaction was monitored for 20 min (Figure 31). A decrease in absorbance at 340 nm in the reaction mixture with no decrease in the control confirmed that the purified His₍₆₎-SsuE was enzymatically active.

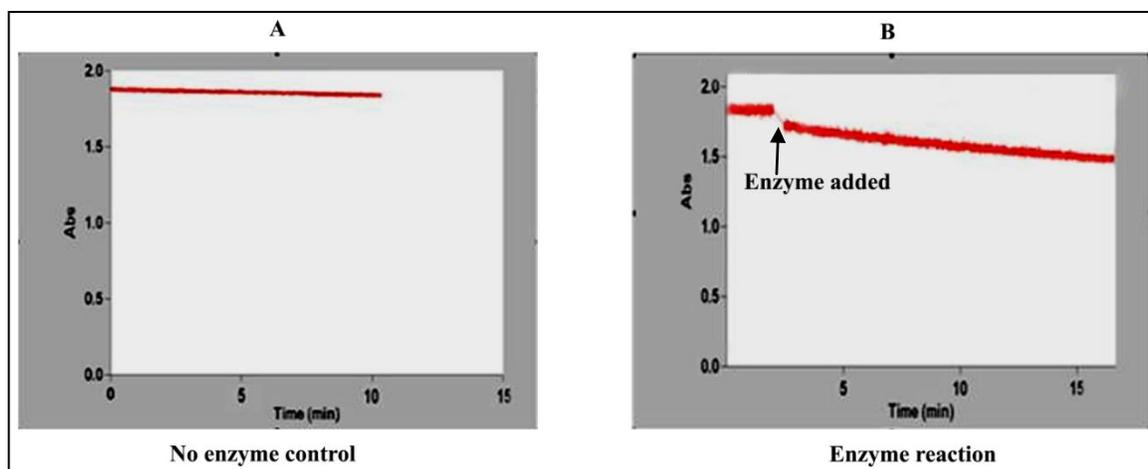


Figure 31. SsuE activity assay with FAD and NADPH (A). No enzyme control, (B). Enzyme reaction

2.7.3. Cloning and expression of flavin reductase Fre

The *fre* gene (702 bp) was amplified with the primers P9 and P10 (Table 32) from *E. coli* BL21 gDNA and was cloned into a *KpnI/XhoI* digested pET-45b(+) expression vector to express an N-terminal His₍₆₎-tagged Fre (pET-KK6). The clone was confirmed by restriction digest of the plasmid and sequencing (Figure 32).

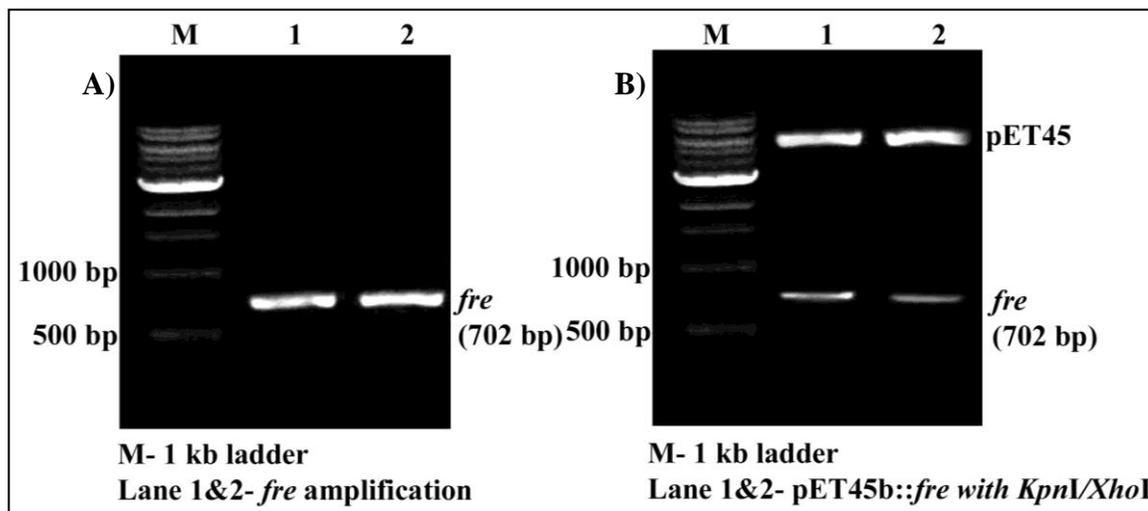


Figure 32. A. PCR amplification of 702 bp *fre* and B. Restriction digestion analysis of *fre* cloning

The construct pET-KK6 was transformed into *E. coli* BL21(DE3), expression was carried out at 30 °C on a large scale volume and protein was purified using Ni²⁺ affinity

chromatography, concentrated and stored as per section 5.15. 20 μl of pure concentrated Fre was analysed by 20% (v/v) SDS-PAGE. A band was observed between 25 kDa and 35 kDa consistent with the expected mass (28.5 kDa) of Fre (Figure 33). 1.2 ml of protein at 3.2 mg ml^{-1} was obtained.

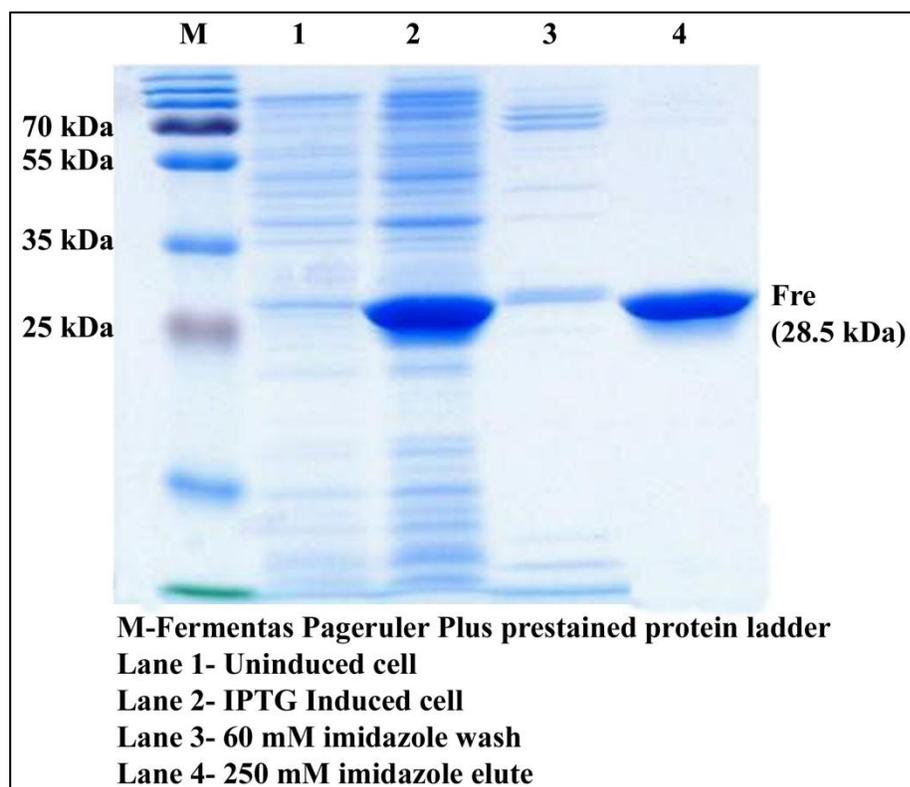


Figure 33. Analysis of pure Fre protein by 20% (v/v) SDS-PAGE.

2.8. Halogenation assay of Ram20 on Hpg, SNAC-Hpg and thioester Hpg

An assay of Ram20's ability to accept free substrates was carried out as mentioned in section 5.22. Literature precedence shows that Ram20 is active with PCP tethered substrates. However, no examples of Ram20 being active with free small molecules are known. The results of the HPLC assay were inconclusive due to co-elution of the Hpg substrates with FAD. The assay mixture was subjected to mass spectrometric analysis which ruled out any chlorination. Attempts towards characterising Ram20 to chlorinate Hpg from ramoplanin are continuing in the group.

In the following sections, we focus on our studies on tryptophan halogenases.

Chapter 3. Flavin dependent Halogenases for biocatalysis: result and discussion

This chapter will focus more on the tryptophan halogenases which accept free substrates particularly the flavin dependent halogenases. In recent years, these enzymes have received much focus due to their regioselective chlorination of substrates. However, only a few studies have been published on these enzymes. We are interested in characterising these enzymes with regards to their substrate scope and regioselectivity. Our studies will focus on five flavin dependent halogenases KtzQ, PrnA, KtzR, SttH and PyrH. These enzymes will be cloned and expressed in *E. coli* for the first time and will be screened for halogenation activity with aromatic and phenolic substrates. Regioselectivity and substrate scope of these enzymes will be improved by carrying out site directed mutagenesis. Structural studies of wild types and interesting mutant enzymes will be undertaken to study the basis for substrate specificity and regioselectivity.

3.1. KtzQ halogenase

KtzQ is a 60.93 kDa protein, first isolated from the kutzneride biosynthetic gene cluster of the soil actinomycetes *Kutzneria spp744* in the Walsh laboratory⁷⁸. They showed that KtzQ is an flavin dependent halogenase that regioselectively chlorinates at the 7-position of tryptophan. BLAST⁹² analysis revealed that this gene has high sequence identity with other flavin dependent halogenases particularly RebH and PrnA.

Table 2 .BLAST⁹² analysis of KtzQ showing identity with tryptophan halogenase

Protein Name	Organism	Accession Number	Identity
tryptophan 6-halogenase	<i>Streptomyces albogriseolus</i>	ABK79936	60%
RebH	<i>Lechevalieria aerocolonigenes</i>	2E4G	60%
PrnA	<i>Pseudomonas fluorescens</i>	AAD46365	53%

Few studies on the regioselectivity and substrate scope of KtzQ have been carried out. We are interested in exploring KtzQ's substrate scope and regioselectivity. Therefore,

KtzQ will be cloned and expressed in *E. coli* and tested for halogenase activity with aromatic and phenolic substrates.

3.1.1. Cloning and expression of KtzQ in *E. coli*

Genomic DNA (gDNA) was isolated from a 7 days old mycelium of *Kutzneria spp744* (Section 5.6). The *ktzQ* gene sequence (1602 bp) was amplified with primers P11 and P12 (Table 32) and cloned into *NdeI/XhoI* restricted pET-28a(+) and pET-30a(+) to express N- and C-terminal His₍₆₎-tagged KtzQ (pET-KK7 and pETKK8, respectively) Similarly, primers P13 and P14 were used amplify *ktzQ* and clone it into pET-45b(+) using the *KpnI/NotI* restriction sites to express an N-terminal His₍₆₎-tagged KtzQ (pET-KK9). The clones were verified by restriction digest (Figures 34 and 35) and sequencing.

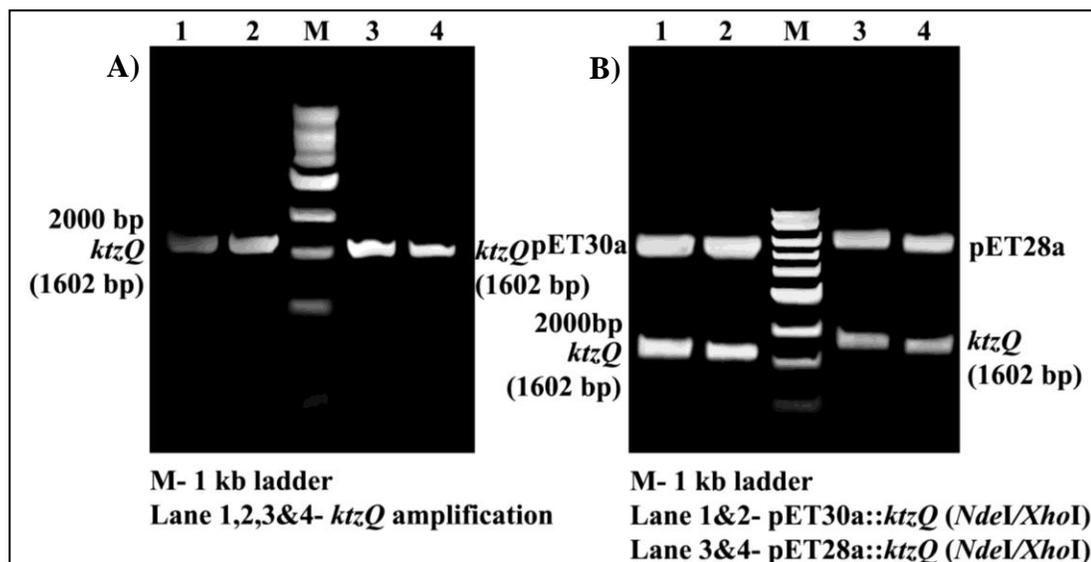


Figure 34. A) PCR amplification of 1602 bp *ktzQ* and B) Restriction digestion analysis of *ktzQ* cloning

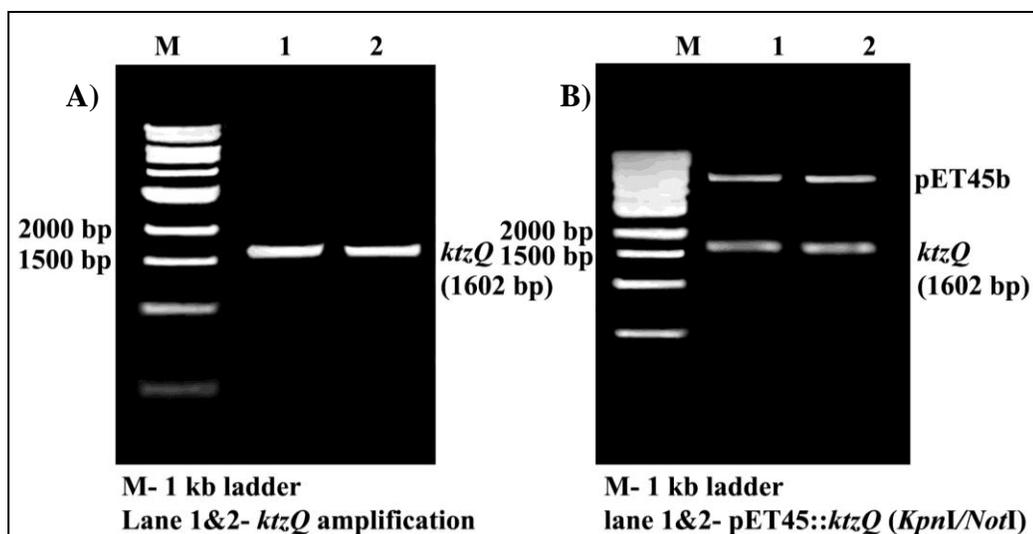


Figure 35. A) PCR amplification of 1602 *ktzQ* and B) Restriction digestion analysis of *ktzQ* cloning

KtzQ was previously expressed from a similar construct in *E. coli* BL21 in Christopher Walsh's group⁷⁸. We attempted to express *KtzQ* from pET-KK7 and pET-KK9 in *E. coli* BL21 (DE3) cells using similar expression conditions but no expression was observed. Expressions tests were then carried out at different temperatures (Section 5.15.1) and in different media including auto-induction medium, M9 medium and LB supplemented with tryptophan. No expression was observed under any of the above conditions. To check the toxicity of protein expression to the cells, growth of the cells was monitored by measuring at OD₆₀₀. An increase in absorbance indicated that protein expression is non-toxic. Due to the continued lack of expression of the protein, *ktzQ* was analysed for rare codons. The results showed the presence of 1 CGA (Arg), 8 CCC (Pro) and 2 AGG (Arg) *E. coli* rare codons that are used rarely in *E. coli* at frequencies of 0.3%, 0.4% and 0.2% respectively. Expression was therefore tested using pET-KK9 in *E. coli* Rosetta-gami (DE3) cells. This engineered strain provides additional copies of 7 tRNAs AGG, AGA, AUA, CUA, CCC and GGA and enhances disulfide bond formation. The results showed no expression under any of the conditions tested. The lack of expression could still be due to the presence of the rare CGA codon present at

the fifth position of sequence which since Rosetta-gami (DE3) cells do not supply an additional copy of this

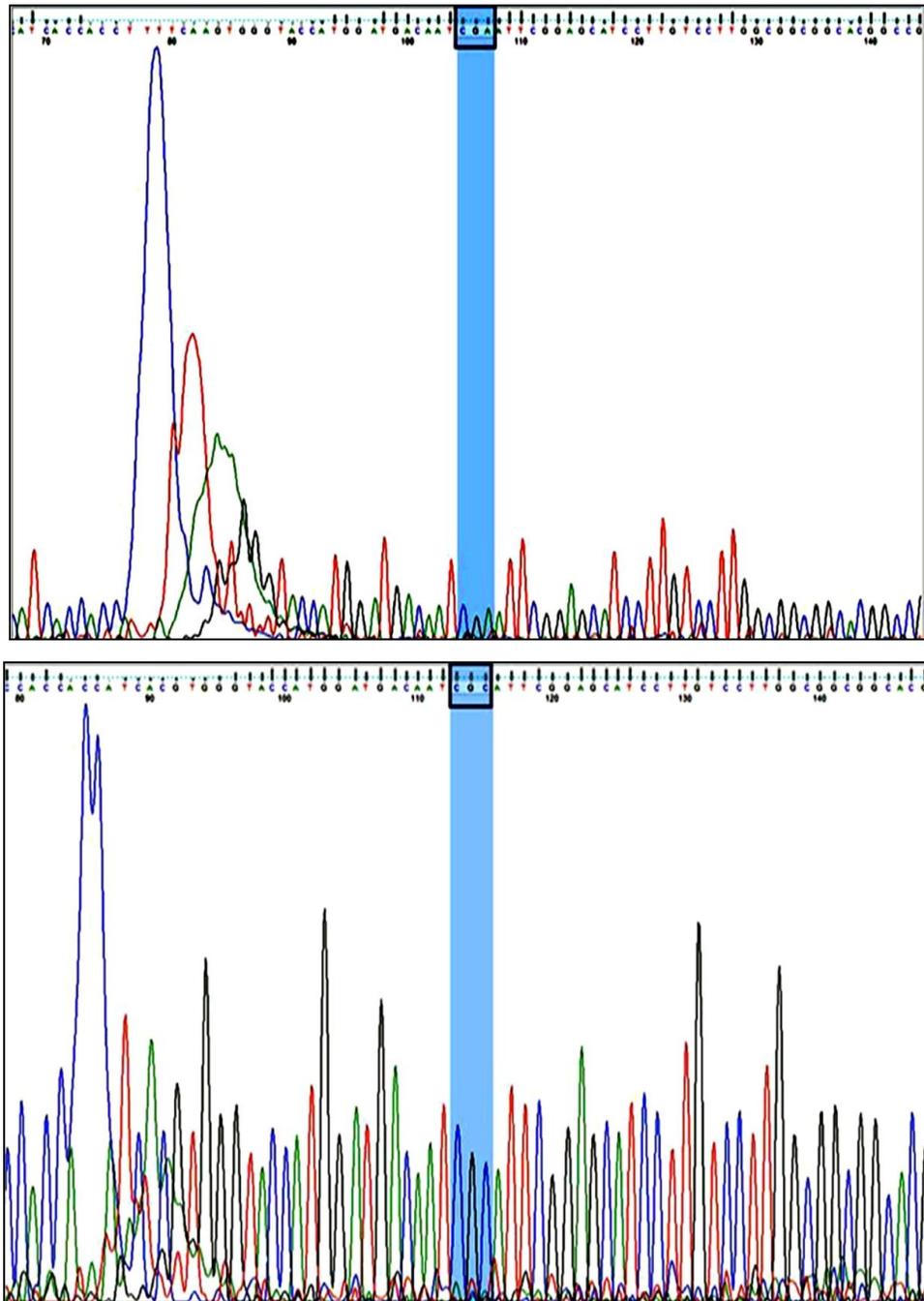


Figure 36. DNA sequencing chromatogram showing the mutation of CGA to CGC in *ktzQ*. Top- wild type *ktzQ* sequence with the CGA codon, bottom- *ktzQ* mutant sequence with the CGC codon. The codons are shown in boxes. Colours- Red-T, blue-C, green-A and black-G

tRNA. Hence a point mutation was introduced by site directed mutagenesis to change the rare CGA codon to CGC (Arg) to increase expression. The construct pET-KK9 was used as template DNA for site directed mutagenesis (section 5.14) using primer P15 (Table 32). The resultant mutant clone (pET-KK10) was confirmed by DNA sequencing (GATC, Germany) (Figure 36).

The mutant construct pET-KK10 was transformed into *E. coli* Rosetta 2(DE3) cells (section 5.13). Expression was tested following section 5.15.1. Protein expression was observed at all temperatures tested (Figure 37).

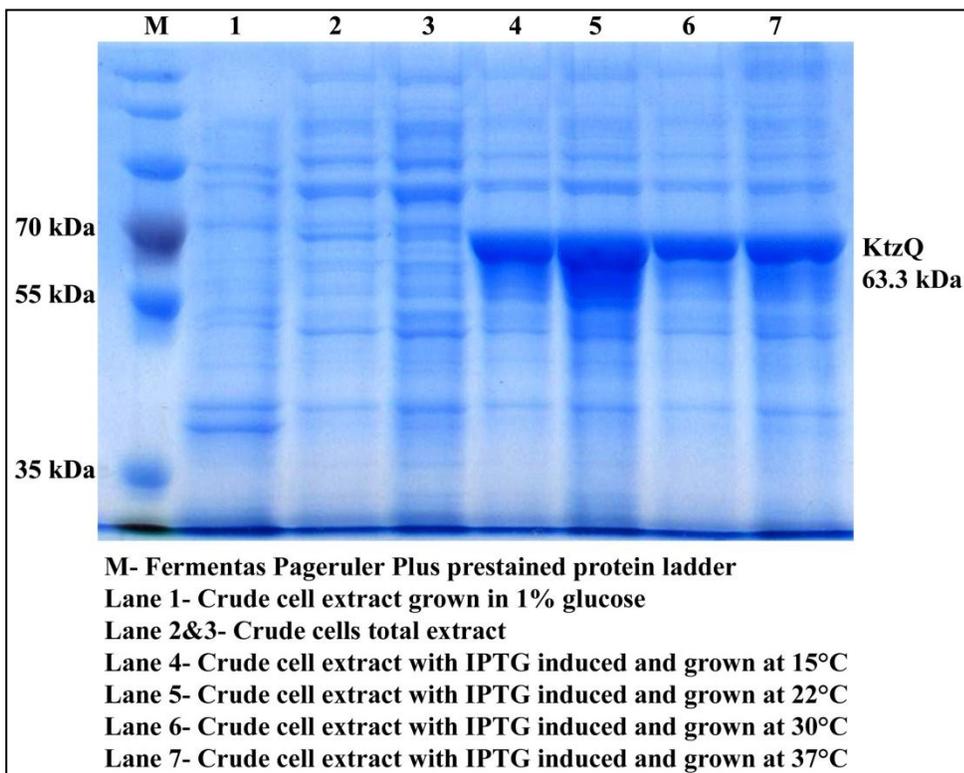


Figure 37. Expression studies of *KtzQ* protein using *E. coli* Rosetta 2 cells analysed by 12% (v/v) SDS-PAGE

Overproduction and a trial purification of *KtzQ* was carried out as per section 5.15. No *KtzQ* was obtained indicating it may form inclusion bodies or be insoluble. This was tested by repeating the experiment and loading the crude supernatant and crude pellet on the 12% (v/v) SDS-PAGE gel. The protein was observed in the cell pellet indicating that

the protein is insoluble. In order to try to obtain soluble protein, the mutant construct pET-KK10 was transformed into *E. coli* ArcticExpress (DE3) RP strain which expresses chaperones to aid protein folding and solubility.

The protein expression, isolation and purification was carried out as mentioned for wild-type KtzQ (section 5.15). 20 μ l of pure KtzQ was analysed by 12% (v/v) SDS-PAGE. A band between 55 kDa and 70 kDa was observed consistent with the expected mass (63.3 kDa) of KtzQ (Figure 38). 1 ml of pure KtzQ at a concentration of 2.2 mg ml⁻¹ was obtained.

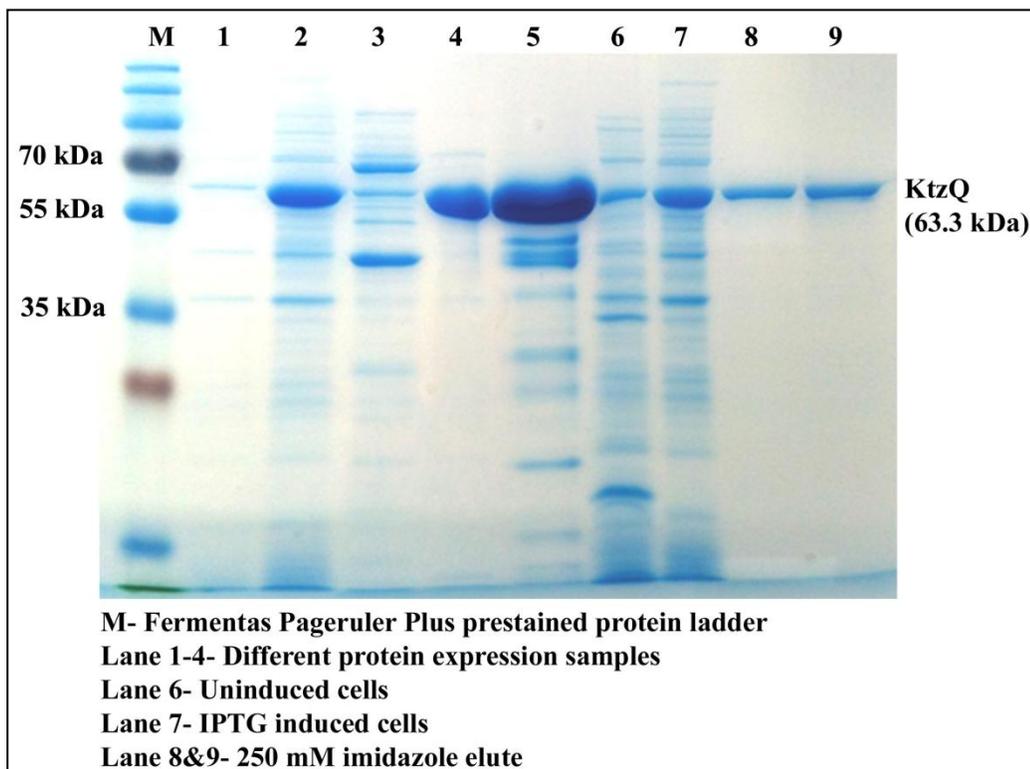


Figure 38: Analysis of pure KtzQ protein by 12% (v/v) SDS-PAGE

3.1.2. Identification of KtzQ by trypsin digestion: mass spectrometry

Mass spectrometry can be used to identify proteins⁹⁵. The method involves using a protease to digest the protein into peptides which can be analysed by matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) or liquid chromatography tandem mass spectrometry (LC-MS/MS). Protein is usually digested from the one or

two dimensional gels. The method involves cutting out the target protein band, removal of the stain, reduction and alkylation, digestion and finally extraction of the peptides for MS analysis. The mass of the peptides are measured by MALDI-MS to generate a peptide map which can be compared to theoretical peptide maps of known proteins. The peptides can also be subjected to MS/MS to determine the sequence of the peptides. Trypsin is the most commonly used enzyme for digestion of proteins. It is a serine protease that specifically cleaves protein at arginine and or lysine residues. Peptides terminating with these charged residues increase ionization leading to improved signal intensity of the peak in MS.

KtzQ was subjected to trypsin digest followed by MS and western blot analysis. The purified protein was subjected to a trypsin cleavage and the resulting sample was analysed by the LC-MS facility in MIB (section 5.18). A MASCOT data file was obtained containing a peak of MS/MS ions and fragment ions. This file can be used to search databases of known proteins to identify the protein. The MASCOT data obtained for KtzQ was run on web based software (http://www.matrixscience.com/search_form_select.html) using MASCOT MS/MS ion search. The results obtained showed a matching score of 86 for KtzQ indicating that the purified protein is KtzQ (Table 3). However, only a subset of the peptides expected based on KtzQ sequence were detected (Figure 39). Although their distribution suggests that the band observed by SDS-PAGE is full-length KtzQ.

Table 3. Scores obtained for matching proteins by MASCOT analysis

S.No	Protein Hits	Score
1	KtzQ [Kutzneria sp. 744]	86
2	KtzR [Kutzneria sp. 744]	59

1	MDDNRIRSI	VLGGGTAGWM	SACYLSKALG	PGVEVTVLEA	PSISRIRVGE
51	ATIPNLHKVF	FDFLGIAEDE	WMRECNASYK	AAVRVFNWRT	PGDQATPRR
101	RPDGRPDHFD	HLFGQLPEHE	NLPLSQYWAH	RRLNGLTDEP	FDRSCYVQPE
151	LLDRKLSPLR	MDGTLASYA	WHFDADLVAD	FLCRFAVQKL	NVTHVQDVFT
201	HADLDQRGHI	TAVNTESGRT	LAADLFIDCS	GFRSVLMGKV	MQEPFLDMSK
251	HLLNDRAVAL	MLPHDDEKVG	IEPYTSSLAM	RS GW SWKIPL	LGRFGSGYVY
301	SSQFTSQDEA	AEELCRMWDV	DPAEQTFNNV	RFRVGRSRRR	WVRNCVAIGV
351	SAMFVEPLES	TGLYFSYASL	YQLVKHFPDK	RFRPILADRF	NREVATMYDD
401	TRDFLQAHFS	LSPRDDSEFW	RACKELPFAD	GFAEKVEMYR	AGLPVELPVT
451	IDDGHYGNGF	EAEFRNFWTN	SNYCYFAGL	GFLPEHPLPV	LEFRPEAVDR
501	AEPVFAAVRR	RTEELVATAP	TMQAYLRLRH	QGT	

Figure 39. Trypsin cleaved proteins obtained by MASCOT analysis (Red-cleaved proteins)

Due to the poor sequence coverage by MS, further confirmation of the identity of the enzyme was attempted by western blot analysis. The western blot analysis of KtzQ enzyme was carried out using anti-His mouse monoclonal primary antibody and mouse hrp conjugate as the secondary antibody for detection by enhanced chemiluminescence (ECL) (section 5.19). The results showed specific binding of His antibody to a His₍₆₎-tagged protein. The protein band was found at the position expected based on the mass of KtzQ, which supports the MS data that the protein is His-tagged KtzQ. (Figure 40).

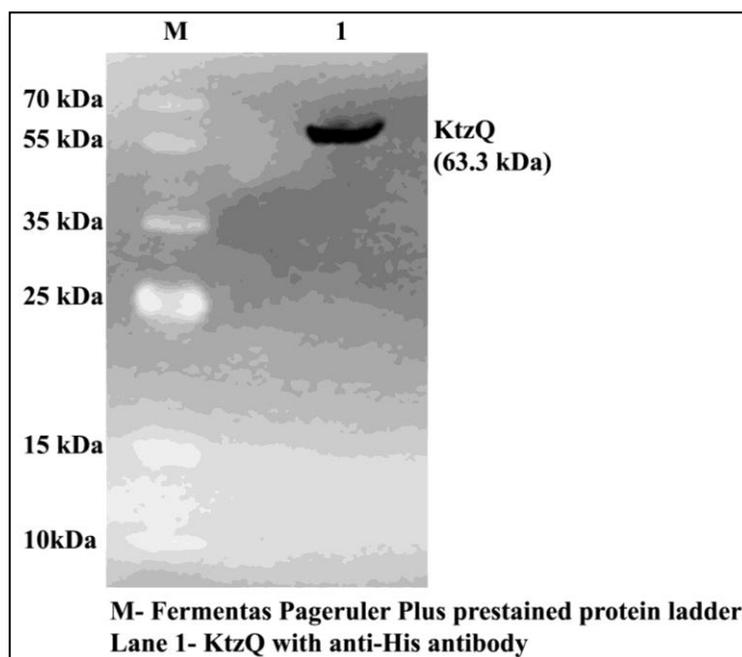


Figure 40. Western blot analysis showing the detection of KtzQ protein with anti His-antibody

3.1.3. Enzymatic assay of KtzQ with L-tryptophan

The activity of KtzQ was assayed in a halogenation reaction using tryptophan as the substrate. The assay was carried out as using 5 μ M of KtzQ, 1.5 μ M of Fre, 1 μ M FAD, 12.5 mM of MgCl₂ and 2.5 mM of NADH as described in section 5.22. The reaction mixtures and a tryptophan standard were separated by HPLC with detection at 280 nm. HPLC and MS analysis indicated that the protein is inactive as no chlorination or

bromination of the tryptophan was observed. This might be due to the mis-folding of KtzQ during expression in *E. coli*. Since we failed to detect halogenation in repeated trials with the protein, we decided to move on to other tryptophan halogenases. Further work on the purification and characterisation of KtzQ tryptophan halogenase was not carried out.

3.2. PrnA

Since the tryptophan 7-halogenase KtzQ did not show activity for chlorination, another tryptophan 7-halogenase, PrnA, was used as an alternate enzyme to study the regioselectively controlled chlorination at the 7-position and substrate specificity of tryptophan-7-halogenases. PrnA is a flavin dependent halogenase from *P. fluorescens* BL915 strain which can chlorinate tryptophan at the 7-position of the indole ring. The enzyme was isolated and characterized in Karl-Heinz van Pee's group^{64,89}. We cloned, expressed and purified PrnA to explore its substrate specificity and regioselectivity. PrnA was previously expressed in *P. fluorescens* BL915 using a broad range vector. *prnA* gene (appendix) cloned into a pPEH vector was obtained from Karl-Heinz van Pee's group. Sarah Shepherd from our group cloned the 1614 bp *prnA* gene into *NdeI/NotI* restricted pET-28a(+) (pET-SAS1) to express an N-terminal His₍₆₎-tagged PrnA protein. PrnA was expressed in *E. coli* BL21 (DE3) cells and was tested for halogenase activity with its natural substrate tryptophan.

3.2.1. PrnA mass spectrometry

PrnA enzyme was obtained from Sarah Shepherd and MS analysis was carried out following trypsin cleavage as described for KtzQ (section 5.18). The data was obtained in MASCOT format and was analysed in web based software (http://www.matrixscience.com/search_form_select.html) using MASCOT MS/MS ion search. The results obtained based on the ion intensity showed the highest matching score for PrnA (186) as expected (Table 4).

Table 4. Scores obtained for matching proteins by MASCOT analysis

S.No	Protein Hits	Score
1	Chain A, The Structure Of Tryptophan 7-Halogenase (PrnA) suggests A Mechanism For Regioselective Chlorination	186
2	tryptophan halogenase PrnA (<i>Burkholderia oklahomensis</i> EO147)	107

Further analysis of the enzyme by western blot analysis using anti His mouse monoclonal antibody confirmed the presence of a His₍₆₎-tagged protein of the expected mass of PrnA (Figure 41).

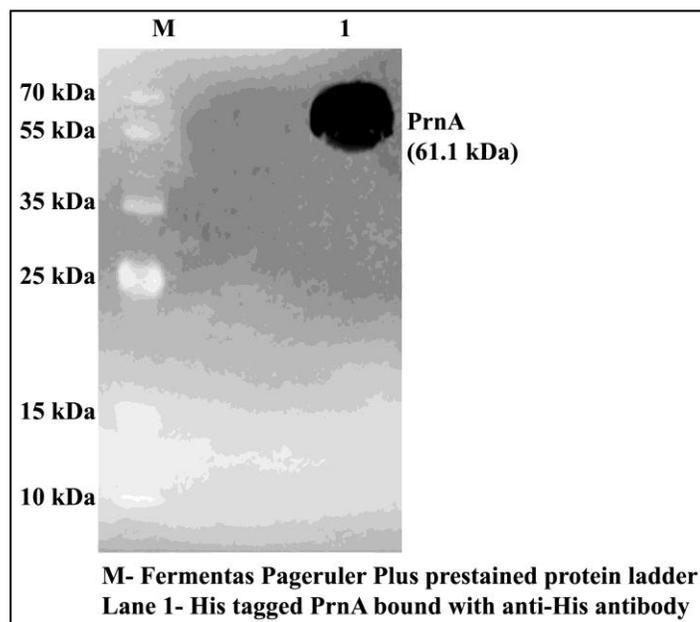


Figure 41. Western blot analysis showing the detection of PrnA protein bound with His-antibody

3.2.2. PrnA enzymatic assay for substrate scope

Sarah Shepherd expressed PrnA in *E. coli* ArcticExpress (DE3) RP and screened PrnA for activity with a series of substrates (Figure 80). PrnA was able to chlorinate and brominate tryptophan, kynurenine, anthranilamide, methyl anthranilate and 2-propyl aniline (Figure 42). The regioselectivity of the halogenation of products mediated by PrnA were analysed by NMR and all the substrate showed chlorination at the position similar to that of 7-Cl-tryptophan. Similar results were obtained for bromination.

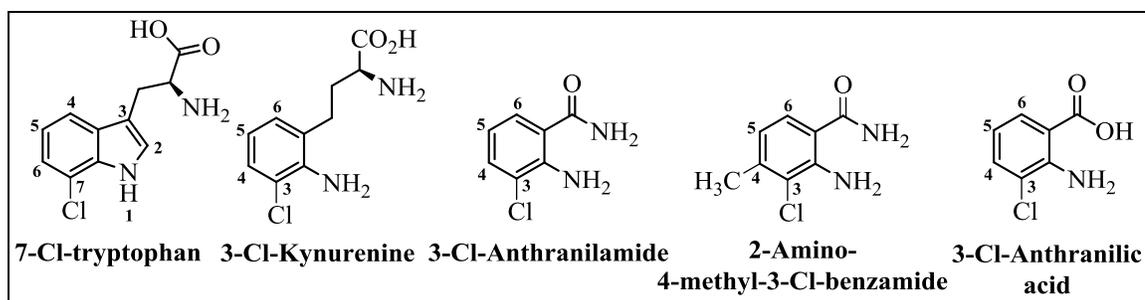


Figure 42. Substrates that showed chlorination by PrnA

In tryptophan, the first highly preferred chemical halogenation is at position 3. But since the position is occupied by extending amino acid chain, the next preference of halogenation will be at position 2, which is not halogenated in enzyme system. If position 3 is blocked by the position 2, the next favourable halogenation of the substrate by EAS will occur on indole ring of the tryptophan which is favoured by PrnA to chlorinate at the 7-position of tryptophan. In kynurenine, the presence of NH_2 group favours more as ortho, para directing group over the extended amino acid chain (also an ortho, para directing group). In chemical halogenation, either position 3 or 5 are preferred to carry out EAS. PrnA is more regio-selective to halogenate at the 3 position rather than producing a mixture of both 3 and 5 tryptophan halogenases. In anthranilamide, NH_2 group favours more ortho, para directing group over the CONH_2 group (meta directing group) preferring either 3 or 5 position for halogenation. The position 3 is ortho to para directing and meta to meta directing while position 5 is para to ortho, para directing group and meta to meta directing group. Chemical halogenation would produce mixture of both 3 and 5 halogenated anthranilamide on EAS but PrnA halogenates more selectively at the 3 position. Similar halogenation is observed for anthranilic acid as seen in anthranilamide. In 2-amino-4-methylbenzamide, position 3 is ortho to ortho, para directing groups (NH_2 and CH_3 respectively) and meta to meta directing group (CONH_2) and would therefore be much preferred position for halogenation. While position 5 will also be preferred because it is ortho to CH_3 (ortho, para directing group), para to NH_2 group (ortho, para directing group) and meta to

CONH₂ (meta directing group). PrnA halogenates more selectively at the 3 position neglecting the 5 position to be halogenated in the enzyme system.

3.2.3. Protein engineering to widen substrate scope

The crystal structure of PrnA (PDB ID: 2AQJ) has been solved previously to study the exact mechanism of chlorination⁶⁴. Karl-Heinz van Pee's group used the crystal structure to identify important residues governing the mechanism and regioselectivity of PrnA and confirmed their role(s) by site directed mutagenesis. The mutations generated by Karl-Heinz van Pee's group and their effects are summarised in Table 5.

Table 5. Previous mutations of PrnA and their effect on for regioselective chlorination

Mutant	Function of the residue and Reasons for mutagenesis	Results	Reference
E346	Required for abstraction of proton and stabilise wheland intermediate for chlorination	-	
E346Q	To increase the charge of Cl ⁻ molecule based on weak interaction with HOCl	Inactive	⁶⁴
E346D	To study the closer interaction with HOCl	Inactive	⁹⁶
S347	Drags HOCl into the tunnel and passes along K79	-	
S347A	Screening for activity	Less active	⁹⁶
K79	Amino group interacts with HOCl to form chloramine	-	
K79A	Screening for activity	Inactive	⁶⁴
H101, W455, F103	Forms sandwich around the indole ring of tryptophan	-	
H101A	Altering regioselective chlorination	No change	⁸⁹
W455A	Altering regioselective chlorination	No change	⁸⁹
F103A	Altering regioselective chlorination	Chlorinated both 7-&5-position of tryptopahn	⁸⁹

Although few mutants have been identified which govern substrate interaction and regioselectivity of halogenation, from a biocatalysis perspective it is imperative to widen the substrate scope for halogenation. Using the available crystal structure of PrnA we identified additional residues that interact with the substrate and could be involved in

determining substrate specificity. E450, Y433, Y444 and N459 which interact with the carboxyl group of the tryptophan (Figure 43) were selected for mutagenesis.

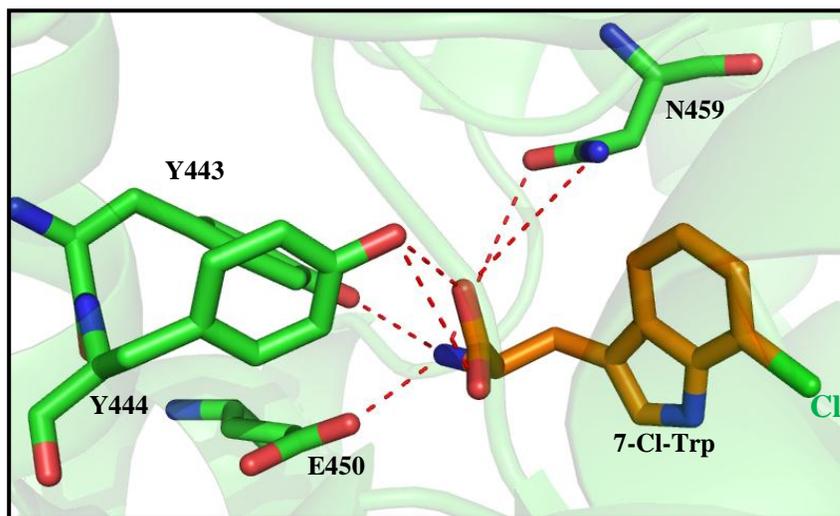


Figure 43. Crystal structure of PrnA (PDB ID:2AR8) generated using PYMOL¹ showing amino acid residues(green) interacting with chlorinated tryptophan(orange) in the active site

The mutants were designed on the basis of creating space in the active site using alanine, reducing the active site size using tryptophan, and changing the polarity at each of the positions by creating the mutations E450K, Y443H, Y444S and N459L. Sarah Shepherd produced many of the mutants for creating space and polarity however I generated the Y444W, N459W and E450W mutants that reduce the space available for substrate binding in the active site. The pET-SAS1 was used as template DNA for site directed mutagenesis (section 5.14) using primers P16-P18 (Table 32). Mutant clones (pET-KK 11-13) were confirmed by sequencing (GATC, Germany). Expression of Y444W, N459W, E450W and Y444S were carried out in *E. coli* ArcticExpress (DE3) RP cells at 15°C except for Y444S which was expressed at 12 °C to overcome solubility issues (Section 5.15.5). All of the mutant proteins were purified (Figure 44) and the Protein concentrations obtained for each of the mutants are tabulated in Table 6.

Table 6. *PrnA* mutant protein concentrations for widening substrates

Mutant proteins	Concentration (mg ml ⁻¹)	Concentrated volume (ml)
Y444W	3.2	1.0
N459W	2.4	1.0
E450W	2.2	1.0
Y444S	2.4	1.0

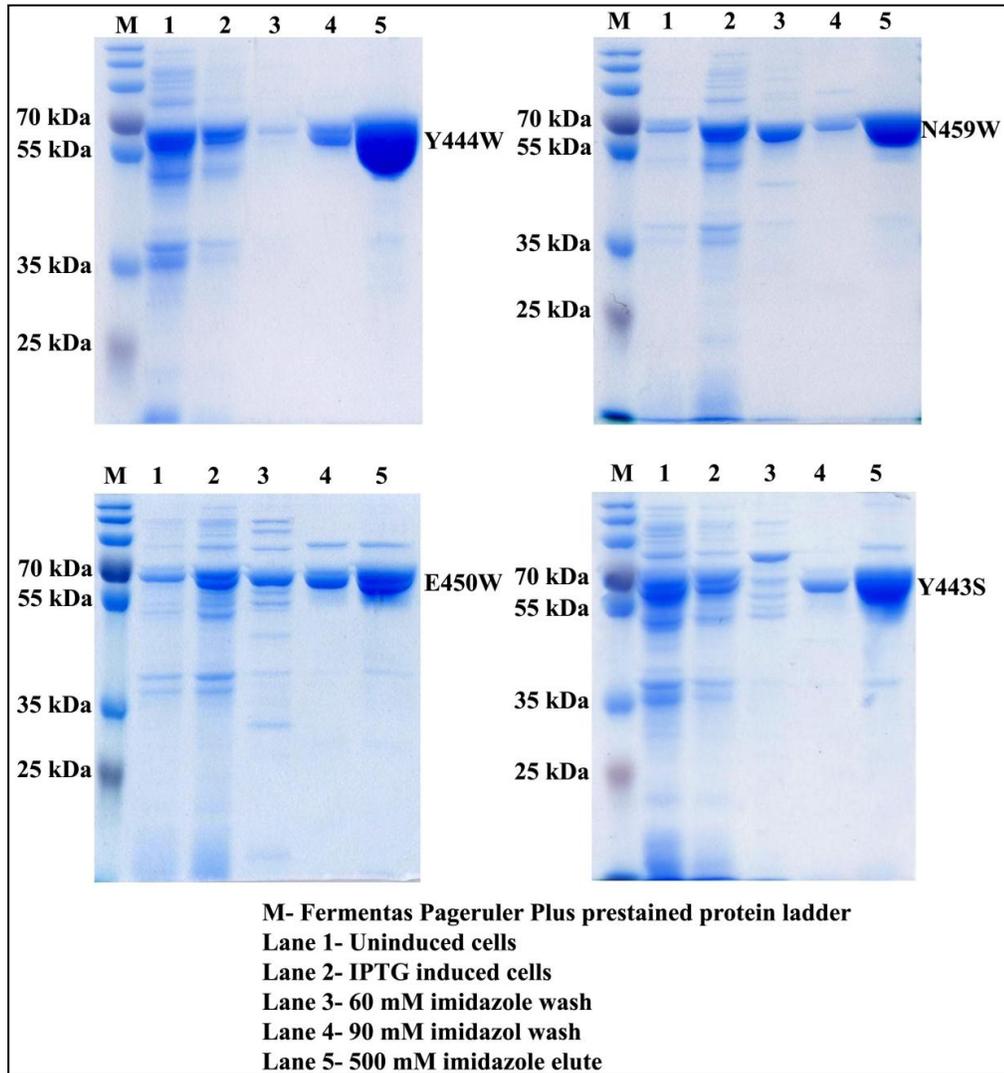


Figure 44. Analysis of purified mutants Y444W, N459W, E450W and Y443S by 12% (v/v) SDS-PAGE

3.2.4. Enzyme assay of mutant proteins with substrates

Sarah Shepherd screened the mutants for halogenase activity with library of substrates (Figure 80) and the results are tabulated in Table 7. All of the mutant proteins, except

Y444S and N459W which appeared to be inactive, chlorinated at least one substrate from tryptophan, kynurenine and anthranilamide. E450W and Y44W were unable to halogenate the natural tryptophan substrate but showed activity with the smaller substrate anthranilamide.

PrnA	Tryptophan		Kynurenine		Anthranilamide		Anthranilic acid
	Cl ⁻	Br	Cl ⁻	Br	Cl ⁻	Br	Cl ⁻
WT	Active	Active	Active	Active	Active	Active	Active
Y444W	Active	Active	Active	Active	Active	Active	Inactive
N459W	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive
E450W	Inactive	Inactive	Inactive	Inactive	Active	Active	Inactive
Y444S	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive	Inactive
E450K	Inactive	Inactive	Inactive	Inactive	Active	Active	Active

Table 7. Table of substrates tested against various PrnA mutants

The mutant E450K was the only mutant to retain activity with anthranilic acid as a substrate and showed higher levels of chlorination (18%) with anthranilic acid compared to that of wild type (6%).

To further check the effect of polar residues interacting with the substrate, the E450 residue was mutated to an R residue using pET-SAS1 as template and primer P19 (Table 32). The mutant clone (pET-KK14) was confirmed by DNA sequencing (GATC, Germany) Expression of E450R was carried out in *E. coli* ArcticExpress (DE3) RP cells (Section 5.15.3) and the protein was purified (Figure 45) and was concentrated to yield 1ml at a concentration of 3.2 mg ml⁻¹.

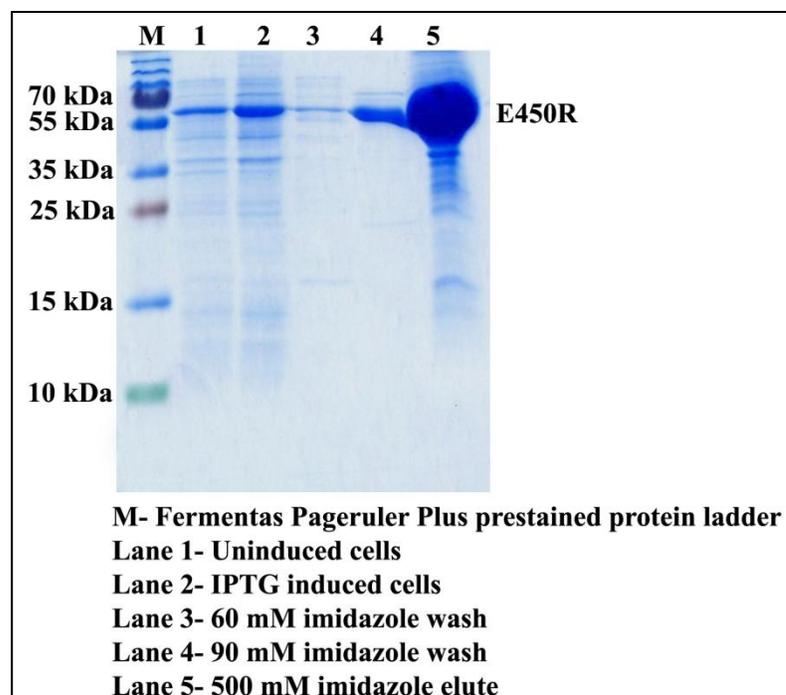


Figure 45. Analysis of E450R mutant protein by 12% (v/v) SDS-PAGE

In an assay with anthranilic acid as the substrate, mutant E450R showed a higher level of chlorination relative to that achieved by the wild type enzyme, but lower levels of chlorination relative to that of the E450K mutant. The positive charge of the lysine residue predicted to strongly interact with the carboxyl terminal of the anthranilic acid resulting in much higher activity of E450K mutant. The positive charge of the arginine residue is more diffuse and may not interact with the anthranilic acid to the same extent as lysine. In order to facilitate higher chlorination levels of anthranilic acid, Sarah Shepherd designed further mutants based on in-silico docking of the anthranilic acid into the crystal structure of PrnA (PDB ID:2AR8). In the crystal structure the carboxyl group of F454 hydrogen bonds with the amino group of tryptophan at a distance of 2.78 Å and Y443 and Y444 interact with the carboxyl of tryptophan. Since the E450K and E450R mutants showed higher level of chlorination compared to that of wild type, mutagenesis was used to change the other tryptophan interacting residues Y443, Y444 and F454 into lysine and arginine residues.

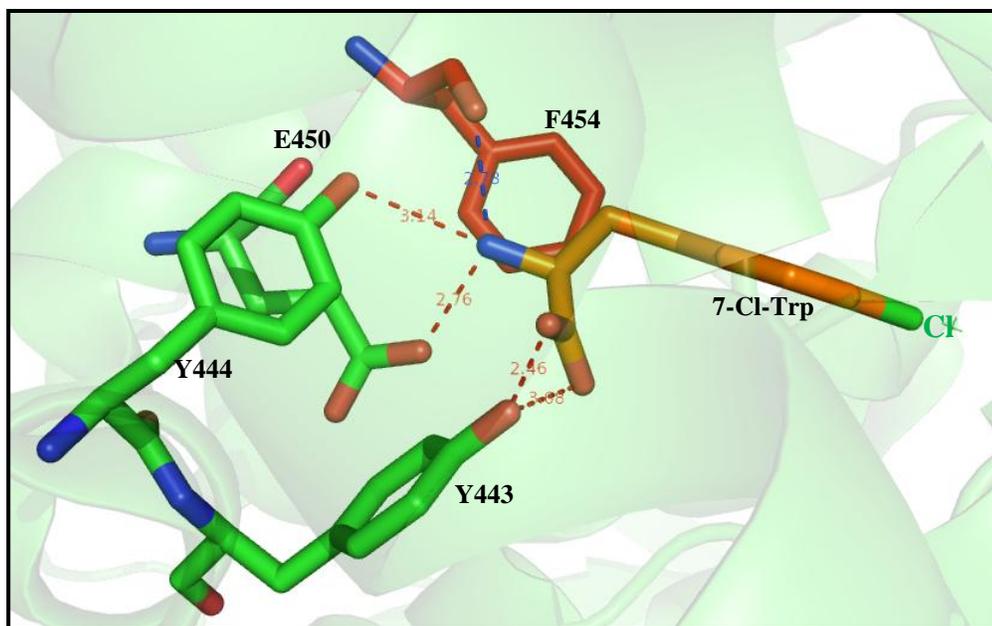


Figure 46. Crystal structure of PrnA(PDB ID:2AR8) generated using PYMOL¹ showing active site F454 (red) and other residues (green) interacting with chlorinated tryptophan

The construct pET-SAS1 was used as template DNA for site directed mutagenesis (section 5.14) using primers P20-P25 (Table 32). Mutants F454K, F454R, Y443K, Y443R, Y444K and Y444R (pETKK15-20) were confirmed by sequencing (GATC, Germany). Expression of the PrnA mutants (pETKK15-20) was carried out in *E. coli* ArcticExpress (DE3) RP cells (Section 5.15.5). Proteins were purified and were concentrated to the following concentrations in a total volume of 1 ml (Table 8).

Table 8. PrnA mutant protein concentrations for anthranilic acid.

Mutant Proteins	Concentration (mg ml ⁻¹)
F454K	7.3
F454R	6.8
Y443K	3.6
Y443R	3.1
Y444K	2.75
Y444R	2.6

3.2.5. Assay of PrnA mutant's with anthranilic acid

The assays with anthranilic acid as the substrate were carried out by Sarah Shepherd and she found that the mutant F454K produced the highest level of chlorination (55%) followed by F454R with 41% and E450K with 35%. The rest of the mutants produced chlorination levels of less than 20%.

3.2.6. Optimisation of protein purification of PrnA for crystallographic studies

So far two crystal structure of PrnA have been resolved: wild type PrnA (PDB ID: 2AR8)⁶⁴ and a PrnA E346D mutant (PDB ID: 2JKC)⁹⁶. Both of these structures were obtained to understand the mechanism involved in the regioselective chlorination of tryptophan. The proteins were previously expressed using *P. fluorescens* BL915⁹⁶. The section will focus on the optimisation of the purification protocol of PrnA expressed in *E. coli*, which will also be used to produce important PrnA mutants for crystallographic studies. Since PrnA wild type had activity on substrates other than tryptophan, our initial focus was to obtain co-crystal structures of PrnA with these substrates. In order to carry out crystallographic studies, large amounts of very pure protein is required. Optimisation of the purification protocol was important to obtain the purest PrnA possible. To express protein for crystal trials, pET-SAS1 was transformed into *E. coli* BL21 (DE3) and *E. coli* ArcticExpress (DE3) RP cells.

PrnA was expressed in *E. coli* BL21 (DE3 strain) at 30 °C and purified as described in (section 5.15). 20 µl of the purified sample was analysed by 12% (v/v) SDS-PAGE and showed the presence of contaminants and degradation of PrnA indicating it that it would be unsuitable for crystal trials. In order to try to reduce the amount of contaminating proteins in the expression system, the cells were grown at the lower temperature of 15 °C. Most naturally occurring proteins in *E. coli* are expressed between 20-35 °C, while at lower temperatures their production becomes minimal. This unfortunately resulted in a similar level of impurities but did reduce the level of protein degradation (Figure 47).

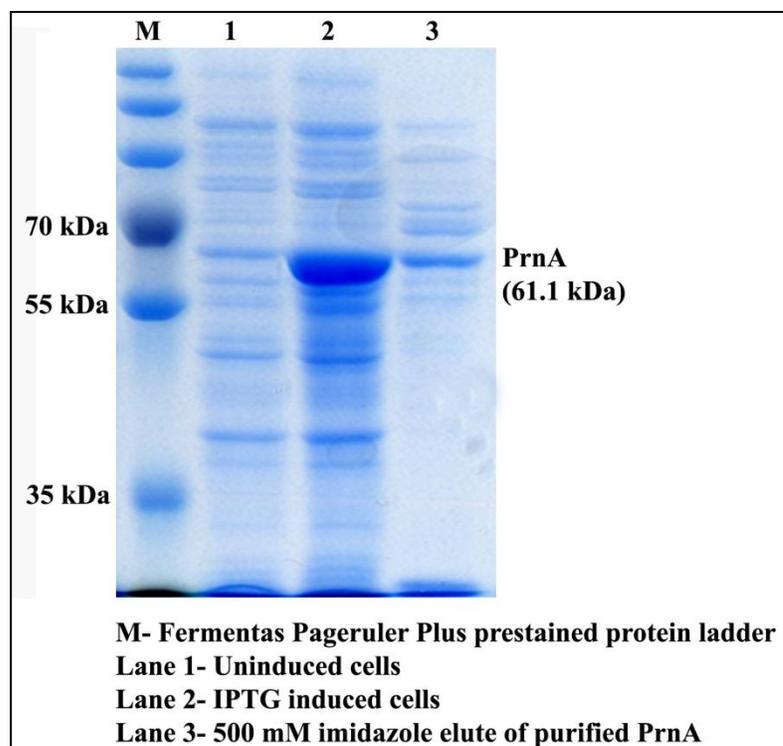


Figure 47. Analysis of purified PrnA from *E. coli* BL21 DE3 cells by 12% (v/v) SDS-PAGE

Following the improvement with low temperature expression, the expression was carried out in the *E. coli* ArcticExpress (DE3) RP strain since it can adapt to very low temperatures (10-15 °C). The cells were grown, harvested, lysed and purified according to section 5.15. The purified protein was analysed by 12% (v/v) SDS-PAGE. The results revealed much purer protein compared to that purified from BL21 (DE3) cells, as both the level of contaminants and degradation was minimal. However, ArcticExpress cells are engineered to produce two cold tolerant chaperones Cpn10 and Cpn 60 that can increase the solubility of expressed recombinant proteins. These chaperones, particularly the Cpn60 (60 kDa) can remain bound to the protein of interest and co-elute from the nickel column^{97,98}. The expressed proteins were difficult to separate from chaperones by gel filtration due to their closely associated sizes. Hence an alternate approach of using ATP to remove these chaperone was adopted^{97,98}.

To remove the chaperone from the PrnA, different methodologies were adopted. Three types of experiments were performed initially to remove the chaperones using ATP. The treatment of ATP provides conformational change for complete removal of protein from chaperones in the system^{99,100}. Three different samples were prepared and a final concentration of 10 mM ATP was added to them at different points of the purification protocol. To one of the tubes, the ATP was added before the cells were lysed with lysozyme (10 mg ml⁻¹). To the second tube, ATP solution was added to the supernatant obtained after lysis and centrifugation. These two tubes were incubated for 2 h at room temperature. To the third tube, ATP solution was added to the supernatant and incubated overnight. Protein from the three tubes was purified separately by FPLC (section 5.15.6) and the purified samples were analysed by 12% (v/v) SDS-PAGE (section 5.16).

The amount of protein obtained when ATP was added prior to lysis was low, but a significant amount of protein was obtained when the ATP was added after lysis and incubated for 2 hours or overnight. Furthermore, the treatment with ATP lead to a reduction in the amount of chaperone co-purifying with PrnA without affecting the yield of PrnA (Figure 48).

Since the ATP treatment showed promise for removal of chaperones from PrnA a different approach was taken that involved using an ATP wash (Section 5.15.6.3). The first fraction of the ATP buffer wash contained both PrnA and the chaperone and the second fraction contained only PrnA. However, the final elution with 500 mM imidazole contained both

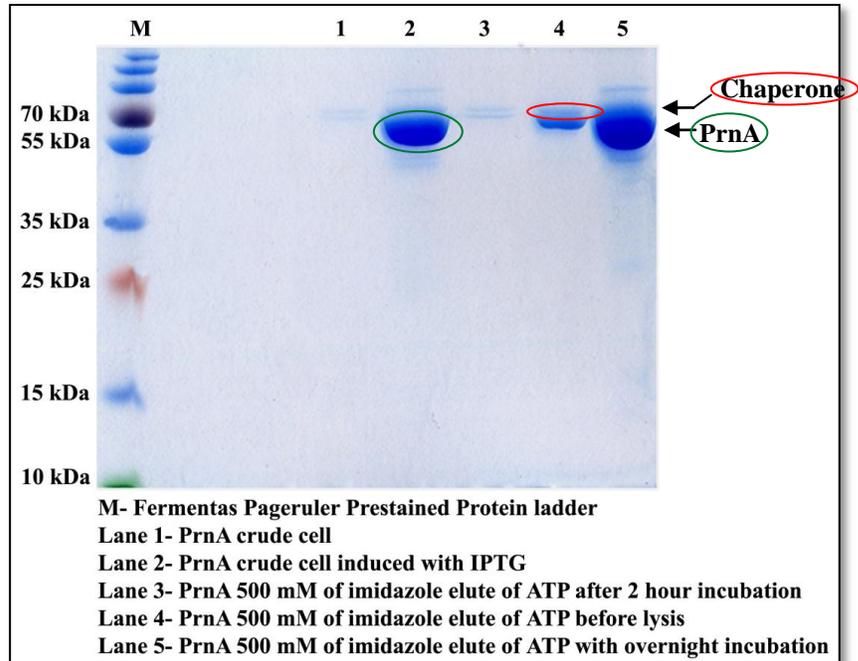


Figure 48. Analysis of PrnA treated with ATP by 12% (v/v) SDS-PAGE

PrnA and chaperone in very low amount (Figure 49). This results indicated that an ATP buffer wash can be used to remove the chaperone from PrnA.

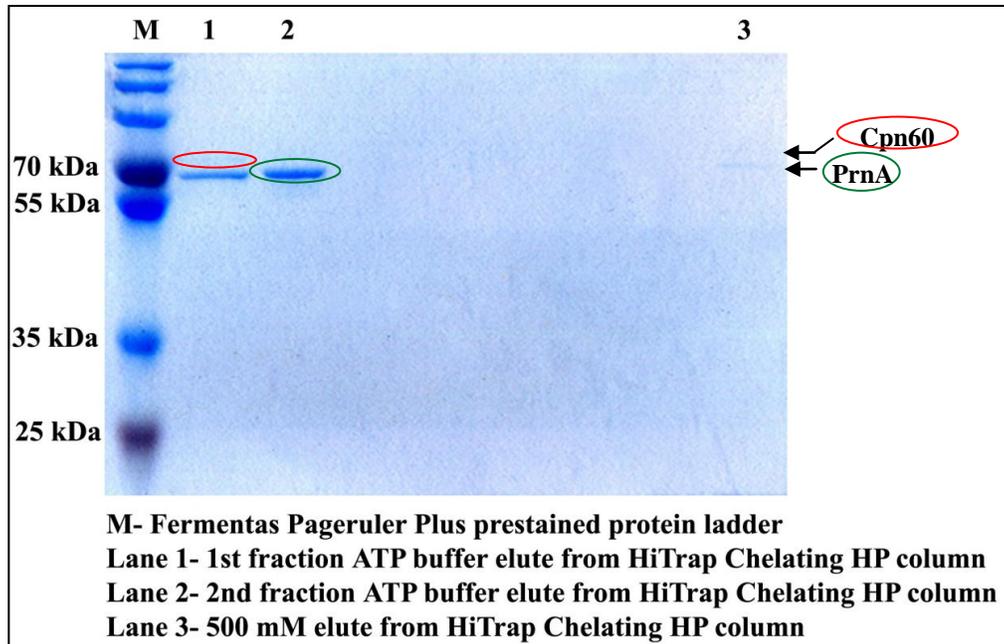


Figure 49. Analysis of PrnA treated with an ATP wash by 12% (v/v) SDS-PAGE

The expression was scaled up and purification was carried out in 3-4 runs as described above. The fraction in which the protein was expected to be pure was collected, pooled, concentrated and further purified on a Superdex™ 200 10/300 gel filtration column. The eluted fractions from the gel-filtration column were pooled together and concentrated to 12 mg ml⁻¹. The protein was analysed by 12% (v/v) SDS-PAGE and unfortunately there was evidence of a small amount of degradation (Figure 50). Due to this degradation, PrnA was not subjected to crystal trials.

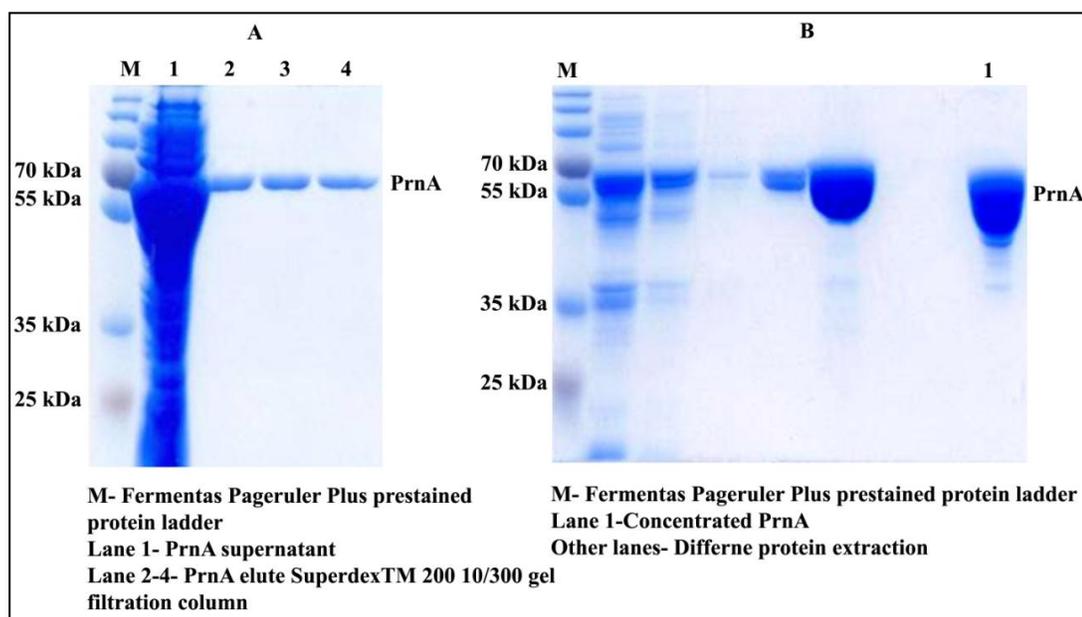


Figure 50. Analysis by 12% (v/v) SDS-PAGE
(A). Purified PrnA following the ATP wash showing removal of the chaperones (B).
Purified PrnA following gel-filtration and concentration

The desired amount of protein required for crystal trials was difficult to obtain using the above method due to the loss of protein with the ATP buffer wash. Hence an alternate method was developed using an engineered vector pET-YSBLIC which was obtained from the Grogan group from York. The plasmid is a pET-28a(+) vector engineered with a HRV 3C protease cleavage site replacing the thrombin cleavage site between the His-tag and the coding sequence of the gene of interest¹⁰¹. HRV 3C protease from human rhinovirus type 14 recognises the cleavage site LeuGluValLeuPheGln/GlyPro and is highly active at 4 °C. It is available as a 22 kDa His-tagged fusion recombinant protein

which allows it to be easily removed from its substrate. *prnA* was subcloned from pET-SAS1 into *NdeI/XhoI* restricted pET-YSBLIC (pET-KK21). Clones were confirmed by restriction digestion and further verified by sequencing (GATC, Germany). pET-KK21 was transformed into *E. coli* ArcticExpress (DE3) RP. The protein was expressed, the cells lysed and the lysate concentrated. A digestion was set up at a ratio of 1 unit: 100 µg HRV3C: PrnA lysate and incubated at 4°C for 48 hr. The protease cleaves the His₍₆₎-tag from the protein and since the HRV 3C protease is His-tagged protein, the cleaved His₍₆₎-tag from PrnA and HRV3C should bind to the Ni²⁺ column, hopefully along with the chaperone, while the tag free PrnA should flow straight through the column. This should remove the chaperones from the protein. The purified protein was analysed by 15% (v/v) SDS-PAGE.

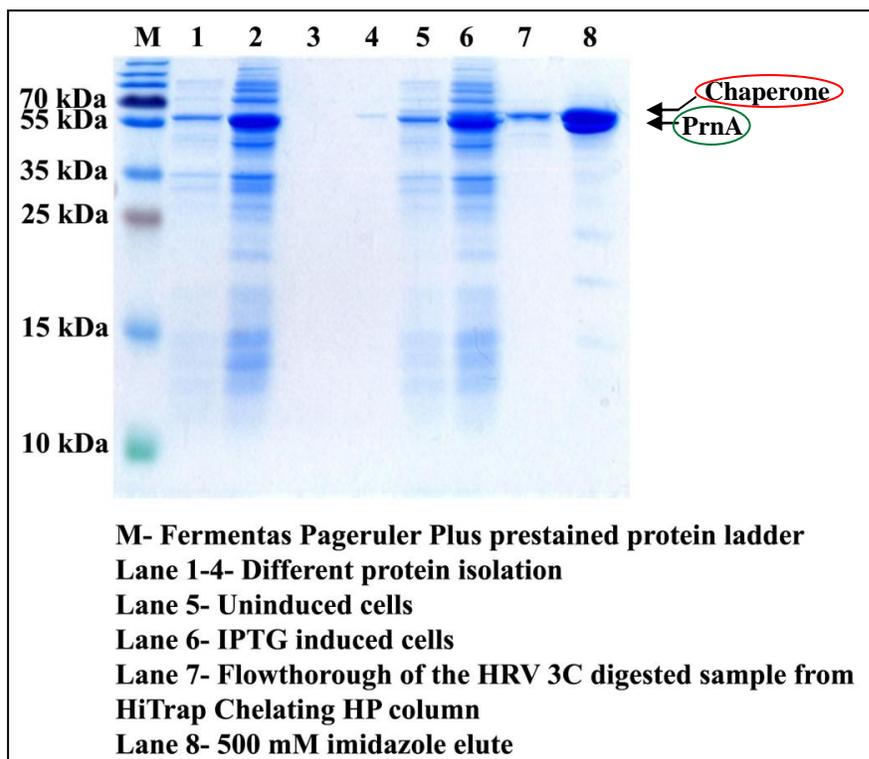


Figure 51. Analysis of HRV 3C cleaved PrnA by 12% (v/v) SDS-PAGE for removing chaperones

The presence of chaperone was observed in the flow through with no PrnA being eluted, while 500 mM elute showed both PrnA and the chaperone. This result indicated that the

digestion of HRV3C protease did not efficiently cleave the His₍₆₎-tag after 48 hr allowing the His₍₆₎-tagged PrnA to bind the nickel column strongly that was eluted by 500 mM imidazole concentration. Later a protocol was optimised using hydrophobic columns to remove the chaperones from PrnA (Section 5.15.6.6). Using this method wild-type PrnA and, the PrnA mutants F454K and E450K were purified. The pure PrnA wild type was subjected to MALS to check the monodispersity of the sample, but due to the low yield it was not submitted for crystal trials. A high yield of the PrnA mutants F454K and E450K was obtained and these proteins were subjected to crystal trials.

3.2.7. Multiangle light scattering (MALS) of PrnA

The pure PrnA wild-type was subjected to multiangle light scattering (MALS) to check for its monodispersity before submitting it for crystal trials. To obtain crystals, the purified protein should be homogenous. MALS is a very sensitive technique based on the scattering of light through proteins in solutions that is used to characterise proteins in solution¹⁰². This technique is used to determine the molecular mass, oligomeric state, protein folding and dispersity. The MALS analysis of PrnA showed that the enzyme preparation is mono-disperse and has a molecular mass of 114.8 kDa consistent with the protein existing as a dimer in solution as has been reported⁶⁴.

3.2.8. Crystal structure of PrnA F454K mutant

All of the crystallographic experiments were performed by Dr. Colin Levy at Manchester Institute of Biotechnology, Manchester, UK. The purified PrnA F454K mutant (34 mg ml⁻¹) was incubated with 2.5 mM of FAD for 1 hr and subjected to crystal trials using a 96 well screen from Molecular Dimensions. Crystals were obtained under similar crystallisation conditions to that reported for the wild-type PrnA enzyme (Section 5.28). A modified crystallisation condition that contained PEG 20K, 0.1M MES pH 6.5 as the precipitant and a 1.1 sitting drop experiment at room temperature was carried out to obtain X-ray diffraction quality crystals. The crystals formed were tetragonal in shape and diffracted to 2.4 Å. Attempts to get co-crystal structure for PrnAF454K mutant were

unsuccessful as the protein failed to crystallise in the presence of either tryptophan or anthranilic acid. Clear electron density was visible for FAD and Cl⁻. The mutant crystal structure was derived using PHASER¹⁰³ software using wild-type PrnA as the structural template. The co-ordinate and structural parameters for derived PrnA F454K structure are shown below.

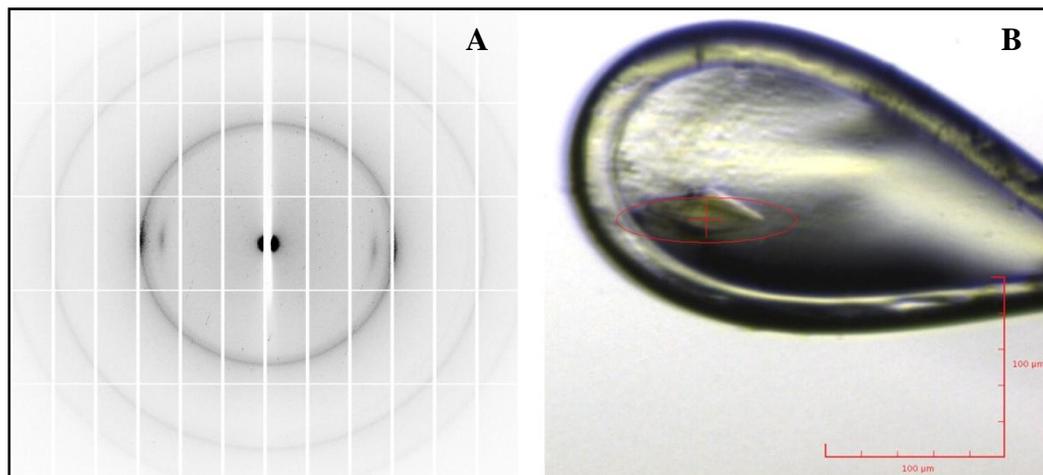


Figure 52. (A). Diffraction image, (B). PrnA F454K mutant crystal in tetragonal shape

Table 9. Structural parameters obtained for PrnA F454K mutant using CCP4 suite software¹⁰⁴

Data collection	Substrate FAD and Cl ⁻
Wavelength (Å)	
Resolution range (Å)	68.84-2.204 2.283-2.204
Space group	P 43 21 2
Unit cell	68.1 68.1 275.38 90 90 90
Unique reflections	33920 (3298)
Completeness (%)	9.96 (99.82)
Mean I/sigma(I)	8.31 (3.80)
Wilson B-factor	31.37
R-work	0.1795 (0.2145)
R-free	0.2325 (0.2937)
Number of atoms	4561
macromolecules	4158
ligands	59
water	344
RMS(bonds)	0.008

RMS(angles)	1.13
Ramachandran favored (%)	97
Ramachandran outliers (%)	0.19
Clashscore	2.41
Average B-factor	26.00
macromolecules	25.70
ligands	24.60
solvent	30.10

The crystal structure of the PrnA F454K mutant contained a bound FAD molecule and a chloride ion. The wild type PrnA (PDB ID: 2AR8) structure was superimposed with the PrnA F454K mutant structure using PYMOL software¹. The mutant structure was found to be in the same overall conformation to that of the wild type with a triangular pyramid shape (Figure 53a). The FAD and Cl⁻ binding sites are also similar to that of the wild type (Figure 53b).

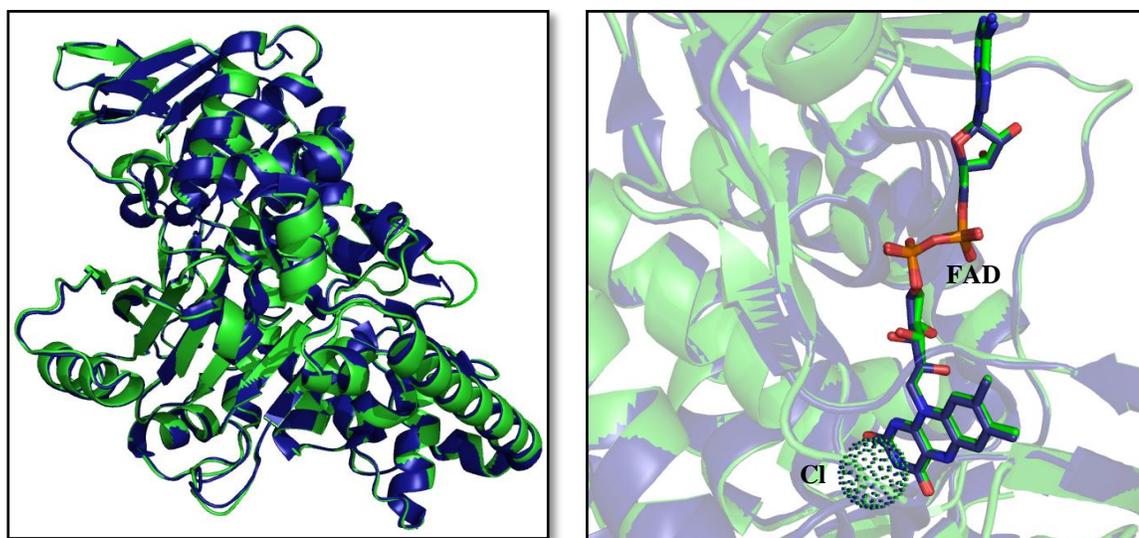


Figure 53. (A).Overlay of PrnA wild type(PDB ID:2AR8) (blue) with PrnA F454K mutant (green) showing similar structures, (B).Similar binding of FAD and Cl⁻ by PrnA (blue) and the F454K mutant (green), generated using PYMOL¹

The Cl⁻ is bound on the face of the isoalloxazine ring of FAD making contact with T348 and G349⁶⁴. The K79 residue that interacts with hypochlorous acid to form the chloramine is also found to be present at the same position as in the wild type. The main difference lies on the position of amino acids from 437-443, which form an extended

loop of an α -helix in the wild-type structure. This loop in the F454K mutant does not align with the α -helix present in wild-type PrnA (Figure 54).

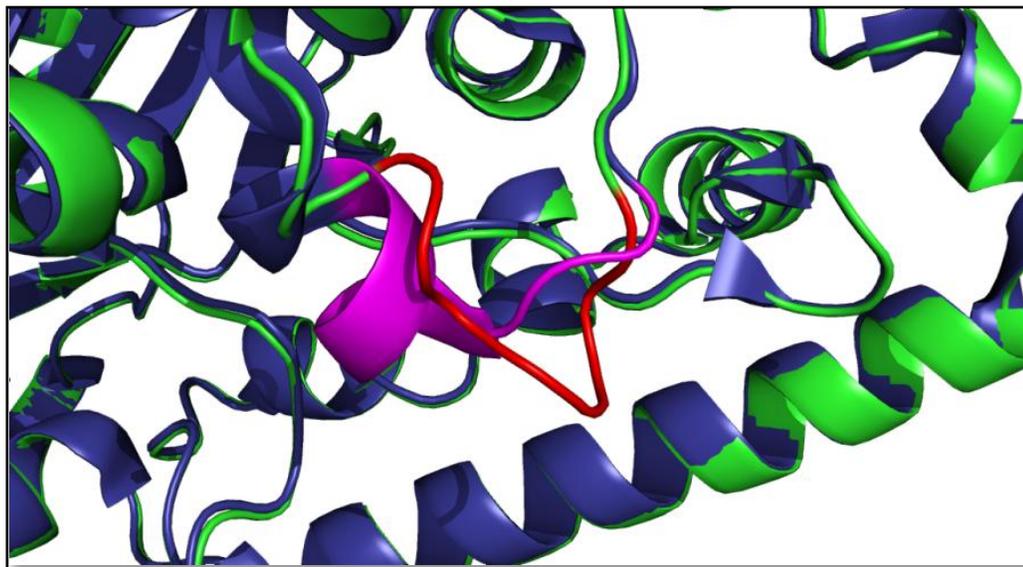


Figure 54. Crystal structure of PrnA wild type(PDB ID:2AR8) (blue) showing the α -helix (magenta) and F454 mutant (green) showing the non-aligning extended loop (red, generated using PYMOL¹)

The α -helix holds the key residue Y444, whose hydroxyl group hydrogen bonds with the carboxyl group of the tryptophan⁶⁴, in position. In the F454K mutant, the residue is pushed away from the active site by 2.3 Å- 2.79 Å such that the hydrogen bonding to the substrate would be disrupted. The Y443 residue also moves away by 0.69 Å. The hydroxyl group of Y443 hydrogen bonds with the amino group of tryptophan in the wild type and this interaction could be impaired (Figure 55). Together, these movements allow more space for the smaller substrates to bind nearer to the lysine residue allowing for a better electrostatic interaction.

Although the attempts to co-crystallise F454K mutant with substrates L-tryptophan and anthranilic acid were unsuccessful. The F454K mutant structure was used as a template to model anthranilic acid using the PYMOL software¹. The docking was carried out by aligning anthranilic acid to the tryptophan in the wild-type structure. The orientation of anthranilic acid at the active site could explain its regioselective chlorination which is aided by the residue E346.

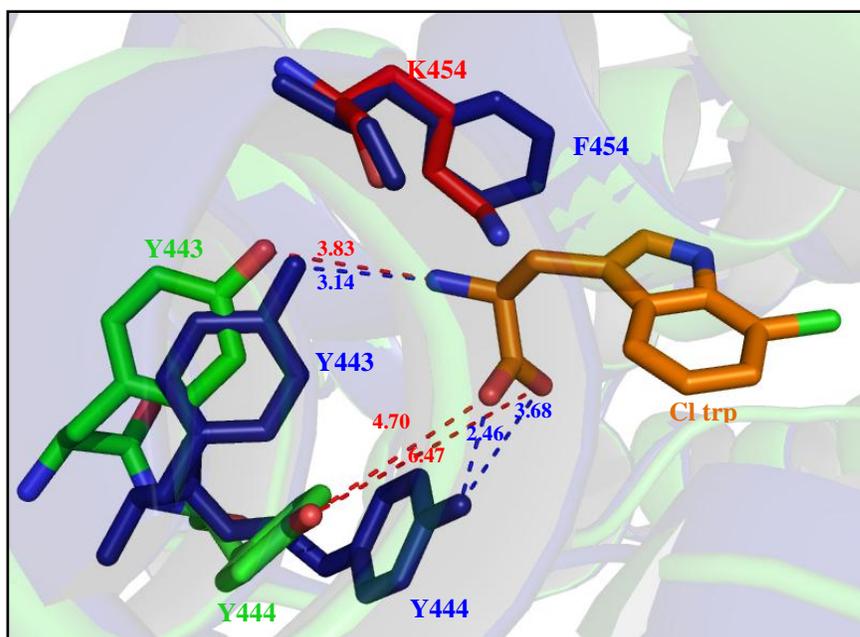


Figure 55. The distance differences of the interaction between PrnA F454K mutant (green) residues Y444 and Y443 and the wild type (PDB ID:2AR8) residues (blue) with tryptophan, the K454 mutation is shown in red and chlorinated tryptophan is orange (generated using PYMOL¹).

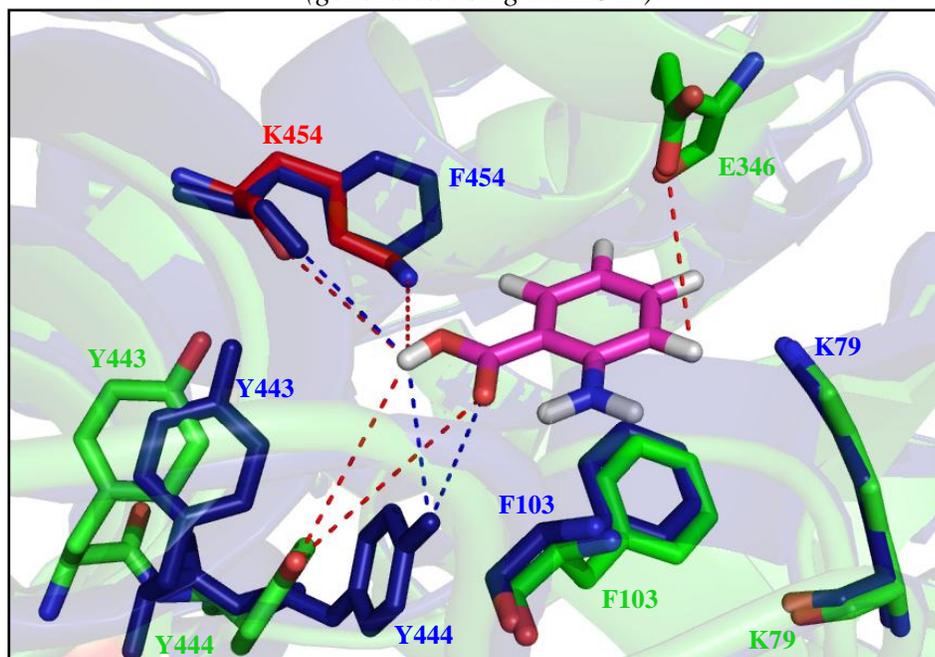


Figure 56. Docking of anthranilic acid in the F454K mutant (green) structure showing interaction with Y444, Y443 and K454 (red), wild type PrnA (PDB ID:2AR8) (blue) used as template, anthranilic acid (magenta), generated using PYMOL¹

The docking experiment shows that the Y443 and Y444 residues are extended outwards allowing more space for the anthranilic acid binding closer to the K454 residue. The previous docking studies suggest that the electrostatic interaction takes place between the hydroxyl group of the anthranilic acid and the amino group of the K454 residue. It would also allow the amino group of the residue to hydrogen bond with the hydroxyl group of the substrate. The strong interaction would allow much higher conversion of anthranilic acid by the mutant. The same interaction in the wild-type enzyme would be weaker leading to a lower conversion of anthranilic acid. This proposed binding model is in agreement with our experimental results. At present efforts are directed towards getting co-crystal structures of the mutant bound to substrates.

3.3. KtzR (tryptophan-6-halogenase)

KtzR was first isolated from the kutzneride biosynthetic gene cluster of soil actinomycetes *Kutzneria spp744*⁷⁸ and was shown to be a flavin dependent halogenase that chlorinates at the 6-position of tryptophan. We cloned, expressed and purified KtzR to study its halogenation mechanism, regioselectivity and substrate scope.

3.3.1. Cloning and expression of KtzR

The microorganism was provided from the Walsh laboratory⁷⁸. Initially *ktzR* was amplified from *Kutzneria spp744* gDNA and cloned into a pET vector for expression in *E. coli* by Sarah Shepherd in our laboratory. Additionally, the gene was also cloned into a pVLT33 vector (pET-SAS3) for expression in *Pseudomonas putida* KT2440 as carried out by Christopher Walsh group⁷⁸. Attempts were made to transform pET-SAS3 into *Pseudomonas putida* KT2440 by chemical transformation and electroporation. Both of these methods were unsuccessful, presumably due to the larger size (10 kb) of the plasmid. The transfer of the plasmid is usually carried out by conjugation, wherein an additional helper strain of *E. coli* S17-1 λ pir is used but this was unavailable in the lab. Due to time constraints in obtaining the helper strain, further attempts at transforming the plasmid were not carried out. Expressions was tested in *E. coli* BL21 (DE3) and

Rosetta-gami (DE3) as per section 5.15.1. Expression studies were also performed using auto-induction medium, M9 medium and LB supplemented with tryptophan or 1% glucose. Various concentrations of IPTG for induction were also investigated. The results revealed no expression in either of the strains under any of the conditions tested. A codon optimised *ktzR* gene with desired flanking regions was designed and obtained in plasmid pMK-RQ (Geneart, Life Technologies, Invitrogen, UK) (Figure 57).

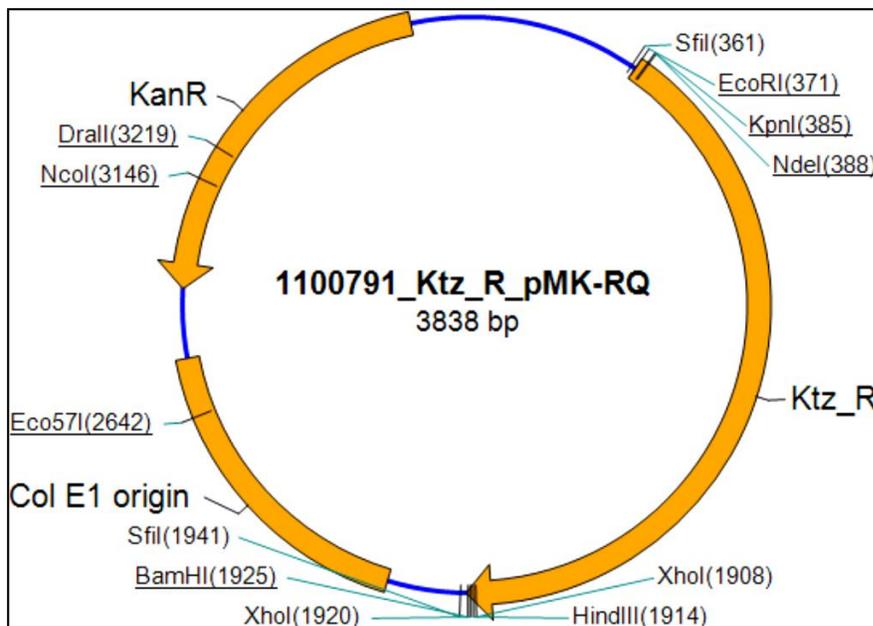


Figure 57. *ktzR* constructed in pMK-RQ (construt map provided by GENEART)

The 1544 bp *ktzR* gene (appendix) was subcloned into *NdeI/XhoI* restricted pET-28a(+) (pET-SAS5) as mentioned in sections 5.11-5.13. Clones were confirmed by restriction digestion and further verified by sequencing (GATC, Germany). The expression of KtzR in *E. coli* BL21(DE3) cells was carried out in LB, LB supplemented with 20 $\mu\text{g ml}^{-1}$ tryptophan, M9 and LB supplemented with unrefined cane molasses¹⁰⁵ at 16 °C for 3 hr. KtzR expression was analysed by 12% (v/v) SDS-PAGE, and expression was observed in all media except M9 (data not shown). On further analysis, the enzyme was determined to be insoluble. To try to obtain soluble protein, cell growth and protein expression was slowed down by addition of 0.4 M sucrose to the media, growth of cells at the lower temperature of 14 °C and inducing with lower concentrations of IPTG. An

alternative cold and heat shock method was also used which revealed the presence of soluble protein by SDS-PAGE (data not shown). An assay was carried out using the crude lysate fraction with tryptophan as the substrate and showed no halogenation suggesting the protein is inactive.

3.3.2. Re-cloning and expression of KtzR

Comparison of the KtzR protein sequence with the sequence of another tryptophan-6-halogenase SttH revealed the potential absence N-terminal amino acids in KtzR. On further analysis of the sequences from the NCBI databases, it was concluded that the gene sequence was deposited incorrectly and was missing the sequence encoding the first 20 amino acids of the protein. Primers were designed to attach the missing 60 nucleotides (20 amino acids) and clone the full length gene into pET-28a(+). Primer P8 was designed to attach the missing nucleotides and primers P27 and P28 were used to amplify the full length *ktzR* sequence (1604 bp). Full length *ktzR* was cloned into *NdeI*/*NotI* restricted pET-28a(+) (pET-KK22) to express an N-terminal His₍₆₎-tagged KtzR protein. The clone was confirmed by restriction digest of the plasmid and sequencing (GATC, Germany) (Figure 58).

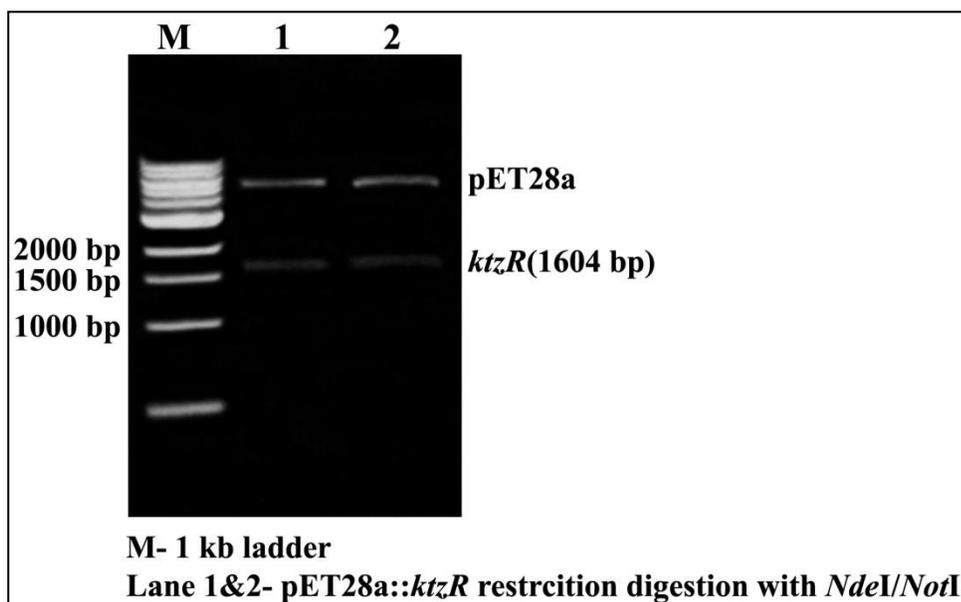


Figure 58. Restriction digestion analysis of pET-KK22 with *NdeI/NotI*

For expression pET-KK29 was transformed into *E. coli* BL21 (DE3) and *E. coli* ArcticExpress (DE3) RP cells. The expression of KtzR in *E. coli* BL21 (DE3) and *E. coli* ArcticExpress (DE3) RP was carried out as described in section 5.15.5. The protein was purified by Ni²⁺ affinity chromatography and concentrated to 1 ml at a concentration of 0.8-1 mg ml⁻¹(Section 5.15). 20 µl of pure KtzR was analysed by 12% (v/v) SDS-PAGE and a band was observed between 55 kDa and 70 kDa consistent with the expected mass (61.4 kDa).

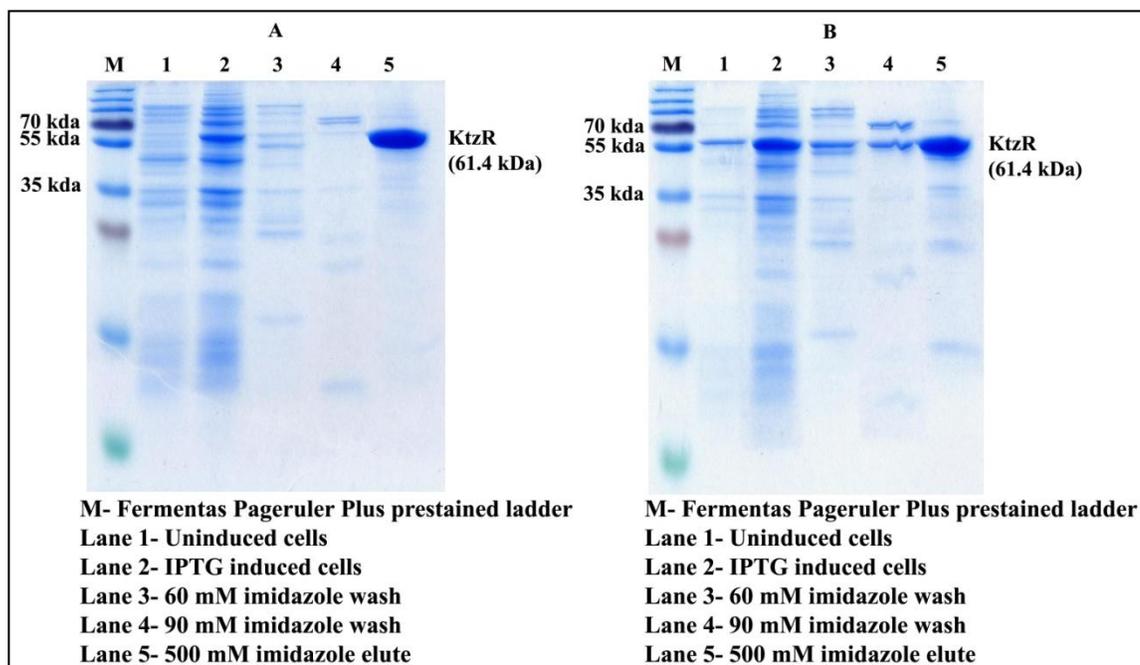


Figure 59. Analysis of pure KtzQ protein from (A). *E. coli* BL 21(DE3) cells, (B) *E. coli* ArcticExpress DE3(RP) by 12% SDS-PAGE

3.3.3. Halogenase activity of KtzR on tryptophan

The KtzR assay was carried out by Sarah Shepherd with L-tryptophan (section 5.22) using 5 µM of halogenase in a total volume of 200 µl in 10 mM potassium phosphate buffer pH 7.2. KtzR showed chlorination of L-tryptophan at the 6-position which was confirmed by LC-MS (239.1 (M+H) and 241.1 in the characteristic 3:1 ratio for chlorine isotopes. Sarah screened KtzR against a library of other potential substrates (Figure 83) and unfortunately found no chlorination or bromination activity.

3.3.4. Purification of KtzR for obtaining crystal structure

The crystal structure of the tryptophan-7-halogenases [PrnA(PDB ID:2AR8)]⁶⁴ and tryptophan-5-halogenases [PyrH (PDB ID:2WET)]⁸⁸ have been solved and provide insight for further mutagenesis studies designed to improve the substrate specificity and regioselectivity of these enzymes. The crystal structure of a tryptophan-6-halogenase has not yet been solved preventing rational mutagenesis of these enzymes.

An effort was undertaken to obtain the crystal structure of KtzR (a tryptophan 6-halogenase). Protein was expressed on a large scale (4 litres) (5.15). The protein purification for crystal trials was carried out using a HiTrap™ Chelating HP column for Ni²⁺ affinity chromatography, HiTrap™ Q HP anion exchange column and a Superdex™ 200 10/300 gel filtration column.

In the initial purification attempt the cells were lysed and purified in batches using the HiTrap™ Chelating HP column and purified fractions were pooled together prior to the next purification steps. The pooled sample had contaminating higher and lower mass proteins. To remove these, the samples were diluted to remove the salt and loaded onto a HiTrap™ Q HP anion exchange. To attempt to remove the contaminating proteins lower salt concentrations washes (20 mM-250 mM NaCl) were used and KtzR was eluted using a higher salt concentration of 500 mM NaCl. However, this failed to remove the contaminants a significant amount of the KtzR protein was lost. Insufficient protein was obtained and the purity achieved was unsuitable for crystallography.

In subsequent attempts the purification was carried out by overloading the nickel column with the lysate. The HiTrap™ Chelating HP column can hold a maximum of 60 mg of protein per ml of resin and the advantage of overloading the column increases the probability of specific binding of His₍₆₎-tagged proteins while reducing the probability of nonspecific binding of other proteins. The protein from the HiTrap™ Chelating HP column had comparatively fewer contaminating proteins. The protein was then subjected to two rounds of purification using a HiTrap™ Q HP anion exchange column, eluting with high salt concentration. The protein obtained was much purer with a single band observed by SDS-PAGE. This was subjected to Superdex™ 200 10/300 gel

filtration column for exchange of the buffer and was concentrated to 600 μl at a concentration of 8 mg ml^{-1} . 20 μl of pure KtzR was analysed by 12% (v/v) SDS-PAGE (Figure 60). The pure protein was subjected to crystallisation trials.

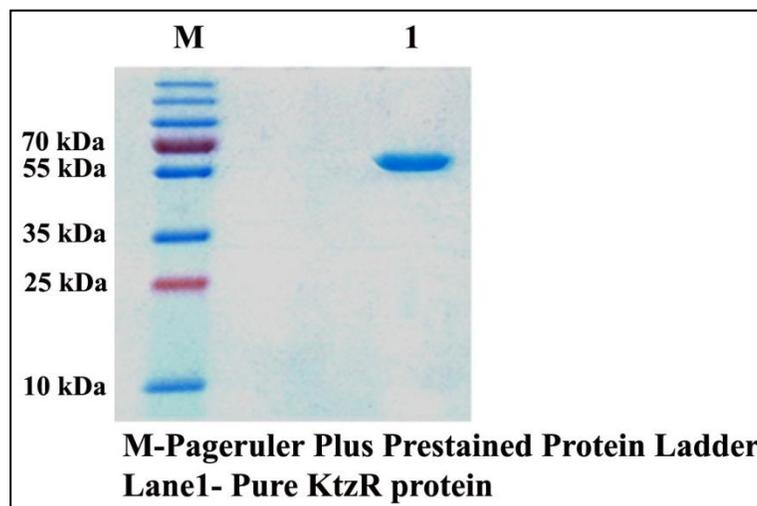


Figure 60. Analysis of pure KtzR protein by 12%(v/v) SDS-PAGE Initial crystal trials did not yield any good crystals for obtaining the structure of the enzyme. The crystallization of this protein was not pursued out due to the inability of this enzyme to halogenate substrates other than the natural substrate tryptophan.

3.4. SttH (tryptophan-6-halogenase)

Since KtzR was an unsuitable enzyme in terms of the amount of protein obtained and its substrate specificity scope, it was necessary to investigate other efficient 6-halogenases for comparative studies with other halogenases. SttH enzyme was isolated in 2011 by Zhang *et al.* from *Streptomyces toxytricini* NRRL 15443, and halogenates both L and D -tryptophan to produce products halogenated at the 6-position⁷⁹. Our work focussed on cloning, expressing and, purifying SttH in *E. coli*, in order to study its mechanism of regioselective halogenation and identifying important residues for the chlorination of tryptophan at the 6-position. We will also explore its substrate specificity and residues governing this.

3.4.1. Cloning and expression of tryptophan-6-halogenase SttH

The gene sequence of *SttH* was obtained from the *S. toxytricini* NRRL 15443 biosynthetic gene cluster deposited in the NCBI database (appendix).

A codon optimised *SttH* gene with desired flanking regions was designed and obtained in plasmid pMK-RQ (Geneart, Life Technologies, Invitrogen, UK) (Figure 61).

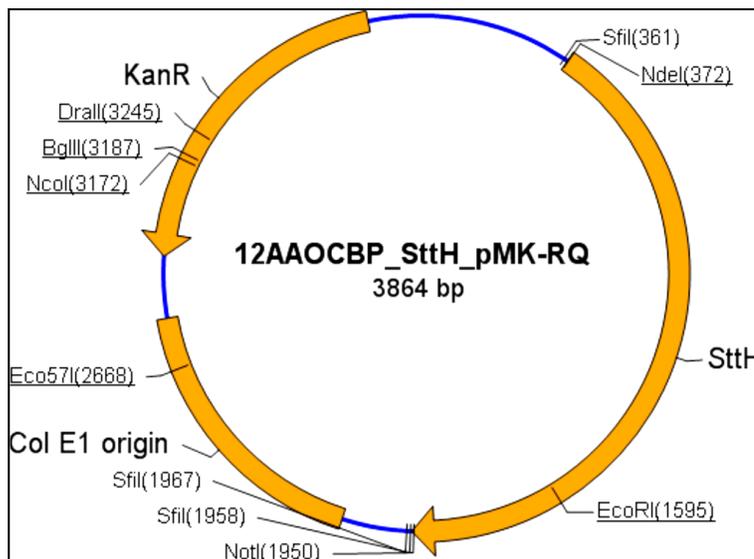


Figure 61. *SttH* constructed in pMK-RQ plasmid (construct map provided by GENEART)

The 1572 bp *SttH* gene was subcloned into *NdeI/XhoI* restricted pET-28a(+) (pET-SAS6) as described in sections 5.11-5.13 by Sarah Shepherd. Clones were confirmed by restriction digestion to verify the presence of insert and further verified by sequencing (GATC, Germany).

For expression, pET-SAS6 was transformed into *E. coli* ArcticExpress (DE3) RP cells and the expression, isolation and purification were carried out as per section 5.15. Pure SttH was eluted with from a Ni²⁺ affinity chromatography with 500 mM imidazole and was subjected to buffer exchange by dialysis (overnight) in phosphate buffer. The SttH was concentrated to 1 ml with a concentration of 6 mg ml⁻¹ and stored in storage buffer at -20°C. Pure SttH was observed between 55 kDa and 70 kDa consistent with the expected mass (61.1 kDa) by 12% (v/v) SDS-PAGE (Figure 62).

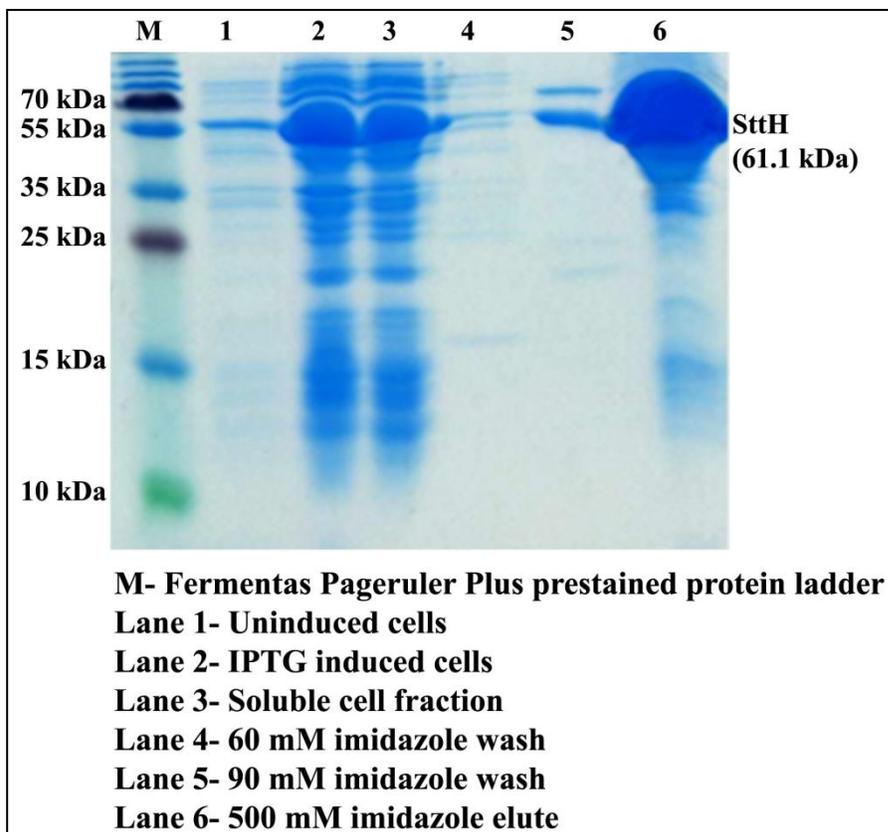


Figure 62. Analysis of pure SttH by 12%(v/v) SDS-PAGE

3.4.2. Substrate specificity of SttH enzyme

Sarah Shepherd carried out the scope of substrate specificity studies on SttH against a wide range of substrates (Figure 80). Briefly, she found that SttH was able to chlorinate and brominate tryptophan, kynurenine, anthranilamide, methyl anthranilate and 2-propyl aniline (Figure 63). The enzyme chlorinated its natural substrate tryptophan the best (65%), followed by kynurenine (50%) and anthranilamide (about 10%). The regioselectivity of the some of the above products produced by SttH was then confirmed using NMR (Figure 63).

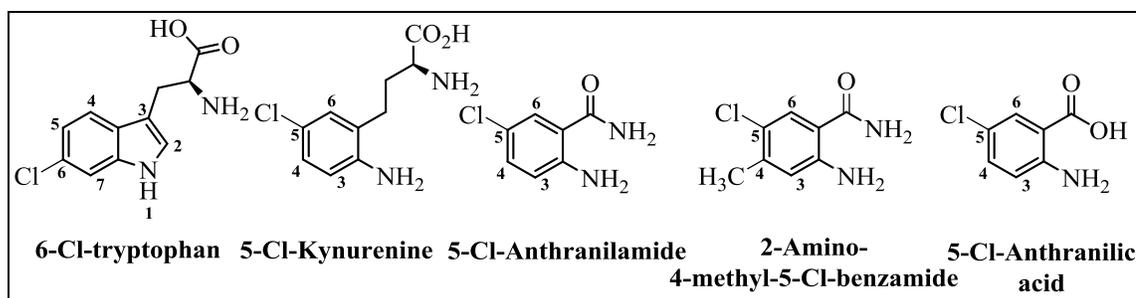


Figure 63. Substrates that were chlorinated by SttH

In tryptophan, the first highly preferred chemical halogenation is at position 3. But since the position is occupied by extending amino acid chain, the next preference of halogenation will be at position 2, which is not halogenated in enzyme system. If position 3 is blocked by the position 2, the next favourable halogenation of the substrate by EAS will occur on indole ring of the tryptophan which is favoured by SttH to chlorinate at the 6-position of tryptophan. In kynurenine, the presence of NH_2 group favours more as ortho, para directing group over the extended amino acid chain (also an ortho, para directing group). In chemical halogenation, either position 3 or 5 are preferred to carry out EAS. PrnA is more regio-selective to halogenate at the 5 position rather than producing a mixture of both 3 and 5 tryptophan halogenases. In anthranilamide, NH_2 group favours more ortho, para directing group over the CONH_2 group (meta directing group) preferring either 3 or 5 position for halogenation. The position 3 is ortho to para directing and meta to meta directing while position 5 is para to ortho, para directing group and meta to meta directing group. Chemical halogenation would produce mixture of both 3 and 5 halogenated anthranilamide on EAS but SttH halogenates more selectively at the 5 position. Similar halogenation is observed for anthranilic acid as seen in anthranilamide. In 2-amino-4-methylbenzamide, position 3 is ortho to ortho, para directing groups (NH_2 and CH_3 respectively) and meta to meta directing group (CONH_2) and would therefore be much preferred position for halogenation. While position 5 will also be preferred because it is ortho to CH_3 (ortho, para directing group), para to NH_2 group (ortho, para directing group) and meta to

CONH₂ (meta directing group). SttH halogenates more selectively at the 5 position neglecting the 3 position to be halogenated in the enzyme system.

3.4.3. Regioselectivity of SttH

As mentioned in section 3.3.4, the structure of a 6-tryptohan halogenase has not yet been solved, which makes it difficult to design rational mutagenesis strategies for these enzymes. We tried to study the regioselectivity of the SttH enzyme based on the sequence alignment with other flavin dependent halogenases. Lang *et al.*, 2011 found that a PrnA F103A mutant changed the regioselectivity of the enzyme from the 7-position to the 5-position of tryptophan⁸⁹. On sequence alignment, this F residue is conserved in all the flavin dependent halogenases (Figure 64). In order to study the regioselectivity of SttH, initially an attempt was made to mutate the corresponding F residue at position 98 of SttH to A.

```

STRHY-5  FFDYLG LSEADWMPECNGTYKLAVRLENWRAPG-----T64YFYHPERMRVV 85
PyrH-5   FFDYLG LDEREWLPRCAGGYKLGIRFENWSEPG-----EYFYHPERLRVV 100
Thal-6   FFDYLG IPEEEWMRECNASYKMAVKFINWRTPGEGSPDPRTLDDGHTDTFHHPGGLLPSA 118
KtzR-6   FFQFLNLREQDWMPCNATYKLGIRFENWRHVG-----HHFYQPEQIRPV 105
SttH-6   FFEFLDLREEEWMPCNATYKLAVRFQDWQRPG-----HHFYHPEQMRSV 104
RebH-7   FFDFLG IPEDEWMRECNASYKVAIKFINWRTAGEGTSEARELDGG-PDHFYHSGLLKYH 117
KtzQ-7   FFDFLG IAEDEWMRECNASYKAAVRFVNWRTPGDQATPRRRPDGRPDHFDHLGQLPEH 119
PrnA-7   FFDFLG IPEREWMPQVNGAFKAAIKFVNWRKSPDPSRD-----DHFYHLGNVPNC 109
**:*:*:* * :*: . :* :*: :* * : * :

```

Figure 64. Sequence alignment of tryptophan 5-, 6- and 7- halogenases showing the conserved phenylalanine residue

The construct pET-SAS6 was used as template DNA for site directed mutagenesis (section 5.14) using primer P29 (Table 32). The mutant clone (pET-KK23) was verified by DNA sequencing (GATC, Germany) and transformed into *E. coli* ArcticExpress (DE3) RP cells. The SttH F98A mutant was expressed in *E. coli* ArcticExpress (DE3) RP cells, isolated, purified and concentrated to 1 ml with a concentration of 2.39 mg ml⁻¹ as described in section 5.15 (Figure 65).

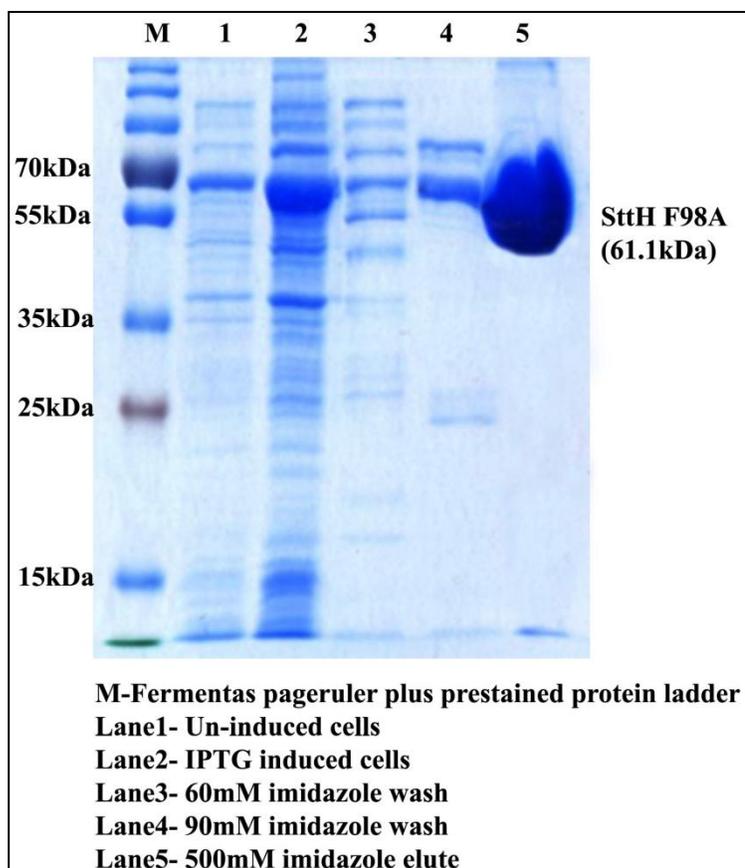


Figure 65. Analysis of pure SttH F98A by 12% (v/v) SDS-PAGE

3.4.3.1. Enzyme assay of SttH F98A

The activity of the SttH F98A mutant was enzymatically assayed (section 5.22). 5-chloro-tryptophan, 6-chloro-tryptophan and 7-chloro-tryptophan were obtained using wild type enzymes PyrH, SttH and, PrnA respectively and were used as standards in this experiment. The comparison of the SttH F98A mutant assay with the standards did not show any change in regioselectivity, showing chlorination of tryptophan only at the 6-position. The mutant also showed a very low activity relative to that of wild type (Figure 66). Attempts at bromination of tryptophan using the SttH F98A mutant were unsuccessful.

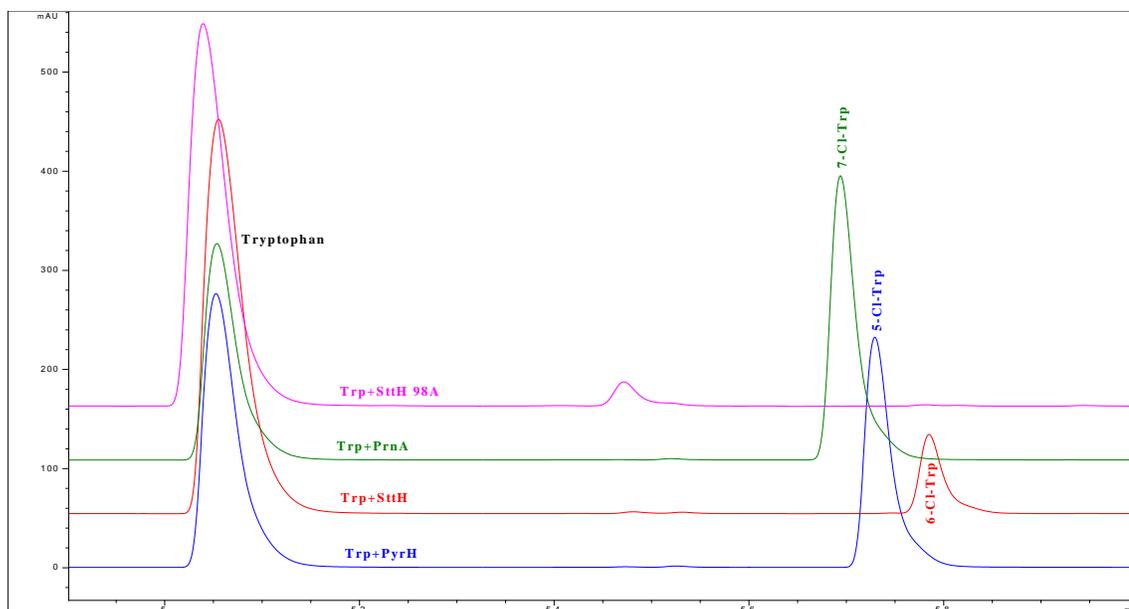


Figure 66. HPLC chromatogram showing low chlorination activity by *SttH* F98A mutant

Since the above results did not show any change in regioselectivity, further mutagenesis was attempted to investigate other potential residues governing regioselectivity. Further analysis by sequence alignment revealed that an 8 amino acid loop region present in tryptophan-7-halogenases was absent in *SttH*. The effect of this loop on the regioselectivity of the enzymes was investigated by inserting the 8 amino acid loop from the 7-halogenases into *SttH* to see if it would then be able to chlorinate tryptophan at the 7-position. Mutagenesis was carried out by site directed mutagenesis (section 5.14) using pET-SAS6 as template DNA and primer 30-31 (Table 32). Mutant clone (pETKK24) were verified by DNA sequencing (GATC, Germany) and transformed into *E. coli* ArcticExpress (DE3) RP cells. The *SttH* loop insert mutant was expressed in *E. coli* ArcticExpress (DE3) RP cells, isolated, purified and concentrated to 0.54 mg ml^{-1} as described in section 5.15.

3.4.3.2. Enzyme assay of *SttH* loop (*SttH* LI) inserted mutant for regioselectivity

The activity of the *SttH* loop inserted mutant protein was enzymatically assayed. (Section 5.22) The HPLC trace of the *SttH* loop insert mutant showed very low chlorination of tryptophan at the 6-position relative to the wild type indicating that

insertion of this loop effects the activity of the enzyme, possibly by disrupting the structure of the active site. Similar results were observed for bromination showing no change in regioselectivity with a very low relative activity.

Overall, the above mutations did not cause any change in regioselectivity for either chlorination or bromination. Also, the mutants showed much lower relative activity (Table 10) indicating that these residues influence the halogenation mechanism or the structure of enzyme.

Table 10. Overall relative activity of SttH mutants obtained for regioselective studies

Mutants	Relative activity (%)	
	Chlorination	Bromination
SttH F98A	0.136	-
SttH LI	0.160	0.265

3.4.4. SttH for crystal trials

Further mutagenesis was not carried due to the unavailability of a crystal structure of SttH. Earlier, KtzR was purified and subjected for crystal trials (section 3.3.4) yielding no crystals. These trials were not repeated because SttH had a better substrate scope than KtzR. It is important to obtain the crystal structure of SttH enzyme to understand the mechanism of the regioselective chlorination of tryptophan at the 6-position which will also help us to study the residues governing substrate specificity and subsequently widen the substrate scope.

3.4.4.1. Purification of SttH for crystal trials

SttH protein expression and purification from *E. coli* ArcticExpress (DE3) RP cells had similar issues with contaminating chaperones as the expression and purification of PrnA (Section 3.2.6). However, SttH expressed at higher levels in the host system compared to PrnA. Therefore, loss of a small amount of SttH would still yield the required amount of protein for crystal trials. Hence, SttH halogenase was purified following a similar method to PrnA (section 3.2.6). Contaminating proteins bound to the column were removed by a 90 mM imidazole wash prior to removal of the chaperones by an ATP wash. SttH was also eluted using an ATP buffer.

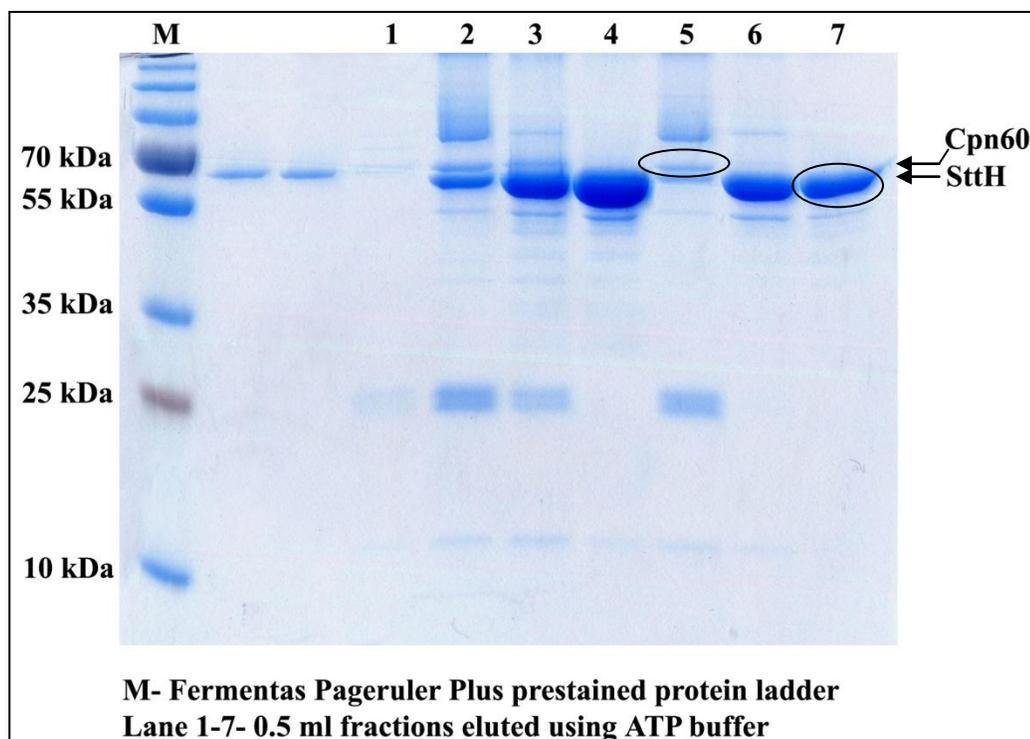


Figure 67. Analysis of ATP treated *SttH* protein fractions by 12%(v/v) SDS-PAGE

The analysis of fractions by 12% (v/v) SDS-PAGE (Figure 67) revealed the release of the chaperones in fractions 2, 5 and 6 of the ATP wash. Pure *SttH* were eluted in fractions 4 and 7 of the ATP wash. Although this produced purer protein, there was still a significant proportion of contaminating proteins. The protein purification using further columns was not carried out due to the low concentration of the pure protein in the cleanest fractions.

The protein was later expressed in and purified from *E. coli* BL21 (DE3) cells (section 5.15). The protein was purified by overloading lysate onto a HiTrap™ Chelating HP Ni²⁺ column to avoid nonspecific binding of contaminating proteins followed by anion exchange chromatography on a HiTrap™ Q HP column. This failed to produce protein of a purity suitable for crystallisation trials.

To obtain pure protein, the protein purification was repeated with modified steps. The HiTrap™ Chelating HP column was first overloaded with the lysate and fractions were analysed by 12% (v/v) SDS-PAGE (Figure 68). All the fractions had similar

amounts of the contaminating proteins. The fractions having the highest amount of SttH protein were 14, 15, 16 and 17 and were pooled together for further purification.

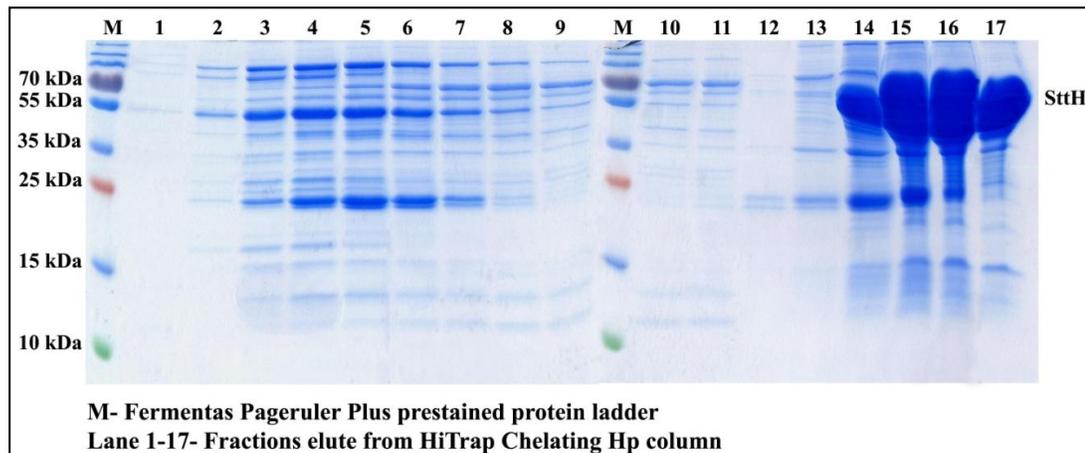


Figure 68. Analysis of fractions purified from a HiTrap™ Chelating HP column by overloading SttH by 12% (v/v) SDS-PAGE

The collected fractions were concentrated down and buffer exchanged into phosphate buffer with no salt to reduce the concentration of the salt prior to anion exchange chromatography. The purification was continued with a HiTrap™ Q HP anion exchange column and protein was eluted at 500 mM salt concentration (Figure 69). SttH protein in fractions 1 and 3 was the purest and these fractions were pooled together. Fraction 2 still contained contaminants and was further subjected to a second round of anion exchange chromatography which failed to improve the purity.

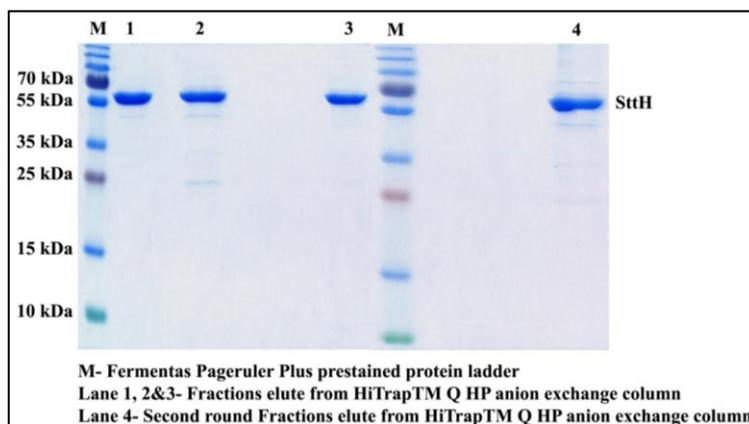


Figure 69. Analysis of purified fractions from the HiTrap™ Q HP anion exchange column by 12% (v/v) SDS-PAGE

The pooled fractions were further purified by gel filtration using the Superdex™ 200 10/300 column. The SttH thus purified was concentrated to 12 mg ml⁻¹. Various amounts of the protein were analysed by 12% (v/v) SDS-PAGE (Figure 70)

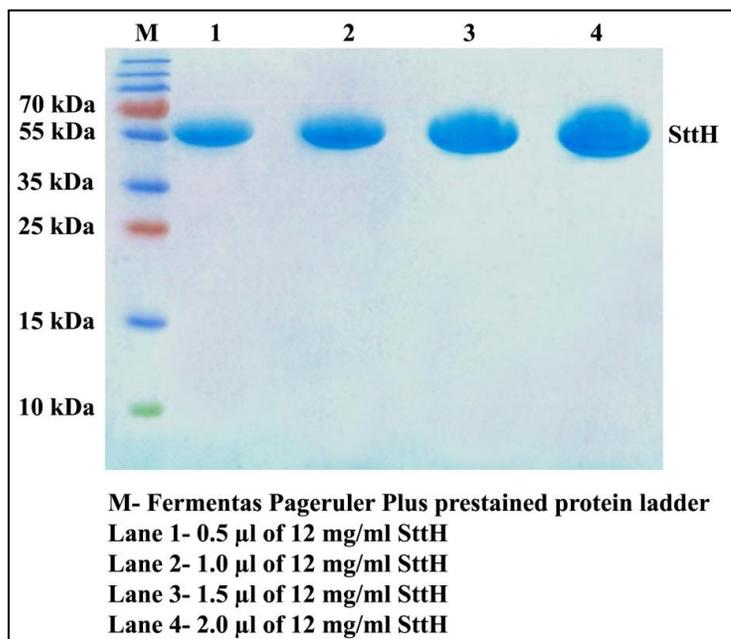


Figure 70. Analysis of pure SttH protein on 12% resolving gel

MALS analysis of SttH revealed a mono-disperse sample. A molecular mass of 117.9 kDa was obtained from MALS, double the mass of a monomer of the protein indicating that the protein is a dimer in solution, like other flavin dependent halogenases.

Crystal trials were carried out by hanging drop method. Initially the crystals formed were too small for determining the structure of the SttH halogenase (Figure 71). Since the protein had a very tiny amount of contamination with other proteins, it might prevent the formation of better crystals. The protein purification was repeated again to obtain the protein at much higher purity.

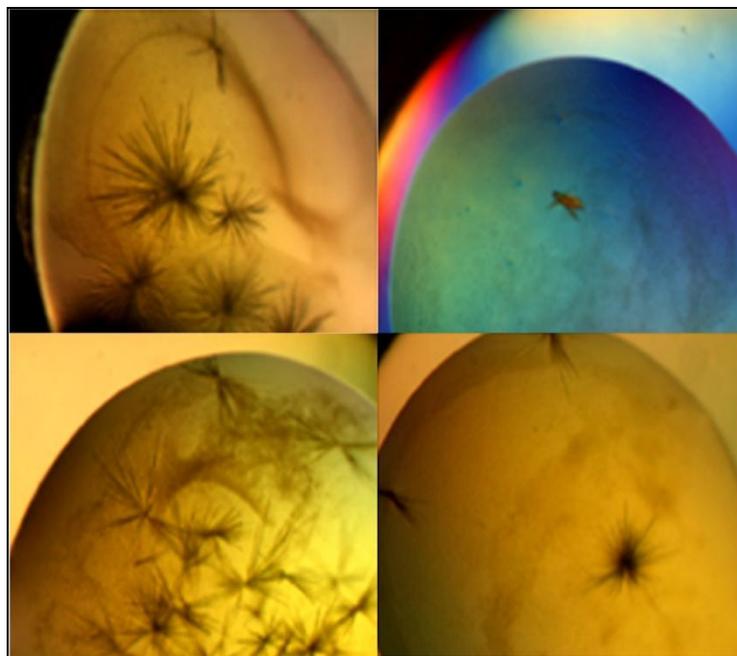


Figure 71. SttH crystals obtained on crystal trials

The SttH halogenase purification was repeated many times following both methods. However, it was not possible to obtain the quantity of protein at the purity required for crystallisation trials using the above purification strategies. The pure proteins obtained at different stages were not concentrated enough for crystal trials. Another purification step was needed to achieve much better results.

Finally a four step procedure was optimised to yield pure SttH. The protein was first purified by Ni²⁺ affinity chromatography using the gravity flow method (Section 5.15.6.2). The protein produced by the gravity flow method had a similar level of purity to that produced using the HiTrap™ Chelating HP column. Analysis by 12% (v/v) SDS-PAGE revealed that fraction 1 was relatively pure compared to fractions (Figure 72). But the other fractions had much higher amounts of SttH and were used further for purification.

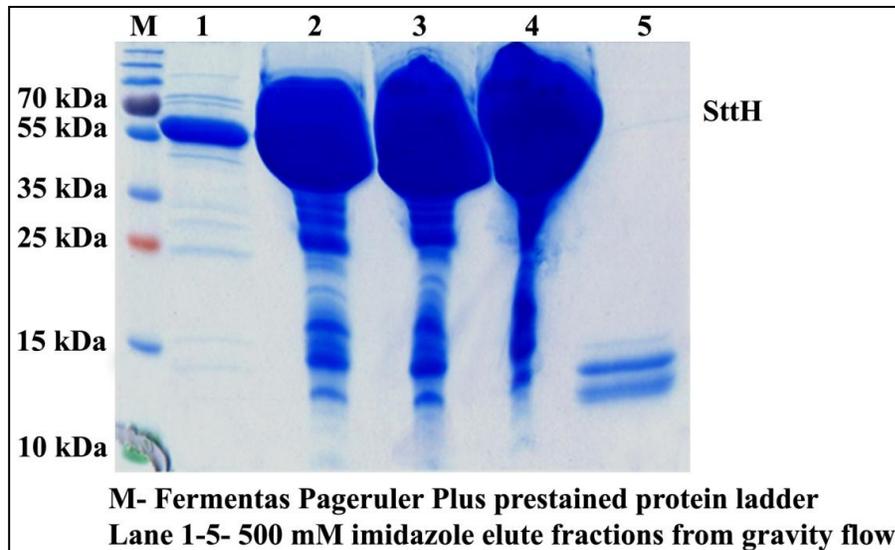


Figure 72. Analysis of purified *SttH* fractions by nickel binding method by 12%(v/v) SDS-PAGE

Fractions 2, 3 and 4 were pooled together and 5 ml of sample was injected onto the HiTrap™ Q HP anion exchange column; *SttH* was eluted using 500 mM salt. The fractions were analysed by 12% (v/v) SDS-PAGE and it was observed that fractions 2, 9 and 10 were much purer than the other fractions (Figure 73).

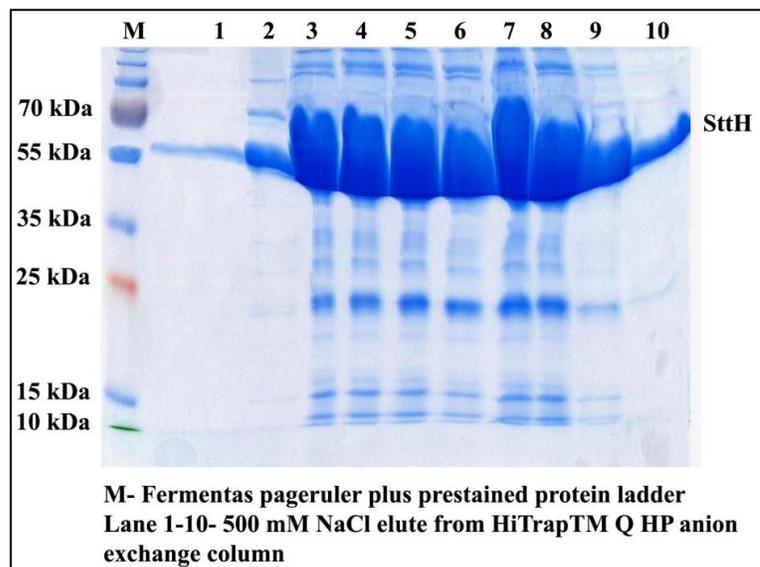


Figure 73. Analysis of purified *SttH* fractions from HiTrap Q HP anion exchange column by 12%(v/v) SDS-PAGE

To obtain a better yield of SttH for crystal trials, fractions 1-10 were pooled together, buffer exchanged and concentrated. The protein was subjected to a MonoQ 4.6/100PE anion exchange column and protein was eluted at 500 mM salt. The fractions showed increased purity compared to that of the HiTrapTM Q HP anion exchange column. Fractions 3 and 4 were much purer and were pooled for further purification (Figure 74a). The other fractions were stored for future assays.

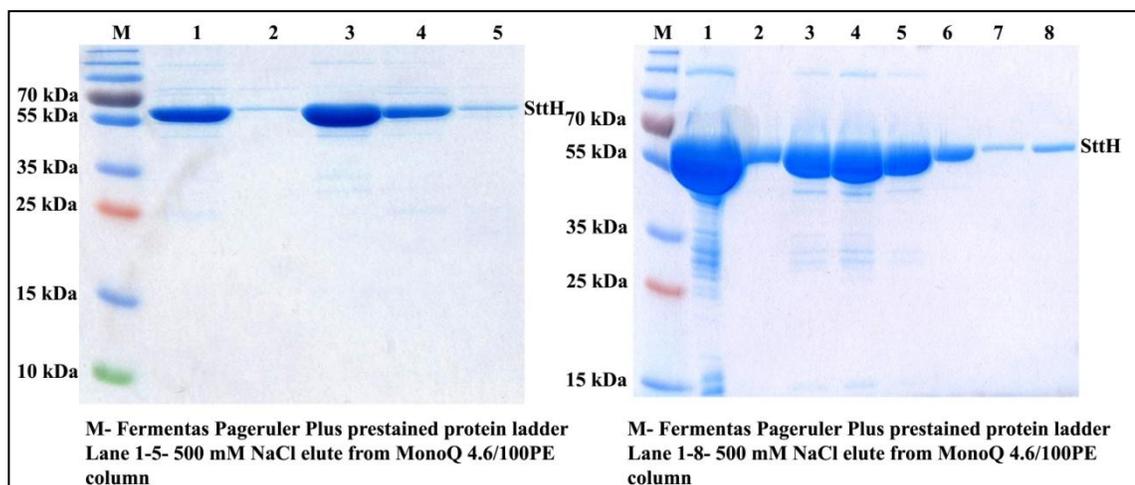


Figure 74. Analysis of purified SttH fractions from the MonoQ HP anion exchange column by 12%(v/v) SDS-PAGE

The pooled fractions were subjected to second round of anion exchange chromatography using the MonoQ 4.6/100PE column. Fractions 2, 6, 7 and 8 contained pure SttH halogenase and were pooled together for gel filtration. Fractions 3, 4 and 5 were pooled as separate fractions for carrying out gel filtration (Figure 74b).

The pooled fractions were concentrated to a volume of 0.5 ml and two fractions of 250 μ l were loaded onto a SuperdexTM 200 10/300 gel filtration column. Lower concentrations of the protein yielded better results than higher concentrations. The fractions were analysed by 12% (v/v) SDS-PAGE. The fractions contained very pure SttH halogenase (Figure 75). The fractions were pooled together, concentrated to 18.53 mg ml^{-1} and subjected to crystallisation trials.

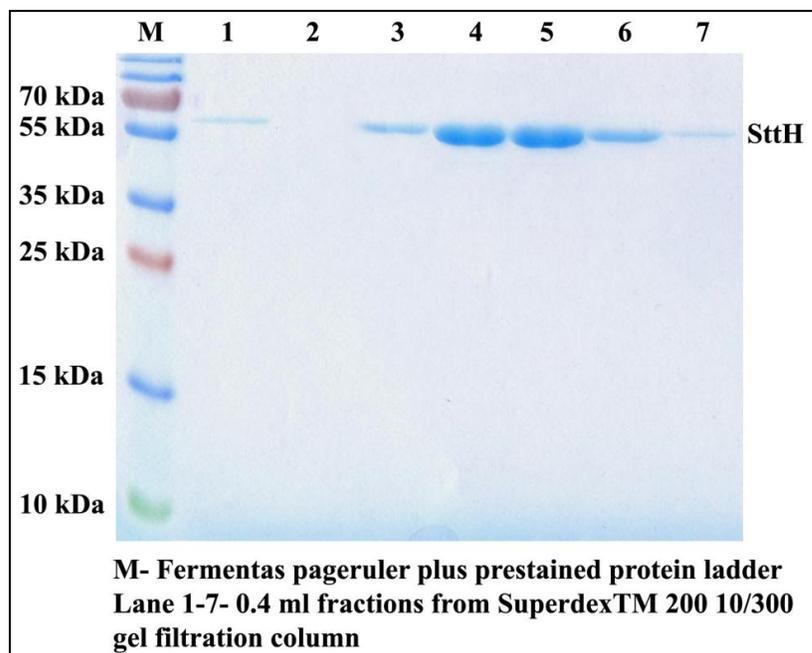


Figure 75. Analysis of purified *SttH* fractions from the Superdex™ 200 10/300 column by 12%(v/v) SDS-PAGE

The crystallisation trials yielded a tiny crystal as observed in previous results (Figure 75). Further protein purification is being repeated at present to obtain better crystals to solve the crystal structure of *SttH*.

3.5. PyrH (tryptophan-5-halogenase)

PyrH is a tryptophan-5-halogenase from pyrroindomycin producing *S. rugosporus* LL-42D005. The enzyme was isolated and characterized previously by Karl-Heinz van Pee's group in 2005⁸⁸ for activity on its natural substrate tryptophan. Their work focussed on studying its mechanism and identifying important residues for the chlorination of the tryptophan at the 5-position. We cloned, expressed and purified PyrH to explore its substrate specificity and regioselectivity.

3.5.1. Cloning and expression of tryptophan-5-halogenase PyrH

The gene sequence of *pyrH* was obtained from the *Streptomyces rugosporus* biosynthetic gene cluster deposited in the NCBI database (appendix).

A codon optimised *pyrH* gene with desired flanking sequences was designed and obtained in plasmid pMK-RQ (Geneart, Life Technologies, Invitrogen, UK) (Figure 76).

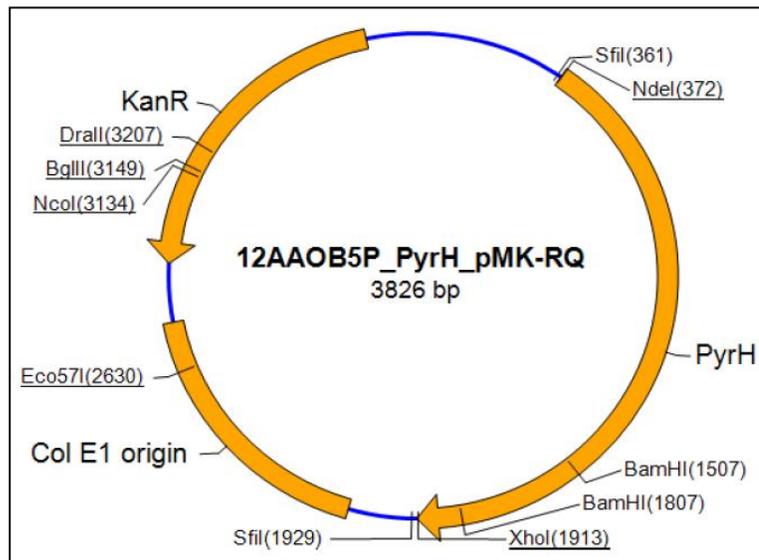


Figure 76. pMK-RQ::*pyrH* synthetic gene construct (construct map provided by GENEART)

The 1536 bp *pyrH* gene was sub-cloned into *NdeI/XhoI* restricted pET-28a(+) (pET-KK25) as described in sections 5.11-5.13. Clones were confirmed by restriction digestion and further verified by sequencing (GATC, Germany) (Figure 77).

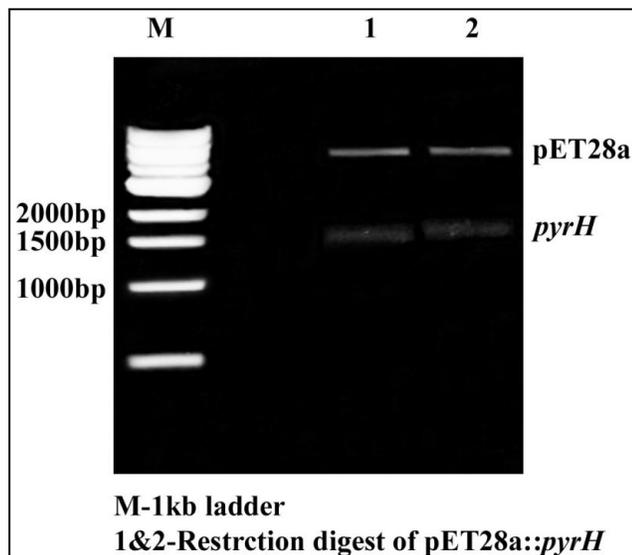


Figure 77. Restriction digestion analysis of sub-cloned construct pETKK25 with *NdeI/XhoI*

For overproduction of PyrH, pET-KK25 was transformed into *E. coli* ArcticExpress (DE3) RP cells. The expression, isolation, purification and concentration were carried out as per section 5.15. Pure PyrH was eluted from Ni²⁺ affinity column with 500 mM imidazole, dialysed, concentrated to 6.2 mg ml⁻¹ and stored in storage buffer (section 5.15). Pure PyrH was observed between 55 kDa and 70 kDa consistent with the expected mass of (60.5 kDa) when analysed by 12% (v/v) SDS-PAGE (Figure 78).

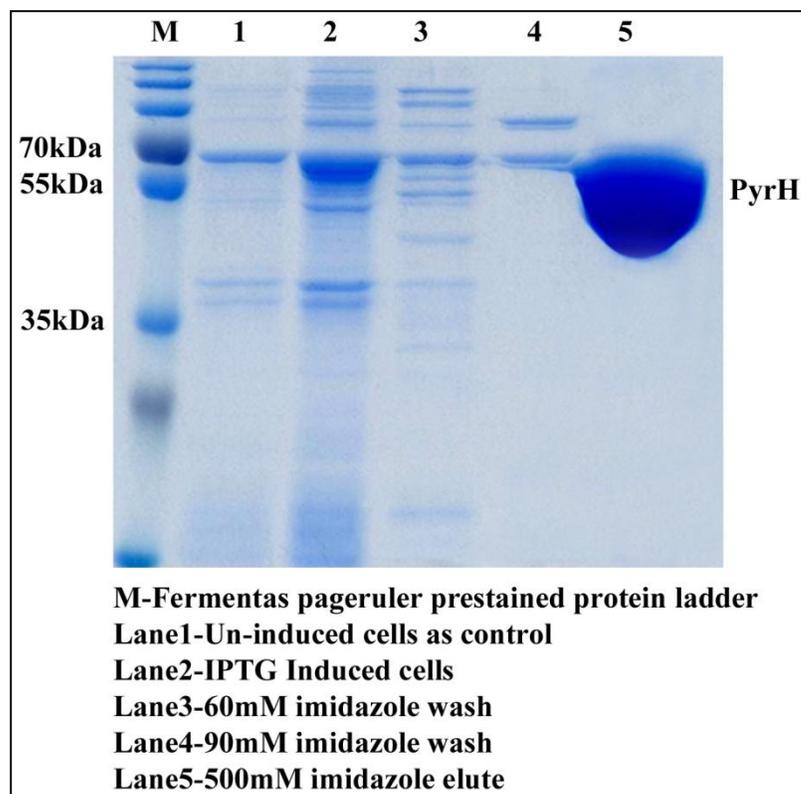


Figure 78. Analysis of purified PyrH by 12%(v/v) SDS-PAGE

3.5.2. Stability of purified PyrH in DMSO and ethanol

The planned assays on a wide range of substrates were constrained by the solubility of the substrates in water. However, substrates were easily soluble in DMSO and ethanol. Solvents could affect the conformational states of enzymes and consequently change their activity^{106,107}. Therefore, the effects of the solvents DMSO and ethanol were tested on purified PyrH before it was used in assays.

Assays were carried out as described in section 5.22 and analysed by HPLC (Figure 79). The relative activity was calculated based on the % area of substrate peak and the % area of the product peak. The relative activity at each time point was plotted against time.

The PyrH assays in 5% and 10% DMSO concentration showed no precipitation in solution. The level of chlorination at 1 hr observed for the PyrH in 5% (v/v) DMSO (80%) was similar to the level seen in assays carried out with 10 mM phosphate buffer. Incubation with 10% (v/v) DMSO resulted in only 50% chlorination at 1 hr relative to the no solvent control. After 2 hr complete chlorination of the substrate was observed for the no solvent control and in 5% (v/v) DMSO. Incubation with 10% (v/v) DMSO yielded only 67-68% chlorination after 2 hr suggesting a decrease in enzyme activity. The efficiency of PyrH was observed to reduce in 10% (v/v) DMSO which could be due to an alteration of the enzyme structure in that particular concentration of solvent. The results indicated that PyrH could withstand a maximum of 5% DMSO without altering its activity.

The activity of PyrH in ethanol was more restricted relative to DMSO with no activity observed in concentrations higher than 5% (v/v). In 5% (v/v) ethanol PyrH chlorinated more of the substrate relative to the 5% (v/v) DMSO (93% compared to 81% in 5% (v/v) DMSO) after 1 hr. This was also higher than the no solvent control (81%). After 2 hr 100% conversion was observed in 5% (v/v) DMSO, 5% (v/v) ethanol and the no solvent control.

The stability studies indicated that the enzyme assays could withstand a maximum of 5% (v/v) DMSO or 5% (v/v) ethanol without loss of enzyme activity. Although, activity in 5% (v/v) ethanol was initially higher than in 5% (v/v) DMSO, the solubility of most of the planned test substrates was higher in DMSO. Therefore 5% (v/v) DMSO was the preferred solvent concentration for PyrH activity assays.

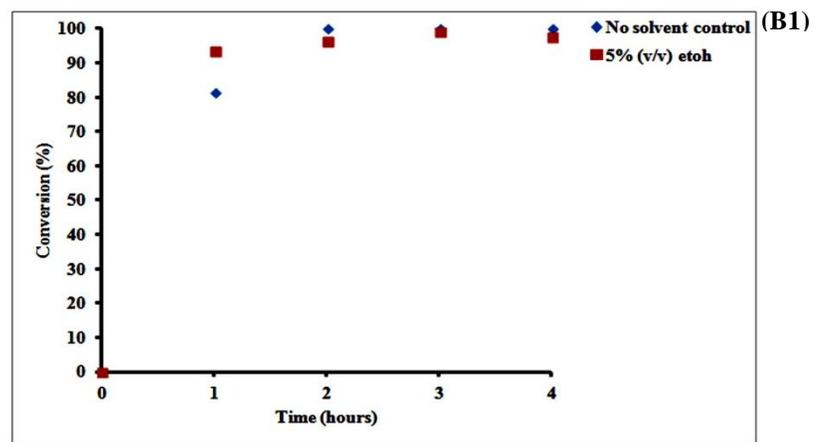
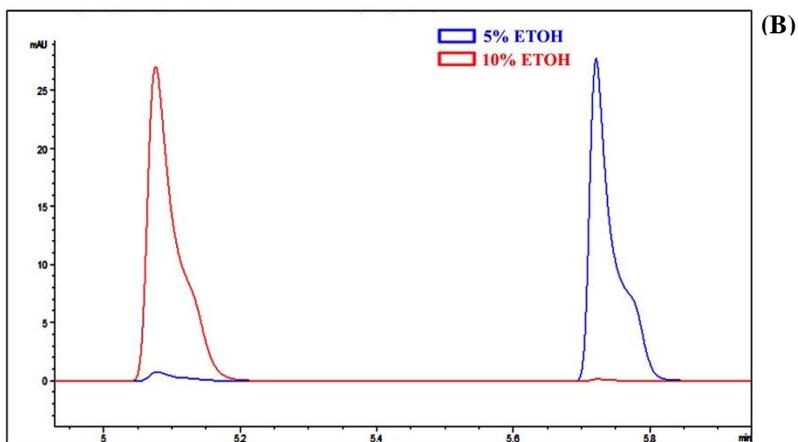
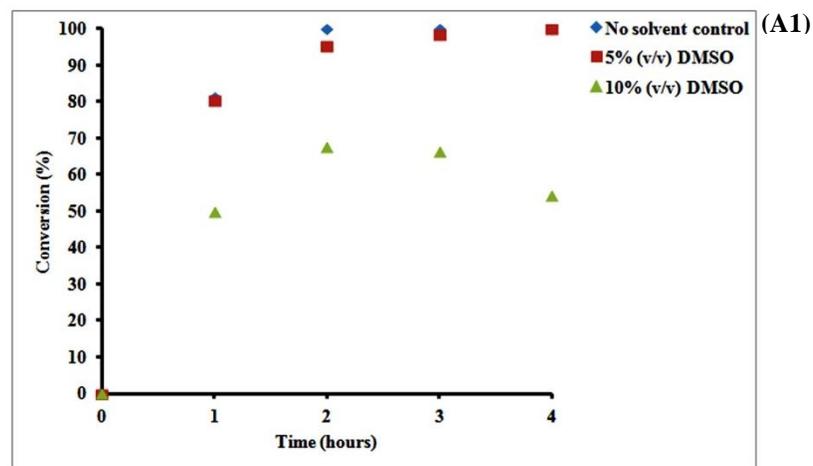
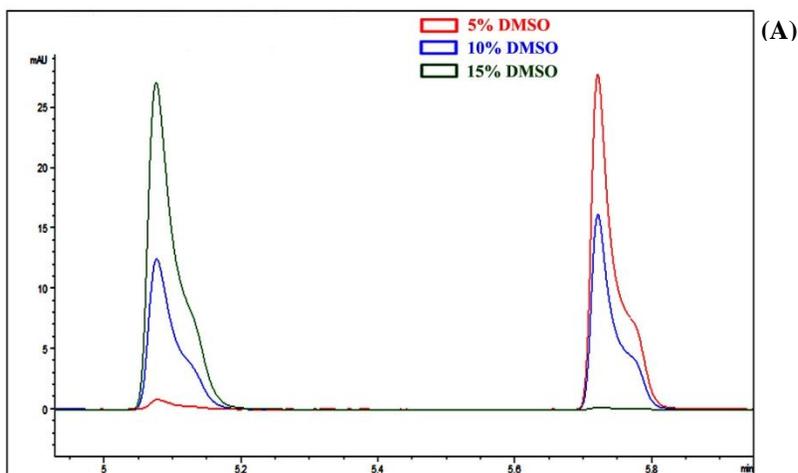


Figure 79. (A). HPLC chromatogram of PyrH chlorination activity in DMSO, (A1). Relative activity of PyrH chlorination activity in DMSO, (B)HPLC chromatogram of PyrH chlorination activity in EtOH, (A1). Relative activity of PyrH chlorination activity in EtOH

3.5.3. PyrH Substrate specificity

Most enzymes are able to efficiently use only a small number of compounds as substrates for catalysis. It has been demonstrated previously that the substrate specificity can be altered by protein engineering without affecting its activity^{108,109}. Site directed mutagenesis of amino acids in the active site of the enzyme can yield mutant proteins with altered substrate specificity. Most of these mutations are carried out based on the existing crystal structure of the enzyme. A wide range of aromatic and indole based/derived compounds were selected to study the substrate specificity of PyrH (Figure 80). The compounds were selected on the basis of similarity to tryptophan such as kynurenine. Some smaller aromatic compounds such as anthranilamide and aniline were also selected.

3.5.3.1. Halogenase activity of PyrH with tryptophan

PyrH activity with tryptophan as the substrate was assayed as described in section 5.22. A no enzyme control sample of the assay was analysed alongside the reaction. The reaction mixtures and tryptophan standard were separated by HPLC with detection at 280 nm. Tryptophan had a retention time of 5.1 min. An additional peak with a retention of time 5.7 min was also observed in the enzymatic reaction mixture. This peak was absent in the no enzyme control suggesting that this was chlorinated tryptophan (Figure 81). This peak was analysed by LC-MS to obtain the mass of the product. Chlorine has two major isotopes of ³⁵Cl⁻ (75.78%) and ³⁷Cl⁻ (24.22%) in relative proportions of ~3:1 respectively. The LC-MS analysis revealed two (M+H) ions at m/z of 239.1 and 241.1 in a 3:1 ratio for mono chlorinated tryptophan (Figure 81).

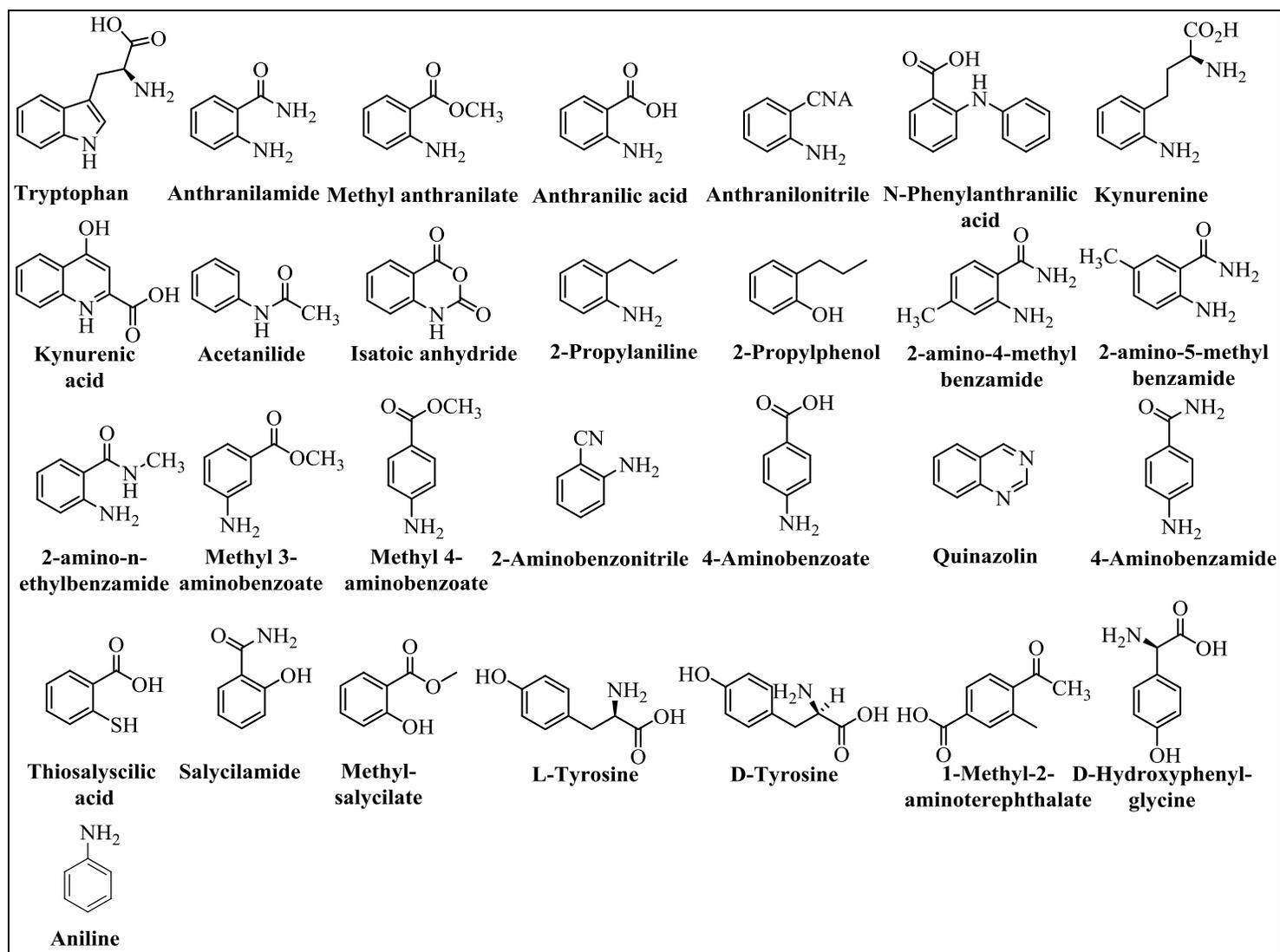


Figure 80. Substrate library screened with purified flavin dependent halogenases and mutants

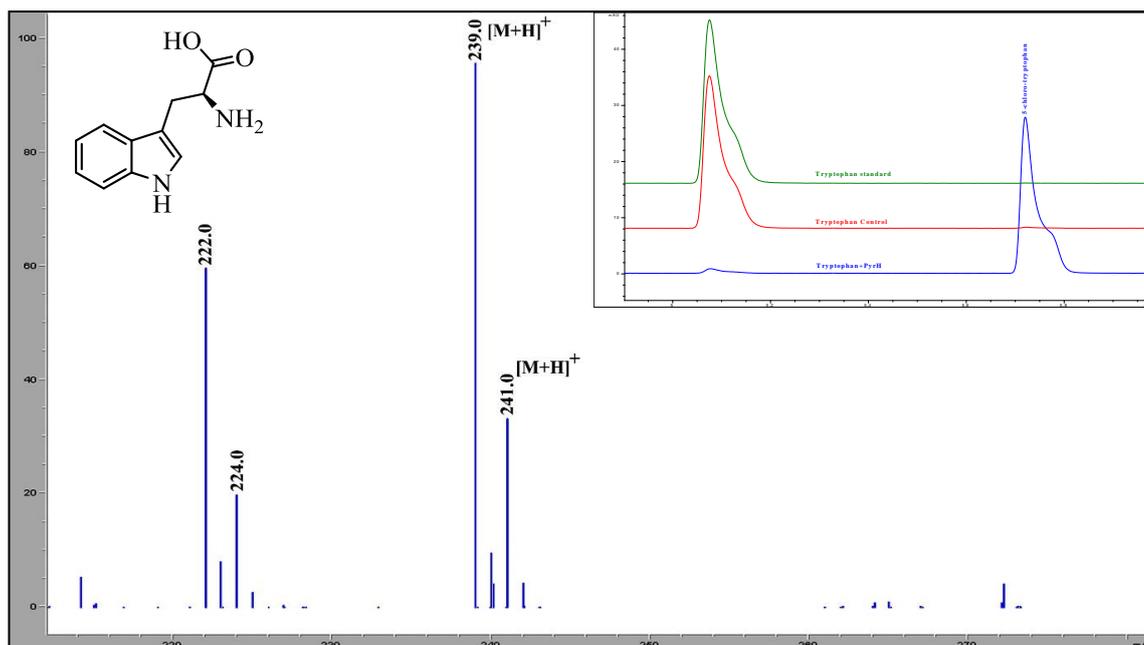


Figure 81. HPLC chromatogram and mass spectrum of tryptophan chlorinated by PyrH with a mass of 239.0 and 241.0

Chlorinated product was purified for NMR using C-18 bond elute column as described in section 5.26. The ^1H NMR data showed the chlorination of tryptophan at the 5-position (Figure 82).

The regioselectivity was also confirmed by 2D NMR. Both tryptophan and chlorinated tryptophan were analysed by COSY experiment (Figure 83). A COSY spectrum shows protons which are coupled to each other. The tryptophan showed coupling of protons at all positions of the aromatic ring while the chlorinated reaction product did not show coupling at the 5-position of tryptophan indicating it to be the chlorinated position.

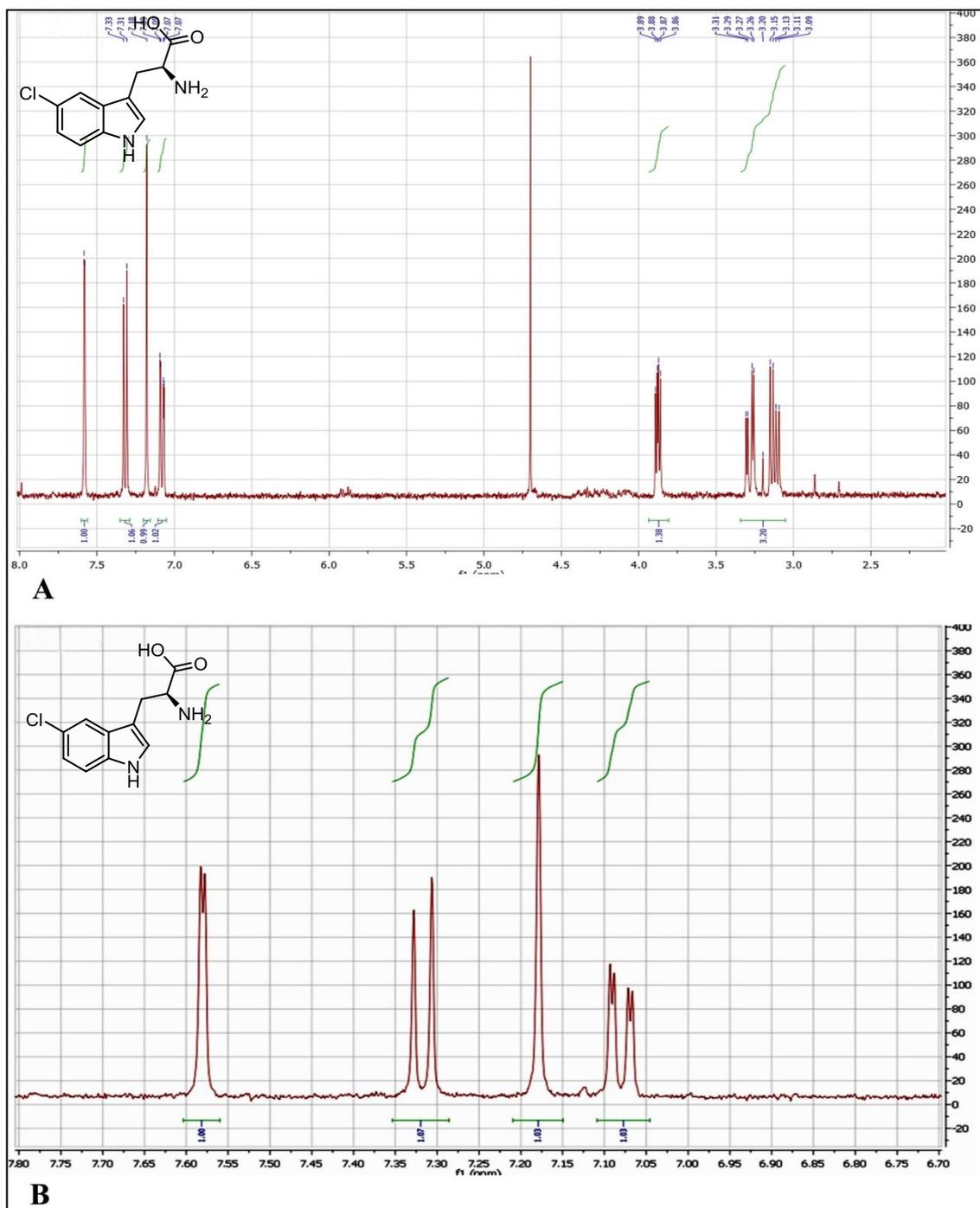


Figure 82. ^1H NMR data showing chlorination of tryptophan at the 5-position by PyrH with the complete ^1H NMR spectrum shown at the top and an enlarged aromatic region shown at the bottom of the figure.

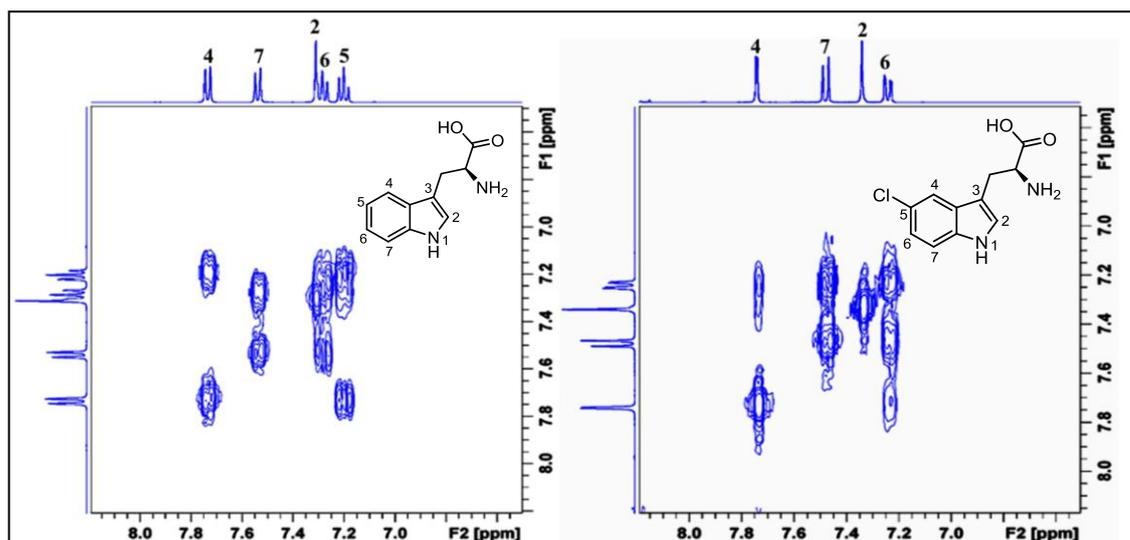


Figure 83. COSY spectra of tryptophan (left) and 5-chlorinated-tryptophan (right)

Tryptophan bromination was also explored using the same enzyme assay. 12.5 mM of NaBr was used as source of bromine in these reaction mixtures instead of MgCl₂. Analysis by HPLC showed a peak for the starting material (tryptophan) at a retention time of 5.1 min and a product peak (brominated tryptophan) at a retention time of 5.7 min. The product peak was absent from both the tryptophan standard and no enzyme control runs. The peak at retention time 5.7 min was analysed by LC-MS to determine the mass of the brominated product. Bromine has two isotopes of ⁷⁹Br and ⁸¹Br with a relative proportion of 50.69% and 49.31% respectively. The LC-MS analysis showed two (M+H) ions at m/z of 283.0 and 285.0 for brominated tryptophan in a 1:1 ratio indicating it is a mono brominated product. The ¹H NMR data confirmed the bromination of tryptophan at the 5-position.

In tryptophan, chemical halogenation by EAS will be most preferred at position 3 but since the position is occupied by extending amino acid chain, the next preference of halogenation will be at position 2. In enzyme system, position 3 and 2 is blocked in the active, allowing the next favourable halogenation of the substrate by EAS to occur on indole ring of the tryptophan which is favoured by PyrH to halogenate regioselectively at the 5-position.

3.5.3.2. Halogenase activity of PyrH with kynurenine

Activity of PyrH with kynurenine as a substrate was assayed as per section 5.22. A no enzyme control was included. The samples were analysed by HPLC. A kynurenine standard had a retention time of 3.1 min. An additional peak at 5.55 min was observed for the reaction mixture. This peak was not observed in the control (Figure 84). This peak was analysed by LC-MS and revealed mass (M+H) ions at m/z of 243.1 and 245.1 in a 3:1 ratio indicating that the kynurenine has been mono chlorinated (Figure 84). Chlorinated kynurenine was purified for NMR analysis on a C-18 bond elute column as per section 5.26. Analysis by ^1H NMR and COSY confirmed chlorination at the 5-position of kynurenine.

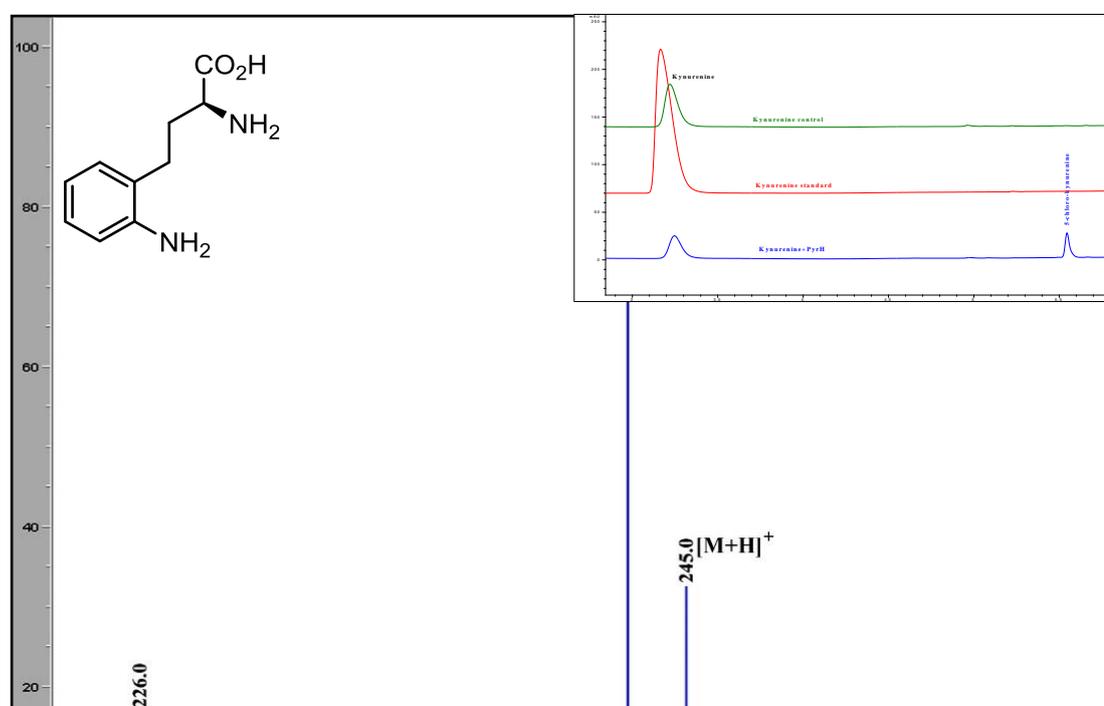


Figure 84. HPLC chromatogram and mass spectrum of kynurenine chlorinated by PyrH showing masses of 243.0 and 245.0

The ability of PyrH to brominate kynurenine was tested in a similar assay. The reaction mixture was supplemented with 12.5 mM of NaBr as a source of bromine instead of 12.5 mM MgCl_2 . An additional peak at a retention time of 5.5 min was detected in the reaction mixture when analysed by HPLC. A kynurenine standard was detected at a retention time of 3.1 min and no additional peak was observed in the no enzyme control experiment.

Analysis by LC-MS revealed two (M+H) ions at m/z of 287.0 and 289.0 for bromine in a 1:1 ratio indicative of mono brominated kynurenine. Brominated kynurenine was purified for NMR analysis on a C-18 bond elute column as per section 5.26. Analysis by ^1H NMR confirmed bromination occurred at the 5-position of kynurenine.

In kynurenine, the presence of NH_2 group favours more as ortho, para directing group over the extended amino acid chain (also an ortho, para directing group). In chemical halogenation, either position 3 or 5 are the preferred sites to carry out EAS. The binding of the substrate could be in such a way that the 5 position lies closer to the glutamic acid that causes rearomatisation allowing PyrH to regio-selectively halogenate at the 5 position. In chemical halogenation, both the mixture of 3 and 5 would be expected, while PyrH has advantage of producing only one halogenated compound more selectively.

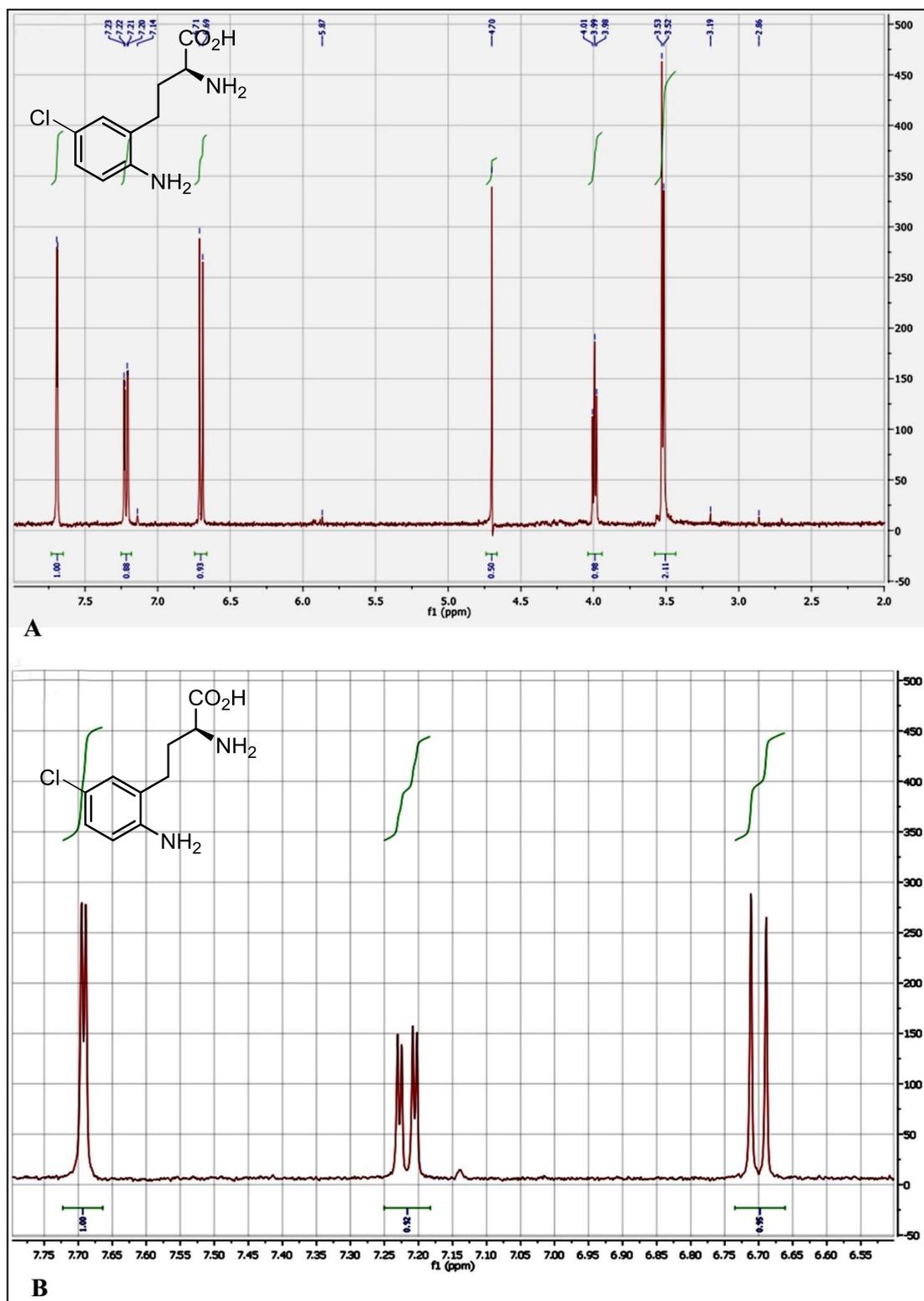


Figure 85. ^1H NMR data showing chlorination of kynurenine at the 5-position by PyrH with the complete ^1H NMR spectrum shown at the top and an enlarged aromatic region shown at the bottom of the figure.

3.5.3.3. Halogenase activity of PyrH with anthranilamide

PyrH activity with anthranilamide as a substrate was assayed as described in section 5.22. A no enzyme control sample of the assay was also analysed. The reaction mixtures and an anthranilamide standard were analysed by HPLC with detection at 280 nm. Anthranilamide had a retention time of 3.6 min and an additional peak with a retention time of 6.5 min was observed for the enzyme assay. This peak was absent in the no enzyme control suggesting that this was chlorinated anthranilamide (Figure 86). The LC-MS analysis revealed two (M+H) ions at m/z of 171.0 and 173.0 in a 3:1 ratio for mono chlorinated anthranilamide with a neutral loss of ammonia (mass-17) (Figure 86).

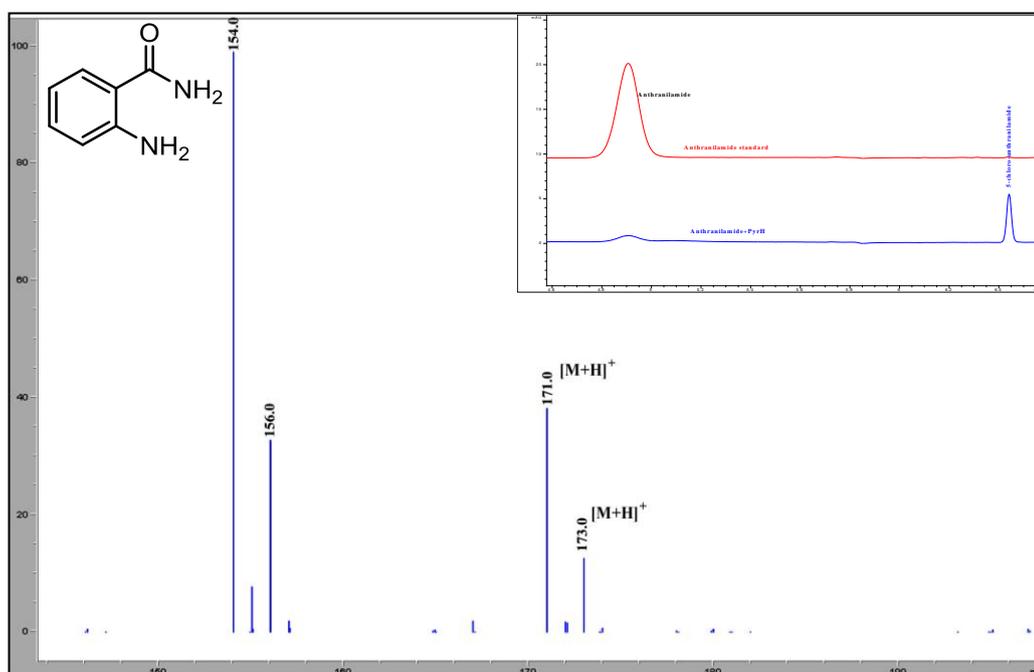


Figure 86. HPLC chromatogram and mass spectrum of anthranilamide chlorinated by PyrH with a mass of 171.0 and 173.0 with a neutral loss of ammonia

Chlorinated product was purified for NMR using C-18 bond elute column as described in section 5.26. The ^1H NMR data confirmed the chlorination of anthranilamide at the 5- position (Figure 87). This was also confirmed by 2D NMR COSY.

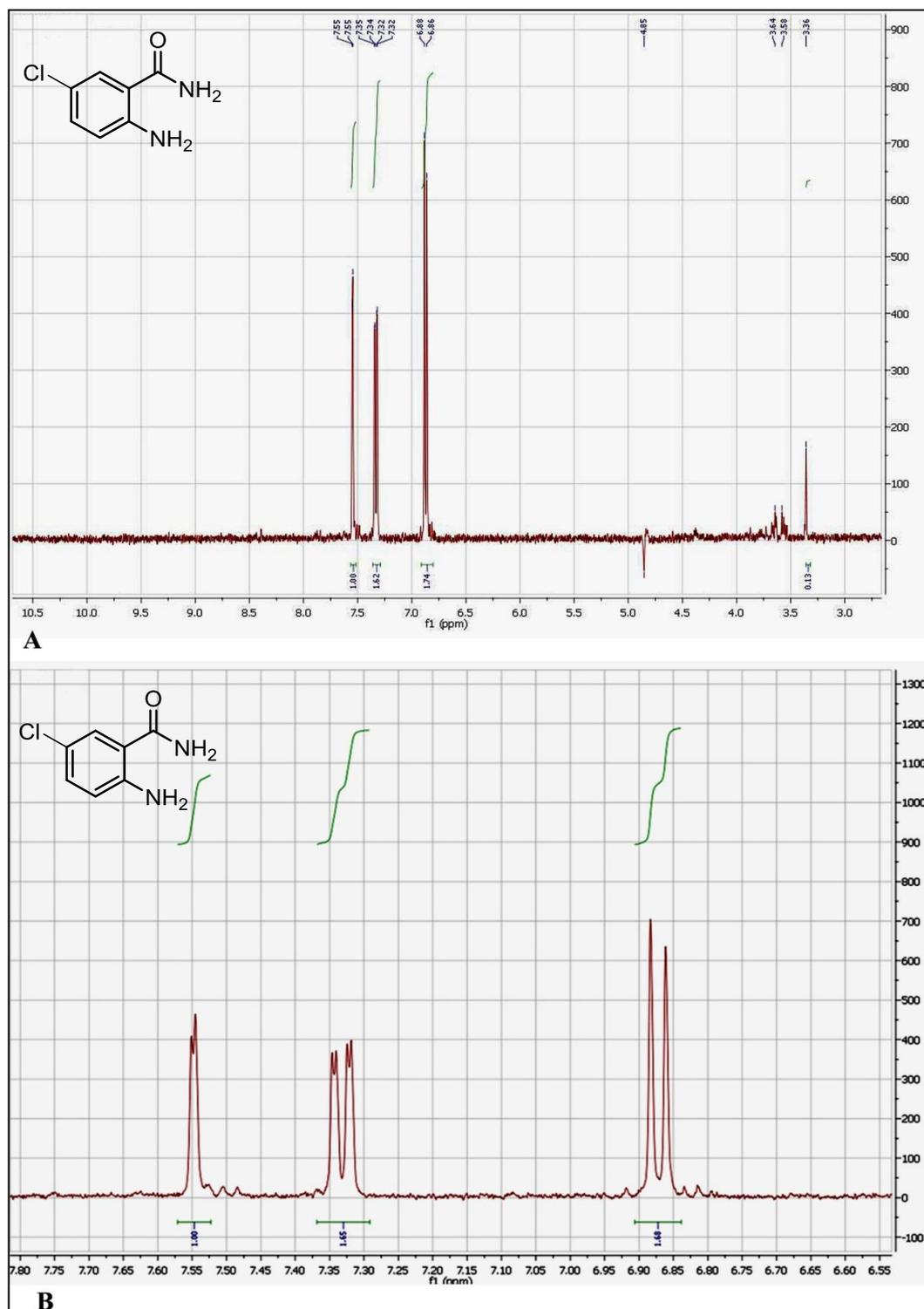


Figure 87. ^1H NMR data showing chlorination of anthranilamide at the 5-position by PyrH with the complete ^1H NMR spectrum shown at the top and an enlarged aromatic region shown at the bottom of the figure.

In anthranilamide, NH₂ group favours more ortho, para directing group over the CONH₂ group (meta directing group) preferring either 3 or 5 position for halogenation. The position 3 is ortho to para directing and meta to meta directing while position 5 is para to ortho, para directing group and meta to meta directing group. Chemical halogenation would produce mixture of both 3 and 5 halogenated anthranilamide on EAS but PyrH halogenates more selectively at the 5 position.

3.5.3.4. Halogenase activity of PyrH with 2-amino-4-methylbenzamide

Activity of PyrH with 2-amino-4-methylbenzamide as substrate was assayed as described in section 5.22. A no enzyme control was included. The samples were analysed by HPLC. A 2-amino-4-methylbenzamide standard had a retention time of 5.5 min. An additional peak at 5.98 min was observed in the reaction mixture. This peak was not observed for the no enzyme control (Figure 88). Analysis by LC-MS detected mass (M+H) ions at an m/z of 185.0 and 187.0 in a 3:1 ratio indicating mono chlorinated 2-amino-4-methylbenzamide (Figure 88).

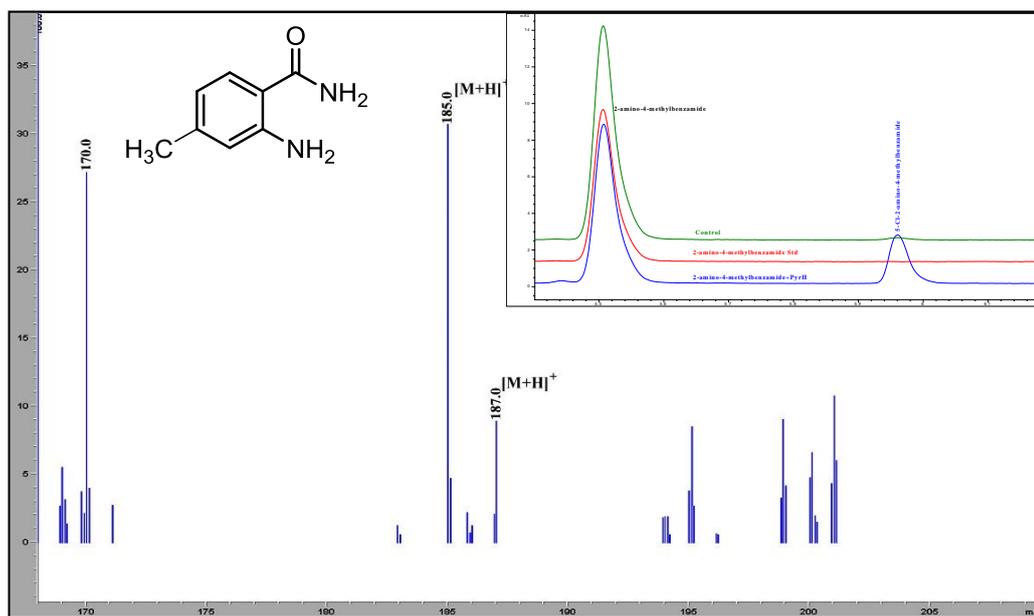


Figure 88. HPLC chromatogram and mass spectrum of 2-amino-4-methylbenzamide chlorinated by PyrH with mass of 185.0 and 187.0

Chlorinated 2-amino-4-methylbenzamide was purified for NMR analysis on a C-18 bond elute column as per section 5.26. Analysis by ¹H NMR showed

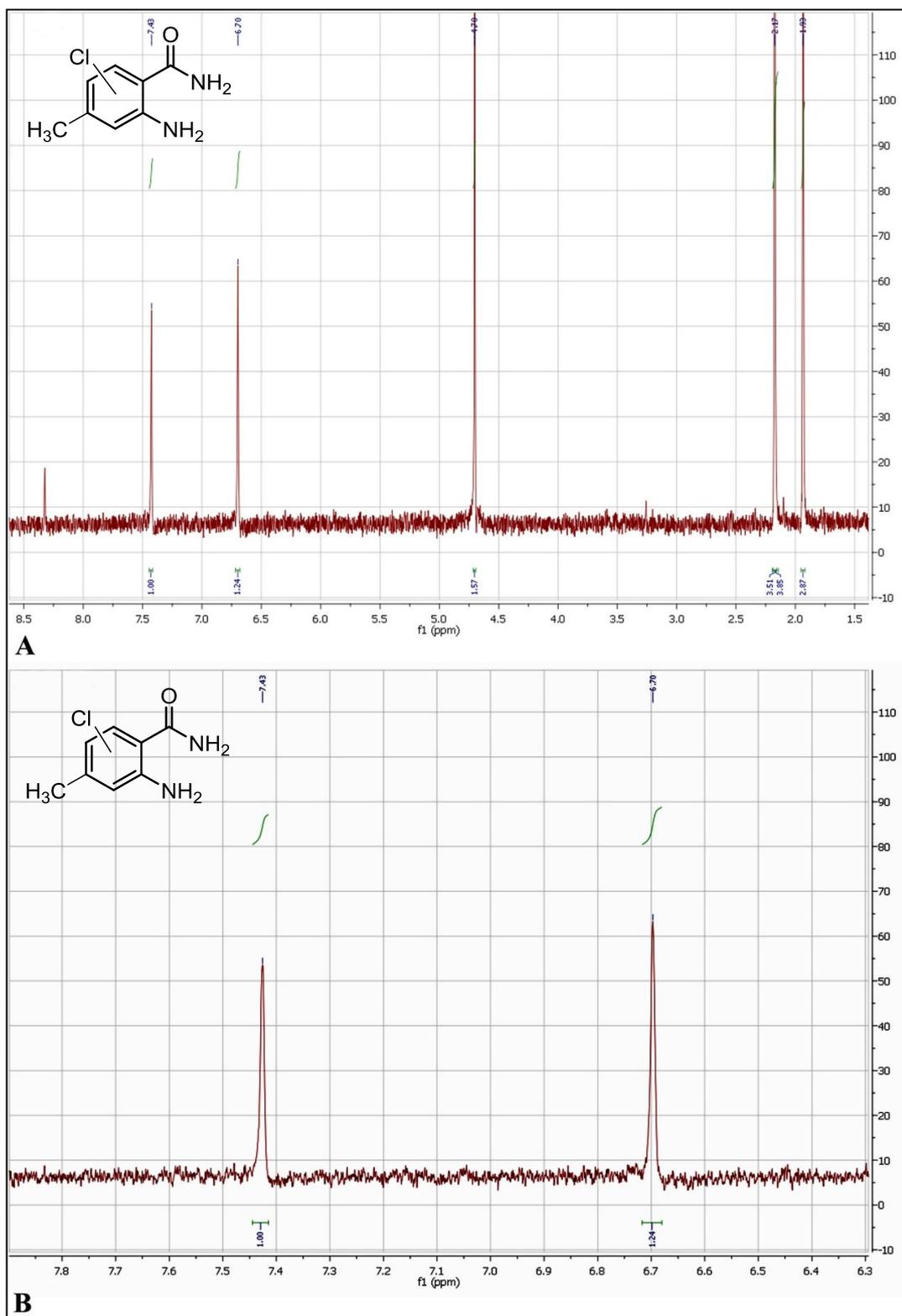


Figure 89. ^1H NMR data showing chlorination of 2-amino-4-methylbenzamide by PyrH with the complete ^1H NMR spectrum shown at the top and an enlarged aromatic region shown at the bottom of the figure

a chemical shift of the aromatic peaks between 6.5 and 7.7 ppm. From the coupling of protons in ^1H NMR, it was difficult to know the exact position of chlorination. The coupling of H protons showed two singlet's, one of which was observed for position 3 while the other singlet's could be either for position 5 or 6 (Figure 89),but the electronics of the ring substituents could direct the chlorination to the 5 position.

2D ROESY NMR was carried out for 2-amino-4-methylbenzamide and the chlorinated 2-amino-4-methylbenzamide sample. ROESY NMR shows the coupling of protons through space. The 2-amino-4-methylbenzamide standard showed coupling of the proton at the 3-position and the proton at the 5-position of the aromatic ring with the methyl group. In chlorinated 2-amino-4-methylbenzamide only the singlet at the 3-position could be observed and the 5-position was observed to be chlorinated which showed no doublet of protons associated with the methyl group (Figure 90).

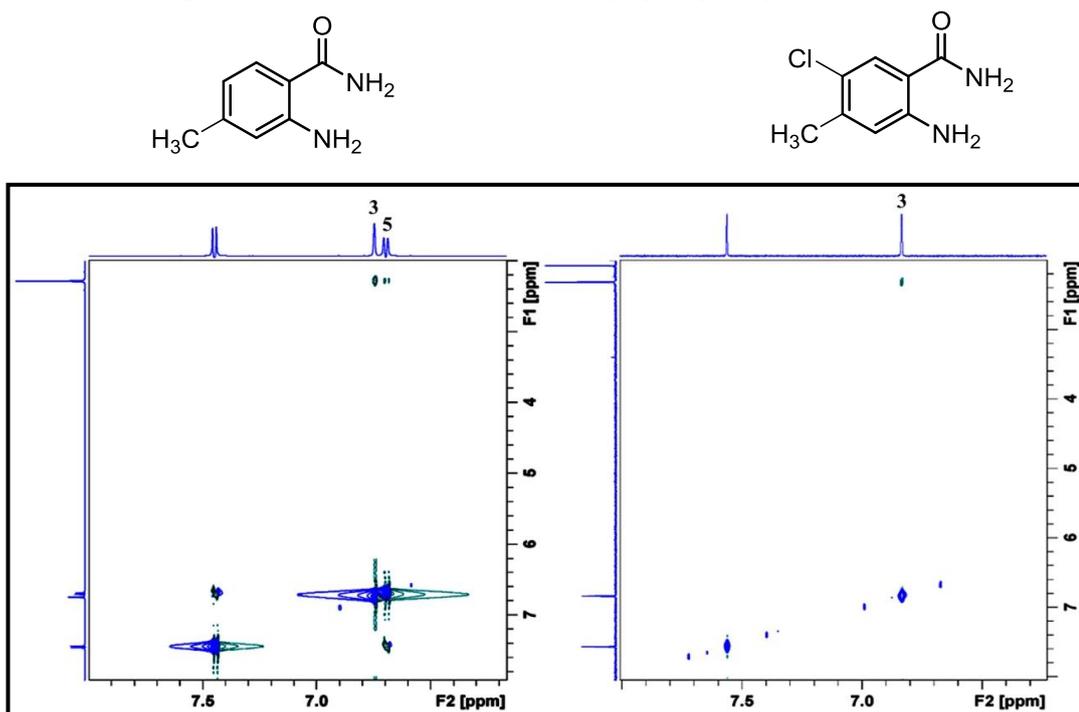


Figure 90. ROESY NMR data showing chlorination of 2-amino-4-methylbenzamide at the 5-position by PyrH

In general while carrying out halogenation chemically on 2-amino-4-methylbenzamide, position 3 would be much preferred position for halogenation since it is ortho to ortho, para directing groups (NH_2 and CH_3

respectively) and meta to meta directing group (CONH₂). While position 5 will also be preferred because it is ortho to CH₃ (ortho, para directing group), para to NH₂ group (ortho, para directing group) and meta to CONH₂ (meta directing group). PyrH halogenates more selectively halogenates at the 5 position neglecting 3 position to be halogenated in the enzyme system.

3.5.3.5. Overall relative activity of PyrH with the substrates

PyrH assays were carried out with various substrates in triplicate to obtain relative activities (RA). The RA was calculated from HPLC analysis based on the % peak area of the substrate relative to that of the product. As shown in Table 11, the RA for tryptophan as a substrate was 100% for chlorination, while the RA for bromination was only 9%. The relative activities for the structurally similar substrate kynurenine were 83% for chlorination and 29% for bromination. The relative activities were much lower for the substrate anthranilamide with 54% for chlorination and 18% for bromination. The relative activities with anthranilamide-like substrates such as 2-amino-4-methylbenzamide, anthranilic acid, methyl anthranilate 21.9%, 2.0% and 1.3% for chlorination respectively. Finally, the relative activity for chlorination of the substrate N-phenyl anthranilic acid was only 0.8%.

Table 11. Overall relative activity of halogenation by PyrH on various substrates

Substrates	Relative activity (%)	
	Chlorination	Bromination
Tryptophan	100.0	100.0
Kynurenine	81.0	29.0
Anthranilamide	54.0	18.0
2 amino -4-methylbenzamide	21.9	-
Anthranilic acid	2.0	-
Methyl anthranilate	1.3	-
N-phenyl anthranilic acid	0.8	-

The docking of kynurenine was carried out using the solved crystal structure of PyrH bound to tryptophan (PDB ID: 2WEU) as a template in PYMOL software¹. The docking (Figure 91) showed that the bigger substrate kynurenine (similar to tryptophan) occupies the active site tightly in such a way that there are minimal changes in the interactions between the protein and substrate. The difference in interactions between kynurenine and tryptophan are tabulated in Table 12.

Table 12. Differences in interaction distance between tryptophan and kynurenine by docking experiments

Interactions	Tryptophan	Kynurenine	Difference in interaction
Carboxyl group of F451 residue with the amino terminal of	6.05 Å	6.91 Å	0.90 Å away
Amino group of Q163 with carboxyl group	2.87 Å	5.07 Å	2.2 Å away
Carboxyl group of S50 to carboxyl group	2.56 Å	3.26 Å	0.6 Å away
Amino group of S50 to carboxyl group	2.84 Å	2.31 Å	0.53 Å closer

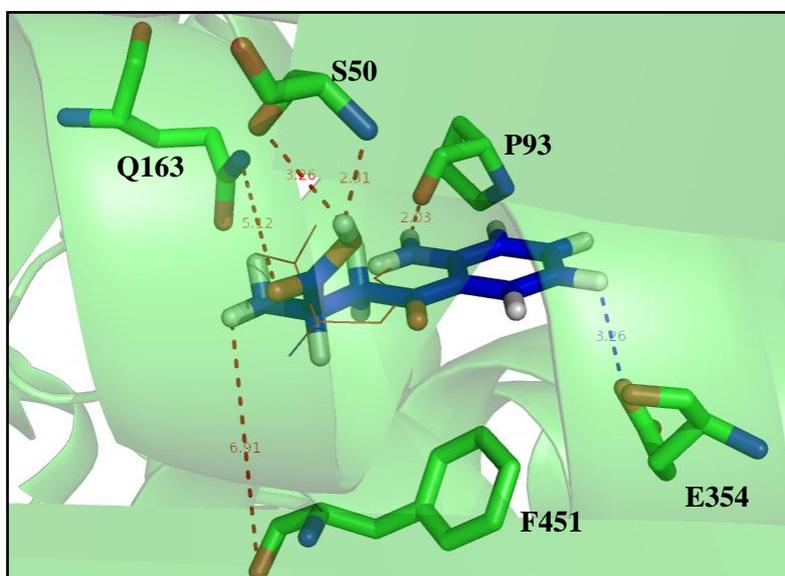


Figure 91. Docking experiment showing the distance interaction of kynurenine in the PyrH(PDB ID:2WEU) active site(red line-tryptophan, blue stick-kynurenine) generated using PYMOL¹

Table 13. Differences in interaction distance between tryptophan and 2-amino-4-methylbenzamide by docking experiments

Interactions	Tryptophan	2-amino-4-methyl benzamide	Difference in interaction
Carboxyl group of F451 residue with the amino terminal of	6.05 Å	-	-
Amino group of Q163 with carboxyl group	2.87 Å	4.68 Å	1.81 Å away
Carboxyl group of S50 to carboxyl group	2.56 Å	5.66 Å	3.10 Å away
Amino group of S50 to carboxyl group	2.84 Å	5.27 Å	2.43 Å closer

Similarly, the docking of 2-amino-4-methylbenzamide was carried out using the PyrH with bound tryptophan crystal structure (PDB ID: 2WEU) in PYMOL¹. The docking was carried out similarly to that of kynurenine. The docking (Figure 92) showed that the smaller substrate 2-amino-4-methylbenzamide occupies the active site in such a way that the interactions at the active site are disrupted. The difference in interactions between tryptophan and 2-amino-4-methylbenzamide are tabulated in Table 13. The active site space is too big for the smaller substrates to bind tightly which might explain the lower relative activity. The interaction between F451 and 2-amino-4-methylbenzamide could not be established due to the presence of a hydroxyl group in the substrate.

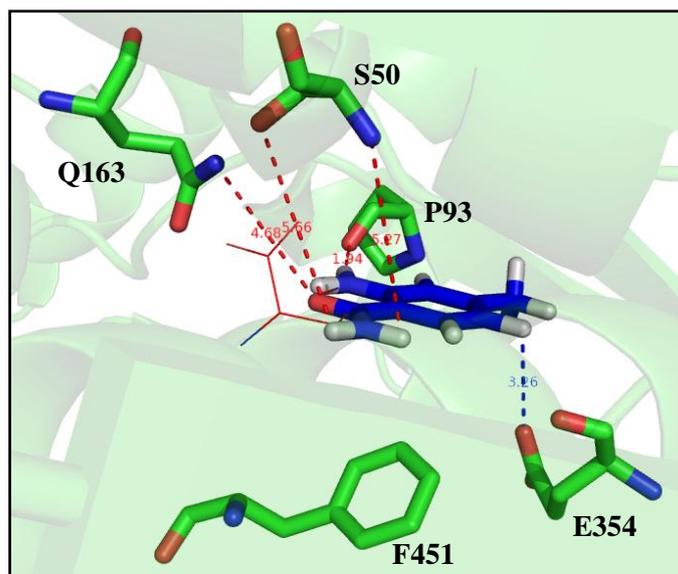


Figure 92. Docking experiment showing the distance interaction of 2-amino-4-methylbenzamide in the PyrH(PDB ID:2WEU) active site(red line-tryptophan, blue stick-2-amino-4-methylbenzamide) generated using PYMOL¹

Based on distance, many of these interactions seem to be very weak. Hence the effect of these interactions on protein-ligand binding is a matter of conjecture. A real picture would emerge when enzyme crystal structures with bound ligands are obtained.

3.6. Regioselective studies on PrnA and PyrH

3.6.1. Engineering PrnA to regioselectivity chlorinate at the 5-position based on PyrH

The regioselectively controlled chlorination of organic compounds is an industrially important chemical process. The study of the regioselectivity of flavin dependent halogenases has been previously discussed (section 1.5.4) based on the studies carried out on PyrH (a tryptophan-5-halogenase) and PrnA (a tryptophan-7-halogenase) by Karl Heinz van Pee group's^{64,88}. The crystal structures obtained by the group predicted that the residues that determine the regioselectivity of the enzymes are those that interact with the substrate through either hydrogen or π stacking. In PrnA, residues F103 and H101 forms a π stacking sandwich around the substrate tryptophan, while

the substrate orientation is maintained by hydrogen bond interactions with residues E450, Y443, Y444, F454 and E346. The indole of the substrate tryptophan is shielded by the F103 and W455 residues. A Lang et al., 2011 mutated the residues W455, F103 and H101 to alanine by site directed mutagenesis and found that the F103A mutant could produce both 7-chlorotryptophan and a low amount of 5-chlorotryptophan. When the same mutant was analysed for bromination they observed higher production of 5-brominated-tryptophan. The other mutants showed slower kinetics indicating that they are important for halogenase activity/substrate binding.

In PyrH the tryptophan binding is very different to that of PrnA. The indole nitrogen of the tryptophan interacts with the carboxyl group of P93 and the hydroxyl group of Y454 positioning the substrate for chlorination at the 5-position. Residues S50, Q163 and F451 also interact with the substrate tryptophan through hydrogen binding. Residues F94 and F49 sandwiches the tryptophan through π stacking interactions. It was observed that residue F49 would allow more space for binding of the tryptophan substrate while Y454 would change the orientation of the substrate, relative to that in PrnA by holding the NH group of the indole. Together, these residues govern the regioselective chlorination at the 5-position of tryptophan-5-halogenases. Zhu et al, 2009 mutated the F49 residue to an A residue and Y454 to an F residue. However they did not observe any change in the regioselectivity of the halogenation.

In an attempt to increase the amount of chlorination that can be achieved at the 5-position of a tryptophan by a tryptophan-7-halogenase, the sequence of tryptophan-7-halogenases was compared to that of 5-chlorotryptophan halogenases (Figure 93). Further analysis by overlaying the crystal structures of both PrnA (PDB ID:2AR8) and PyrH (PDB ID:2WET) revealed that residues I52, P53 and N459 in the active site of PrnA are replaced by F49, S50 and S455 in the active site of PyrH (Figure 94). It is also predicted that residue N459 plays a important role in adopting the conformation of the tryptophan in PrnA⁸⁹.

```

RebH-7-  MSG-KIDKILIVGGGTAGWMAASYLGKALQGTADITLLQAPDIPTLGVGEATIPNLQTA 58
KtzQ-7-  MDDNRIRSIIVLGGGTAGWMSACYLSKALGPGVEVTVLEAPSI SRIRVGEATIPNLHKV 59
PrnA-7-  MNK-PIKNIIVIVGGGTAGWMAASYLVRALQQANITLIESAAIPRIGVGEATIPSLQKV 58
STRHY-5  -----MTASYLRAAFADRVDITVVESRRIGTIGVGEATFSTVVR-H 39
PyrH-5   ----MIRSVVIVGGGTAGWMTASYLKAAFDDRIDVTLVESGNVRRIGVGEATFSTVVR-H 54
          *:*** * : . . . . . : : ***** . . :
RebH-7   LYYGNFEEEFNFWNNSNYCVLAGLGLVPDAPSPRLAHMPQATESVDEVFGAVKDRQN 512
KtzQ-7   HYYGNFEEAEFRNFWTNSNYCIFAGLGLPEHPLPVLEFRPEAVDRAEPVFAAVRRRTEE 514
PrnA-7   TYYETFDYEFKFWLNGNYCIFAGLGMPLDRSLPLLQHRPESIEKAEAMFASIRREAER 501
STRHY-5  PYYHAFEP-----YSYVCMALGLGGIPLRHPPALDLFDT--RAARAELTRVREQARK 484
PyrH-5   PYYHGFET-----YSWITMNLGLGIVPERPRPALLHMDP--APALAEFERLRREGDE 495
          ** : . . : * * : . * . . . . . : : .

```

Figure 93. Sequence alignment of tryptophan 7-halogenases with tryptophan 5-halogenases

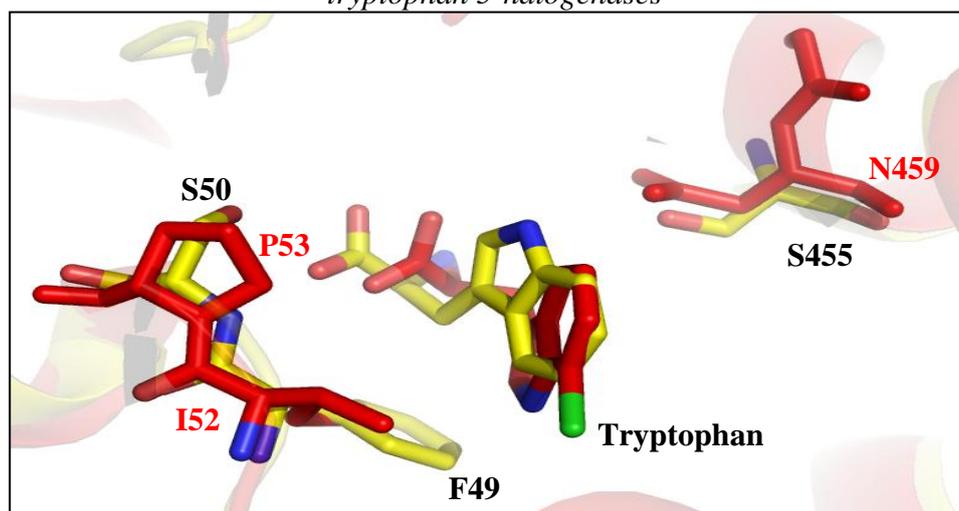


Figure 94. The three important residues selected for mutation studies Red-PrnA (tryptophan 7-halogenase, Yellow-PyrH (tryptophan 5-halogenase)

pET-SAS1 was used as template DNA for site directed mutagenesis (section 5.14.1) using primers P32-P35(Table 32). Clones for the single mutants F103A, I52F, P53S, N459S, (pET-KK26-29) were verified by DNA sequencing (GATC, Germany). All the mutant proteins were expressed in *E. coli* ArcticExpress (DE3) RP cells, purified and concentrated to the following concentrations in a volume of 1.0 ml (Table 14).

Table 14. PrnA mutant protein concentrations obtained for regioselective studies

Mutant	Concentrations mg ml ⁻¹
I52F	3.66
P53S	2.3
N459S	3.9
F103A	0.486

3.6.2. Enzymatic assay of PrnA mutants with tryptophan for regioselectivity

The activities of the PrnA mutant proteins were enzymatically assayed (section 5.22). 5-chloro-tryptophan and 7-chloro-tryptophan were obtained using wild type enzymes PyrH and PrnA, respectively, and were used as standards in this experiment. The PrnAI52F, PrnAP53S and PrnAN549S mutants were less active than the wild type enzyme and showed chlorination of tryptophan at the 7-position only indicating no change in regioselectivity (Figure 95).

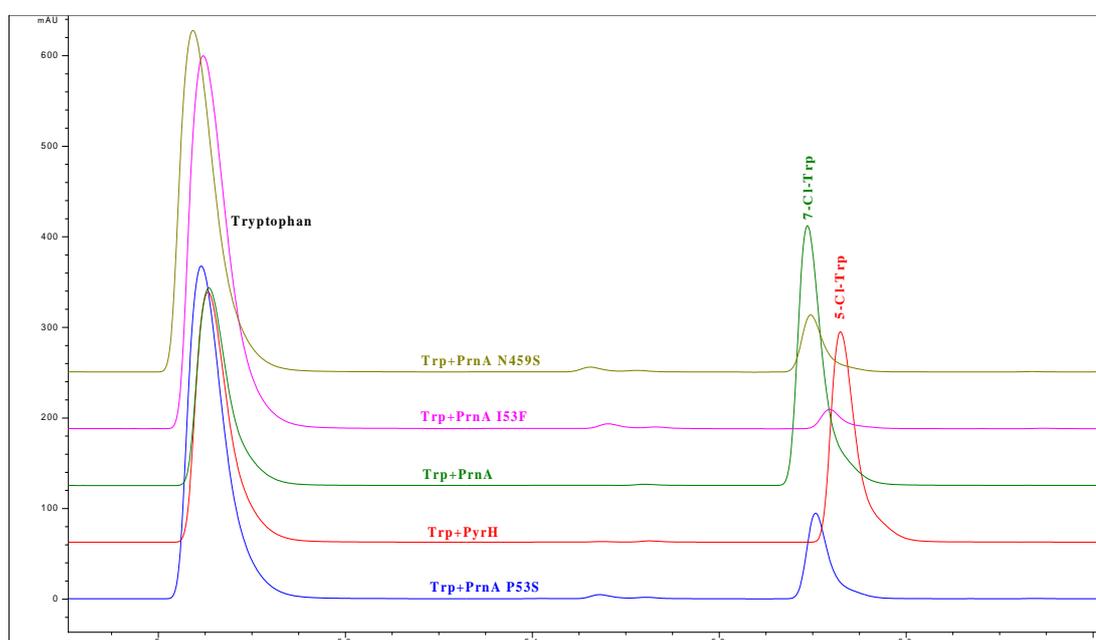


Figure 95. HPLC chromatogram of PrnA mutants with tryptophan for regioselective chlorination

The PrnA mutant proteins were also assessed for their ability to brominate tryptophan. Assays were carried out as described in section 5.22. 5-Br-Trp and 7-Br-Trp were used as standards for detection by HPLC. Similar results were observed to that of chlorination, wherein PrnA I52F, PrnA P53S and PrnAN459S mutants showed bromination of tryptophan but only at the 7-position indicating no switch in regioselectivity (Figure 96).

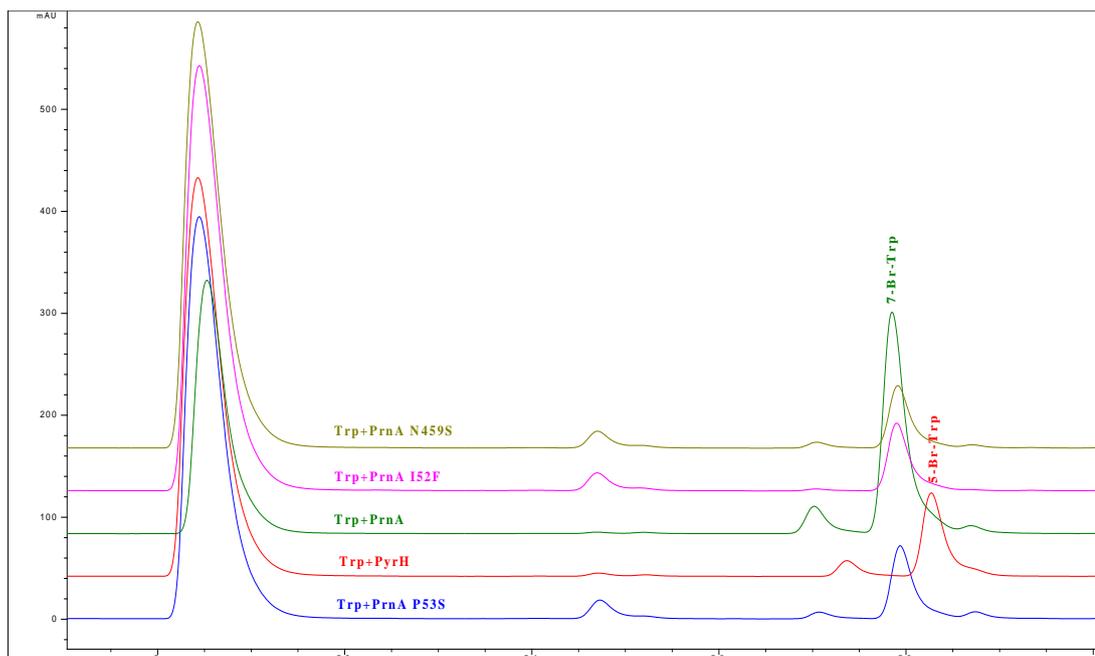


Figure 96. HPLC chromatogram of PrnA mutants with tryptophan for regioselective bromination

Overall the mutants constructed to attempt to change the regioselectivity of PrnA tryptophan-7-halogenase to a tryptophan-5-halogenase did not cause any change in the regioselectivity of halogenation. Also the relative activities of the mutants were lower than for the wild-type enzyme. Relative activities were found to be highest for chlorination (14.84%) than for bromination (11.67%).

Table 15. Overall relative activity of the PrnA mutants obtained for regioselective studies

Mutants	Relative activity	
	Chlorination	Bromination
I52F	3.30	9.38
P53S	14.84	11.22
N459S	10.11	11.67

3.6.3. Engineering PyrH to chlorinate 7-position based on PrnA

To switch the regioselectivity of PyrH from the 5-position to the 7-position, the sequence was compared to that of other flavin dependent halogenases which are known to halogenate at the 7-position. As shown in Figure 94, analysis by overlaying the crystal structures of both PyrH and PrnA revealed

that the residues F49, S50 and S455 in the substrate binding pocket of PyrH are replaced by I52, P53 and N459 residues in PrnA.

Site directed mutagenesis was carried out to switch these amino acids in PyrH to those found in the tryptophan-7- halogenases and therefore determine if they influence the regioselectivity of the enzyme. pET-KK25 was used as template DNA for site directed mutagenesis (section 5.14.1) using primers P36-P40 (Table 32). Mutant clones (pETKK30-34) were verified by DNA sequencing (GATC, Germany). The PyrH mutants, F49I, S50P, S455N, double mutants and triple mutants were expressed in *E. coli* ArcticExpress (DE3) RP cells, isolated, purified and concentrated (Table 16) as described in section 5.15.

Table 16. Protein concentration of PyrH mutants for switching regioselectivity

S.No	Mutants	Protein (mg ml ⁻¹)
1	F49I	0.44
2	S50P	1.21
3	S455N	4.72
4	F49I/ S50P	1.18
5	F49I/ S50P/S455N	0.24

3.6.4. Enzyme assays of single, double and triple mutant for regioselectivity

The activities of the PyrH mutant proteins were enzymatically assayed as per section 5.22. The mutants S50P, double mutant F49I/S50P and triple mutant failed to show any chlorination peak for tryptophan indicating that they are inactive. However, the HPLC trace of the F49I and S455N mutants provided evidence that these proteins were active but showed chlorination of tryptophan at the 5-position only, showing no change in regioselectivity (Figure 97). The relative activity of the mutants are tabulated in Table 17.

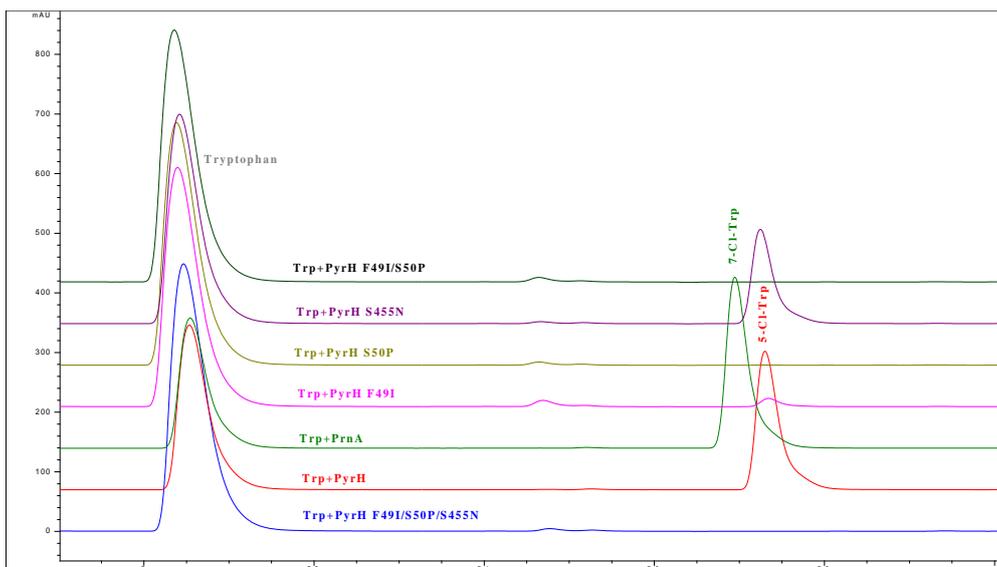


Figure 97. HPLC chromatogram of chlorination activity of the PyrH mutants with tryptophan for regioselectivity

The PyrH mutant proteins were also assessed for their ability to brominate tryptophan following section 5.22. Bromination of tryptophan, albeit at low relative activity could be observed with all the mutant proteins. However no change in regioselectivity was observed for any of the mutants (Figure 98).

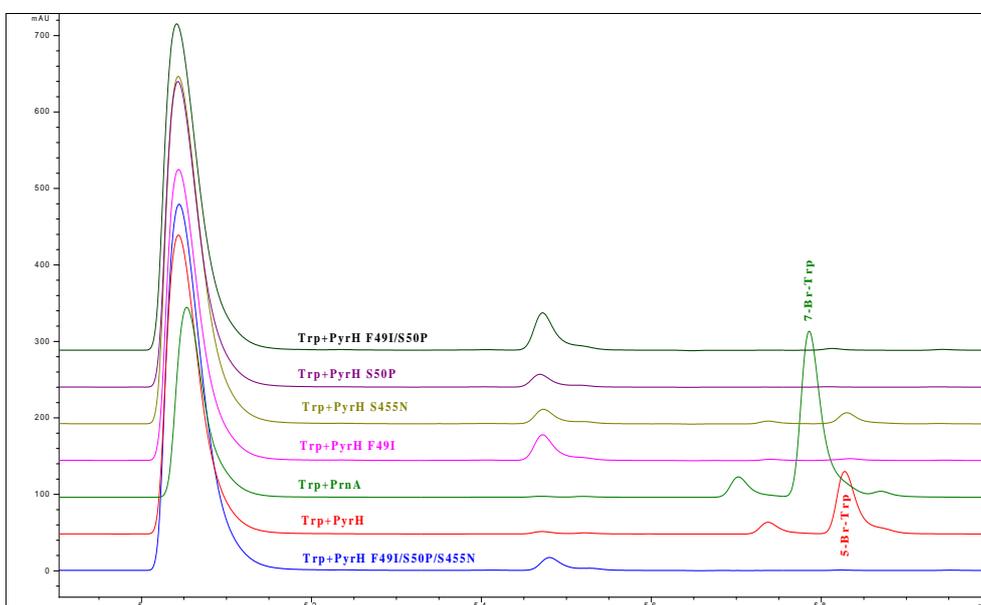


Figure 98. HPLC chromatogram of bromination activity of the PyrH mutants with tryptophan for regioselectivity

Taken together these data indicated that the mutations F49I, S50P and S455N explored did not have any effect on the regioselectivity of the tryptophan-5-halogenase PyrH although they did affect the relative activity of the enzyme in terms of both chlorination and bromination.

Table 17. Relative activity of PyrH mutants with tryptophan for halogenation

Mutant	Relative activity %	
	Chlorination	Bromination
F49I	2.1	0.6
S50P	-	-
S455N	23.8	1.9
F49I/ S50P	-	0.3
F49I/ S50P/ S455N	-	0.4

the above results did not show any change in regioselectivity, further mutagenesis was carried out to investigate other potential residues governing regioselectivity. Further analysis by sequence alignment (Figure 99) revealed presence of an 8 amino acid loop region in tryptophan-7-halogenases, such as PrnA, close to substrate binding site. Few of these residues interact directly with the substrate in the active site (Figure 100). This loop region was found to be absent in tryptophan-5-halogenase. The effect of this loop on the regioselectivity was investigated by deleting the 8 amino acid loop from PrnA.

```

RebH-7      LYYGNFEEFRNFWNNSNYICVLAGLGLVPDAPSPRLAHMPQATESVDEVFGAVKDRQRN 512
KtzQ-7      HYYGNFEAEFRNFWTNSNYICIFAGLGFLPEHPLPVLEFRPEAVDRAEPVFAAVRRRTEE 514
PrnA-7      TYYETFDYEFKNFWLNGNYICIFAGLGMLPDRSLPLLQHRPESIEKAEAMFASIRREAER 501
STRHY-5     PYYHAFEP-----YSYVCMALGLGGIPLRHPPALDLFDT--RAARAELTRVREQARK 484
PyrH-5     PYYHGFET-----YSWITMNLGLGIVPERPRPALLHMDP--APALAEFERLRREGDE 495
          ** :                .: : * * :      . *      . : : .

```

Figure 99. Sequence alignment of tryptophan-5-halogenases with tryptophan-7-halogenases (yellow-loop region)

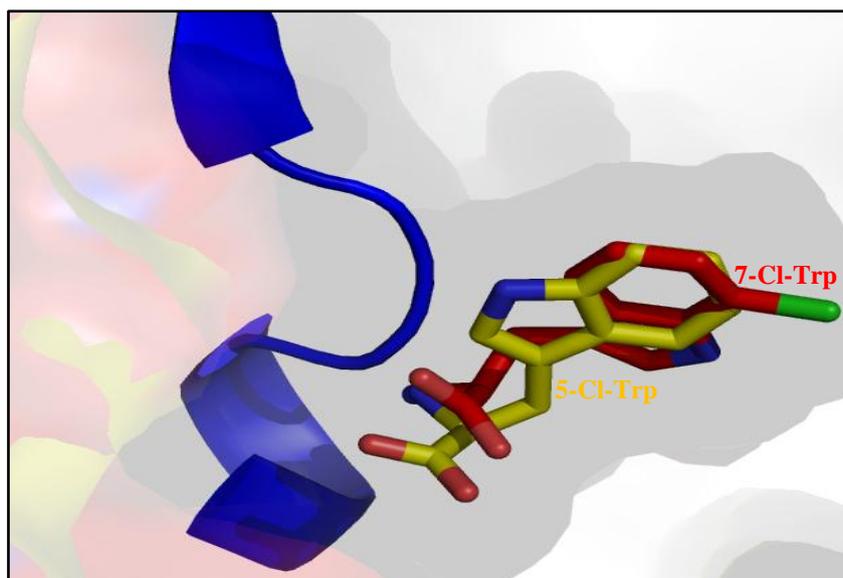


Figure 100. The 8 amino acid loop present in PrnA (Blue) is absent in PyrH, Red-7-chloro-tryptophan, yellow- 5-chloro-tryptophan

Primers P41 and P42 were used for site directed mutagenesis using pET-SAS1 as the template. The mutant clone (pET-KK35) was verified by DNA sequencing (GATC, Germany). The PrnA loop deleted mutant was expressed in *E. coli* ArcticExpress (DE3) RP cells, isolated, purified and concentrated to 800 μ l with a final concentration of 1.72 mg ml⁻¹.

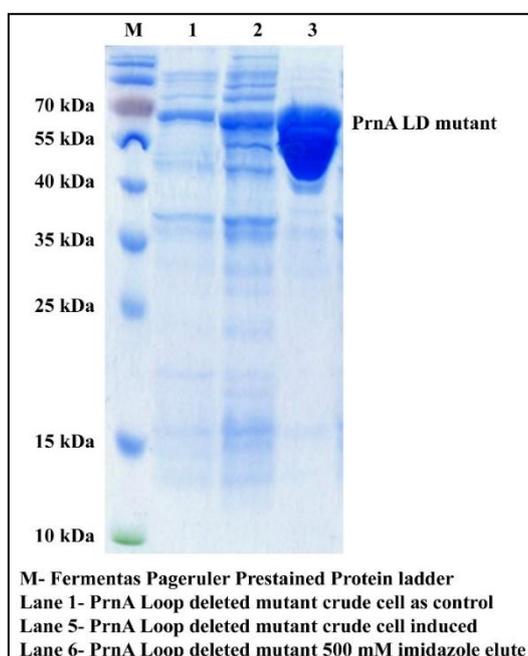


Figure 101. Analysis of the PrnA loop deleted mutant protein by 12% (v/v) SDS-PAGE

3.6.5. Enzyme assay of PyrH loop inserted mutant for regioselectivity

The activity of the PrnA loop deleted mutant protein was assayed following section 5.22. The HPLC trace of the PrnA loop deleted mutant showed no chlorination of tryptophan thus indicating that this loop is important for retaining the activity of enzyme and the structure at the active site. Similar results were obtained when the PrnA loop deleted mutant was assayed for bromination with NaBr as the halogen source with no halogenation observed.

3.6.6. Insertion of 8 the amino acid loop in PyrH for regioselectivity

As described in section 5.14.1, in PyrH, the effect of the loop on the regioselectivity was investigated by inserting an 8 amino acid loop from the 7-halogenases into PyrH and testing to see if it would be able to chlorinate tryptophan at the 7-position.

Site directed mutagenesis was carried out with pET-KK25 as the template and primers P43-P44 (Table 32). The mutant clone(pET-KK36) was verified by DNA sequencing (GATC, Germany) and was expressed in *E. coli* ArcticExpress (DE3) RP cells, isolated, purified and concentrated to 0.24mg ml^{-1} as described in section 5.15 (Figure 102).

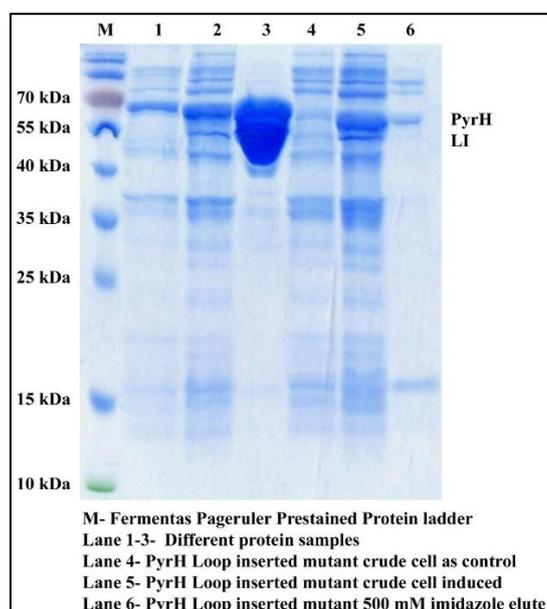


Figure 102. Analysis of pure PyrH loop inserted mutant by 12% (v/v) SDS-PAGE

3.6.7. Enzyme assay of PyrH loop inserted mutant for regioselectivity

The activity of the PyrH loop inserted mutant protein was enzymatically assayed (section 5.22). The HPLC trace of the PyrH loop insert mutant showed no chlorination of tryptophan. This implies that the insertion of the loop is affecting the structure of the substrate binding pocket in the active site. Similar results were obtained when the PyrH loop inserted mutant was assayed for bromination with NaBr as the halogen source with a very low level of halogenation of 0.416% detected.

CHAPTER 4. SUMMARY AND FUTURE WORK

Halogenation is an important process in the functionalisation of industrially important compounds. Chemical halogenation has been widely used to produce halogenated compounds which often results in unwanted by-products causing pollutants in the environment. The method is also expensive, laborious and often lacks regio-control. Bio-halogenases may solve the above issues in producing much greener halogenated products. Enzymes are being used in producing halogenated products and recent identification of flavin dependent halogenases have brought about regio-controlled halogenation of the compounds. The mechanism governing the regio-controlled halogenation of tryptophan at the 7-position by PrnA, RebH and the 5-position by PyrH have been solved using protein crystallography.

To widen the substrate scope of halogenating enzymes, KtzQ and KtzR (tryptophan 6 and 7-halogenases) were cloned and expressed for the first time in *E. coli* Arctic Express (DE3) cells and were obtained in a soluble form. The repeated trials of halogenation by KtzQ was un-successful with tryptophan while KtzR had a very low relative activity. Halogenation reactions performed for other substrates with KtzR resulted in no relative activity. Simultaneously, optimisation of protein purification of KtzR was performed for obtaining a crystal structure of a tryptophan-6-halogenase. Pure KtzR was obtained at 12 mg ml⁻¹ and repeated crystal trials failed to form crystals. Since both of the above enzymes were inefficient, we explored other tryptophan halogenases. PrnA, SttH and PyrH (tryptophan 7-, 6- and 5-halogenases) were cloned and expressed in *E. coli* Arctic Express (DE3) cells in a soluble form. The work on PrnA and SttH was carried out along with our group member Sarah Shepherd. Mutagenesis was performed on PrnA with the aim of widening the substrate scope. Site directed mutagenesis was carried out on the residues interacting with tryptophan mainly Y443, Y444, N459S and E450. The screening of the mutants showed similar activity to that of wild type, but a single mutant E450K was

identified to increase the activity of anthranilic acid. Mutagenesis was performed on Y443, Y444, E450 and F454 to change them into a polar residues. Mutant clones were obtained in soluble form and it was found that the F454K, F454R mutants had the highest conversion of 54% while other mutants had conversion of less than 20% with anthranilic acid. Optimisation of protein purification was carried out to obtain the crystal structure of F454K and E450K mutants with anthranilic acid and pure F454K was obtained at 34 mg ml⁻¹. The crystal structure of F454K mutant with FAD and chloride bound structure was obtained, but unfortunately the substrate bound crystal structure determination was unsuccessful. On comparison of the mutant to the wild type structure showed that the FAD and Cl⁻ binding sites were at the same position. The difference lies in residues Y443 and Y444 which form an extended loop. The mutant shows a closer electrostatic interaction between the hydroxyl group of anthranilic acid and amino group of K454. The experiments were also performed on switching amino acid residues between PrnA and PyrH to produce regioselective halogenation. The mutants resulted in no change in the regioselective halogenation and showed a low relative activity indicating the residues to be important for better halogenation. To further change the regioselectivity, an 8 amino acid loop was deleted from PrnA which was mostly absent in the other tryptophan halogenases. The resulting mutant had no halogenation activity indicating that it was very important for function of these enzymes. SttH was characterised for the substrate scope by Sarah Shepherd and she found that the enzyme halogenated tryptophan, kynurenine, anthranilamide, 2-propylaniline and methyl anthranilate. Further mutagenesis was not feasible due to the absence of a crystal structure. An effort was taken to optimise the purification procedure to obtain pure SttH for crystal trials. Pure SttH of 18.53 mg ml⁻¹ was obtained and MALS analysis showed the protein to be a dimer. The repeated crystal trials yielded tiny crystals, but unfortunately it was not suitable for resolving the full crystal structure of SttH. The regioselectivity halogenation by SttH was analysed by sequence alignment with other tryptophan halogenases. The SttH98A and SttH loop deletion (8

amino acid) mutant resulted in no change in regioselectivity with a very low relative activity indicating it be important in halogenation. PyrH was screened for substrates and it showed chlorination and bromination for tryptophan, kynurenine and anthranilamide while it showed only chlorination for 2-amino-4-methylbenzamide. All the above assays were characterised by mass spectrometry and NMR. Other substrates aniline, anthranilic acid, methyl anthranilate and N-phenyl anthranilic acid showed relative activities by HPLC and mass spectrometry. The regioselectivity of PyrH was carried out similar to that of PrnA, wherein mutants showed no change in the regioselectivity with low relative activity. The PyrH loop inserted mutant also had a similar result to that of PrnA, wherein no relative activity was observed indicating the loop to be important in retaining halogenation activity.

As future work, the enzymes PrnA, SttH and PyrH will be used to screen new substrates for producing regioselective halogenated compounds. Efforts are being made in the lab to obtain the kinetic data of these enzymes and mutants with the substrates characterised so far. Mutagenesis is being performed on PyrH to widen the substrate scope based on the interaction of residues with tryptophan. Work is also being carried out to obtain anthranilic acid bound F454K and E450K mutants and the SttH (tryptophan-6-halogenase) crystal structure. The crystal structure obtained will be used to understand the exact regioselective halogenation of the substrate at the 6-position by this enzyme. The structure will also be used to widen the substrate scope and produce regioselective halogenated compounds. Recently palladium based cross coupling reactions have been carried out using bio-orthogonal conditions. A palladium cross coupling could be possible with these enzymes as a single reaction with substrates aiding in carbon-carbon bond formation. This process could be developed to incorporate fluorescent aromatic substituents in halogenated products. This would provide development of a high-throughput technique which could be used in directed evolution. Random-mutagenesis could be performed and

high-throughput screening technique could be used to screen halogenated products in a rapid fashion.

CHAPTER 5. GENERAL MATERIALS AND METHODS

5.1. General information

The procedures for molecular biology experiments were carried out following Sambrook *et al.*, 2001¹¹⁰.

5.1.1. Chemicals, enzymes and kits

All the chemicals were purchased from Sigma Aldrich (UK), Fisher (UK) and Merck (UK). Phusion high-fidelity DNA polymerase, restriction digestion enzymes and DNA ladders were bought from New England Biolabs, UK. Molecular biology kits were purchased from Qiagen, UK and Sigma Aldrich, UK. The protein ladder was bought from Fermentas life Sciences, UK. The complete mini EDTA-free protease inhibitor cocktail tablets were obtained from Roche, UK. The *E. coli* ArcticExpress (DE3) RP, *E. coli* Rosetta 2(DE3) competent cells, site directed mutagenesis kit, FPLC and HPLC columns were bought from Agilent technologies, UK. All the primers were synthesised and obtained desalted from Sigma Aldrich, UK. The genes were synthesised from GeneArt, Life Technologies, Invitrogen divisions, UK.

5.2. Stock solutions for experiments

5.2.1. Antibiotics solution

Table 18. Antibiotic stock solutions

Antibiotic	Stock solution (mg ml ⁻¹)		Final concentration (µg ml ⁻¹)	
	Concentration	storage	Medium	Broth
Ampicillin	50	-20 °C	50	50
Kanamycin	10	-20 °C	10	10
Gentamycin	20	-20 °C	20	20
Chloramphenicol	24	-20 °C	24	24

5.2.2. Substrates stock solution

Table 19. Substrates stock solution

S. NO	Substrates	Molecular mass (g)	For 50 mM stock solution (mg ml⁻¹)	Dissolved
1	Tryptophan	204.2	0.01020	Water/DMSO
2	Kynurenine	208.2	0.01040	Water/DMSO
3	Anthranilamide	136.1	0.00680	Water/DMSO
4	Anthranilic acid	137.1	0.00685	Water/DMSO
5	Methyl anthranilate	151.1	0.00755	Water/DMSO
6	2-propyl aniline	135.1	0.00675	Water/DMSO
7	2-propyl phenol	136.1	0.00680	Water/DMSO
8	Kynurenic acid	189.1	0.00945	Water/DMSO
9	Quinazoline	130.1	0.00650	Water/DMSO
10	Acetanilide	135.1	0.00675	Water/DMSO
11	2-aminobenzonitrile	118.1	0.00590	Water/DMSO
12	Isatoic anhydride	163.1	0.00815	Water/DMSO
13	Benzodiazepine	248.7	0.01243	Water/DMSO
14	Salicylamide	137.1	0.00685	Water/DMSO
15	Thiosalicylic acid	154.1	0.00770	Water/DMSO
16	Ethyl 2-aminobenzoate	165.1	0.00825	Water/DMSO
17	Aniline	93.1	0.00465	Water/DMSO
18	Methyl 3-aminobenzoate	151.1	0.00758	Water/DMSO
19	Methyl 4-aminobenzoate	151.1	0.00758	Water/DMSO
20	Methyl salicylate	152.1	0.00760	Water/DMSO
21	2-amino-5-methylbenzamide	150.18	0.00750885	Water/DMSO
22	2-amino-4-methylbenzamide	192.08	0.009604	Water/DMSO
23	2-amino-N-methylbenzamide	150.18	0.007509	Water/DMSO
24	N-phenylanthranilic acid	213.23	0.016066	EtOH/DMSO
25	1-methyl-2-aminoterephthalate	195.17	0.009755	Water/DMSO

5.3. Preparation of competent cells using calcium chloride

5.3.1. Materials required

100 mM calcium chloride solution

11.09 g of calcium chloride was dissolved in 1L of dH₂O, autoclaved and stored at 4 °C.

20 mM calcium chloride with 20% glycerol solution

0.2208 g of calcium chloride was dissolved in 100 ml of dH₂O, 20 ml of glycerol was added, autoclaved and stored at 4 °C.

LB broth

10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl in 1 litre of distilled water was adjusted pH to 7.0 with NaOH and autoclaved before use.

5.3.2. Procedure

The preparation of competent cells was slightly modified from Sambrook *et al.*, 2003¹¹⁰. Either a single colony of the desired cells or a glycerol stock was used to inoculate 3 ml of LB medium with the required antibiotic (if necessary for selecting the plasmid) and grown overnight at 37 °C. The overnight culture were scaled up to 300 ml of LB broth and grown at 37 °C with shaking until the OD₆₀₀ 0.3-0.4. The culture was kept on ice for about 15-20 min and centrifuged at 5000 *xg* for 10 min. The cell pellet was re-suspended in 50 ml of 100 mM CaCl₂ solution. The process of spinning down and resuspending was the repeated 3-4 times and the final pellet was resuspended in 1-2 ml of 20 mM CaCl₂ with 20% glycerol solution. The cell suspension was dispensed in 200 µl aliquots into 1.6 ml microfuge tubes and immediately frozen in liquid nitrogen and stored at -80 °C.

5.4. Preparation of competent cells using Inoue solution

5.4.1. Materials required

0.5 M PIPES (piperazine-1,2-bis[2-ethanesulfonic acid]) solution

15.1 g PIPES was dissolved in 80.0 ml of dH₂O, KOH was added until the PIPES had dissolved. The pH was adjusted to 6.7 and was made up to 100.0 ml using dH₂O.

Inoue solution

10.9 g MnCl₂, 2.2 g CaCl₂, 18.7 g KCl, 20.0 ml of 0.5 M PIPES solution were dissolved in 1L of dH₂O and filter-sterilize using 0.25 µM syringe filter. .

SOB medium

0.5 g NaCl, 20.0 g tryptone, 5.0 g yeast extract, 0.186 g KCl was dissolved in 1litre of dH₂O, autoclaved and stored at room temperature.

SOC medium

0.5 g NaCl, 20.0 g tryptone, 5.0 g yeast extract and 0.186 g of KCl was dissolved in 1.0 litre of dH₂O. After autoclaving, 20.0 ml of sterilised 1.0 M glucose was added and stored at room temperature.

5.4.2. Procedure

This method was modified from that described in Inoue *et al.* (1990)¹¹¹. A single colony of *E. coli* (DH5α) from an LB agar plate was used to inoculate 3 ml SOB medium and cultured overnight at 37 °C. 3 ml of overnight culture was then used to inoculate 250 ml SOB, and grown at 18 °C with shaking until OD₆₀₀ 0.4. The culture was put on ice for 10 min, and then centrifuged (10 min, 4000 *xg*, 4 °C). The cell pellet was re-suspended in 80 ml ice-cold Inoue solution and pelleted by centrifugation (10 min, 4000 *g*, 4 °C). The cells were re-suspended in 20 ml cold Inoue solution with DMSO (final concentration 7%) and incubated on ice for 10 min. The cell

suspension was then dispensed in 200 µl aliquots into 1.6 ml microfuge tubes and immediately frozen in liquid nitrogen and stored at -80 °C.

5.5. Growth of cells

5.5.1. *Kutzneria spp.744*

Merlin Morkans medium

Ammonium phosphate monobasic 0.184 g, calcium chloride 0.05g, ferric EDTA 0.02g, magnesium sulphate 0.0358 g, potassium phosphate monobasic 0.5 g, sodium chloride 0.025 g, malt extract 3.0 g, sucrose 10.0 g and thiamine 0.0001 g in 1 litre of water, pH was adjusted to 5.5, autoclaved and store at 2-6 °C. For solid medium, 15% agar was added prior to autoclaving.

Kutzneria spp.744 culture was inoculated in MNM modified medium using a thick wire sterile loop. The cultures were grown for 7-14 days at 22 °C at 180 rpm.

5.5.2. Growth of *E. coli* BL21 (DE3) cells

A single colony of *E. coli* BL21 (DE3) cells from the glycerol stock of competent cells was used to inoculate 3 ml of LB medium (5.3.1) with required antibiotic (if necessary for selecting plasmid) and grown overnight at 37 °C.

5.5.3. Growth of *Actinoplanes* ATCC30766

Bouillon broth

Meat Extract 10.0 g, Peptone 10.0 g, NaCl 5.0 g, K₂HPO₄ 2.0 g was dissolved in 1.0 litre of dH₂O, adjusted to pH 7.0 and autoclaved.

The *Actinoplanes* ATCC30766 culture was inoculated in Boullin broth using a thick wire sterile loop. The cultures were grown for 7-14 days at 22 °C at 180 rpm.

5.6. Isolation of genomic DNA

Tris-EDTA (TE) buffer

10 mM tris, 1 mM EDTA was dissolved in 100 ml of dH₂O, pH adjusted to 7.0. The 1 ml mycelium of *kutzneria* spp.744 and *Actinoplanes* ATCC30766 cells were taken and centrifuged at 5000 *xg* for 10 min to obtain a pellet. Similarly 5 ml of *E. coli* BL21 (DE3) cells were centrifuged at 5000 *xg* for 10 min to obtain a pellet. The supernatant was completely discarded and pellet was further used to extract the gDNA using the Qiagen genomic tip DNA extraction kit by following the method described in the kit. The isolated DNA was dissolved in 200 µl of TE buffer and stored at 4 °C.

5.7. Plasmid DNA extraction

A single colony of *E. coli* TOP10 cells or *E. coli* DH5a cells with plasmid was inoculated in 3 ml of LB broth containing the desired antibiotics with concentrations mentioned in section 5.2.1. The culture was grown overnight at 37 °C at 200 rpm. The 3 ml cells were centrifuged at 5000 *xg* for 5 min to obtain the pellet. The supernatant was discarded and pellet was used for plasmid DNA extraction using Qiagen spin miniprep kit by following the method described in the kit.

5.8. Agarose gel Electrophoresis

DNA Electrophoresis Buffer

10x TBE buffer containing 890 mM Tris-Cl, 890 mM Boric acid and 20 mM EDTA, pH 8.2.

DNA loading dye

25 mg of bromophenol blue, 3 ml of glycerol was dissolved in 10 ml of d.H₂O.

To prepare 1.2% of agarose gel, 1.2 g of agarose was added to 100 ml of 1X TBE and boiled in a microwave until agarose was completely dissolved. The boiled solution was allowed to cool and ethidium bromide (2 µl of 10 mg ml⁻¹) was added and poured onto gel casting tray with appropriate comb. The gel was allowed to

polymerize and the comb was removed to form wells. The gel casting tray was placed in a gel running tank immersed with 1X TBE. Samples were loaded with DNA loading dye and were run at 110 V for 45 min. The DNA was observed by using a UV transilluminator.

5.9. PCR amplification of genomic DNA

The sequence data of the genes were obtained from the NCBI database and suitable primers were designed clone manager software for potential amplification. The primers were designed to incorporate the desired restriction enzyme sites with overhangs and were synthesised by Sigma Aldrich, UK. The primers used to amplify the products are listed in Table 32 .

5.9.1. PCR amplification and conditions

All the genes were amplified using Phusion high fidelity DNA polymerase and the following standard PCR mixtures were prepared.

Table 20. Cocktail mixture of PCR reaction

Component	Amount (ml)
5x Phusion™ HF Buffer	10
10 mM dNTPs	1
10 mM Forward primer	1
10 mM Reverse primer	1
Template DNA	50-100ng
DMSO (if required)	1.5
Phusion DNA Polymerase	0.5
Total volume	50

Based on the Primer design the PCR amplification was carried out as followed

Table 21. PCR cycle used in thermocycler

Cycle step	Temperature (°C)	Time
Initial denaturation	98	2 min
Denaturation	98	20 sec
Annealing	42-60	20 sec
Extension	72	30 sec/Kb
Final extension	72	5 min

5.10. Gel-extraction of DNA from Agarose Gels

The desired DNA was excised from agarose gel with a sterile scalpel blade and the DNA was extracted using Qiagen Gel Extraction Kit following the standard method described.

5.11. Digestion, Ligation and Transformation

The amplified genes were digested with respective enzymes overnight at 37 °C. Similarly the vectors used to clone were digested with respective enzymes at 37 °C for 3 hrs. The digested samples were run on a 1.0% agarose gel and were extracted using a Qiagen gel extraction kit. The following standard reaction was used for restriction digestion

Table 22. Restriction digestion reactions mixture

Component	Amount (µl)
Buffer	5
Template DNA	40
Enzyme 1	2.5
Enzyme 2	2.5
BSA(if needed)	0.5
Total	50

5.12. Ligation Reaction

A ligation reaction was performed using T4 DNA ligase at the insert to vector ratio of 3:1. The ligation reaction was carried out at 16 °C overnight and the following standard reaction was used.

Table 23. Ligation reactions mixture

Components	Amount(μ l)
DNA vector 50 ng	X
Insert (calculated ng)	Y
10 X Buffer	2
T4 DNA Ligase NEB	1
H ₂ O	to 20
Total	20

The ligated product was transformed either into *E. coli* TOP10 or *E. coli* DH5 α cells.

5.13. Transformation

The SOC medium was used in all transformation experiments. 10 μ l of the ligated mixture was added to the 50 μ l of *E. coli* TOP10 or *E. coli* DH5 α cells competent cells and placed on ice for 30 min. For transforming the constructs, 2 μ l of the plasmid solutions were used. The cells were heat shocked at 42 °C for 60 sec and placed on ice for 5 min. After 5 min 300 μ l of SOC broth was added to the cells, incubated at 37 °C for 1 hr and plated on LB agar plate containing the desired antibiotic. The plate was then incubated overnight at 37 °C to obtain colonies.

5.13.1. Analysis of the transformed colonies

The transformed colonies were analysed for the gene insert using colony PCR and restriction digestion analysis. For colony PCR, *Taq* DNA Polymerase, NEB, UK was used and the standard reaction was prepared as follows

Table 24. *PCR reactions mixture of performing colony PCR*

Component	Amount (μ l)
10X <i>Taq</i> buffer	10
10 mM dNTPs	1
10 mM Forward primer	1
10 mM Reverse primer	1
DMSO	1.5
<i>Taq</i> DNA polymerase	0.5
Total volume	To 50

A Single colony from the plate was picked up using a sterile pipette tip and added to the reaction mix. The tip was twirled well to break the cells. The PCR cycle was performed as mentioned below

Table 25. *Thermocycler conditions used for colony PCR*

Cycle step	Temperature (°C)	Time	Cycles
Initial denaturation	95	5 min	1
Denaturation	95	20 sec	
Annealing	40	20-30 sec	30
Extension	68	1 min/Kb	
Final extension	68	5 min	

The restriction digestion was carried out as mentioned in section 5.11. The colony with insert was further confirmed by sequencing.

5.13.2. Sequencing

The sequencing was carried out in GATC, Germany using 30-100ng of plasmid DNA material.

5.14. Site directed mutagenesis

To introduce a single base point mutation in the DNA sequences, the Quick change Site-Directed Mutagenesis kit was used. The mutation used a single primer for point mutation. The constructs were used as templates for

mutagenesis and a list of primers used for creating various mutants is mentioned in Table 32. The reaction mixture was carried out as mentioned below

Table 26. *PCR reactions mixture for performing site directed mutagenesis*

Component	Templates >5 Kb (μ l)
10X Quick change multi reaction buffer	12.5
Double distilled water	To final volume of 12.5
DMSO	1.5
Ds DNA(Construct)	2.0 (50 ng)
100 mM Mutagenic primer	0.5
dNTP mix	0.5
Quick change multienzyme	0.5

The PCR reaction is carried out as follows

Table 27. *Thermocycler conditions used for site directed mutagenesis*

Cycle step	Temperature (°C)	time	cycles
Initial denaturation	95	1 min	1
Denaturation	95	1 min	30
Annealing	55	1 min	
Extension	65	2 min/Kb	
Final extension			

The PCR amplification reaction was allowed to cool, 1 μ l of *Dpn1* restriction enzyme was added to the reaction and incubated at 37 °C for 1 hr to digest parental ds DNA. The mutation reaction was transformed (Section 5.13) and the colonies were analysed for mutation by sequencing.

5.14.1. Site directed mutagenesis to insert and delete an 8 amino acid loop region

A different methodology was used to insert the loop and delete the loop within these enzymes.

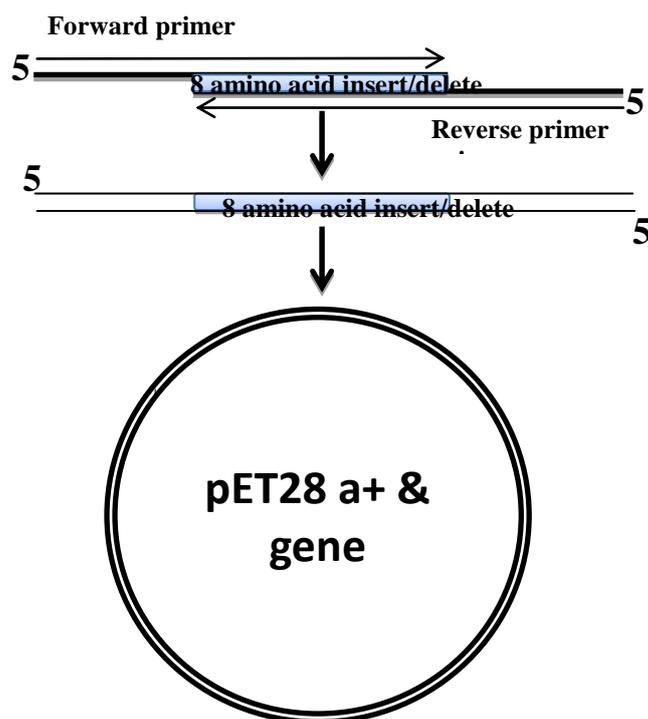


Figure 103. Site directed mutagenesis strategy to insert/delete an 8 amino acid loops

Forward and reverse primers were designed with 55-60 nucleotides having flanking region of the genes of 20 nucleotides at each end followed by the sequence with its complement in the middle (Figure 103). The PCR reaction was carried out for 16 cycles at 40 °C with extension of 30" and a mega primer was obtained. This mega primer was later used for site directed mutagenesis with the desired construct as template for mutation reactions.

5.15. Protein Expression and purification

1 M Isopropyl β-D-1-thiogalactopyranoside (IPTG) solution

0.283 mg of IPTG was dissolved in 10 ml of dH₂O to obtain 1 M stock solution and was stored at -20 °C

HiTrap™ Chelating

5.15.1. Protein Expression Studies

Expression was tested in a 50 ml culture volume. A 1:50 dilution from an overnight culture was grown to log phase (OD₆₀₀ 0.6). The cells were then

induced with IPTG (1mM final concentration) and grown at various temperatures 37 °C, 30 °C, 22 °C and 18 °C for 3 hr and overnight.

5.15.2. Protein expression using *E. coli* BL21 (DE3) cells

3 ml LB broth was inoculated with single *E. coli* BL21 (DE3) colony with the constructed plasmid, antibiotic and was grown overnight in 37 °C at 200 rpm. The overnight culture was diluted 1:100 in fresh LB with antibiotic and grown at 37 °C shaking until the OD₆₀₀ 0.6 and IPTG (final concentration 0.5-1.0 mM) was added to the culture. The culture was incubated under optimised condition.

5.15.3. Protein expression using *E. coli* RosettaGami (DE3) RP cells

RosettaGami (DE3) cell consists of T7 RNA with *lac* promoter and can be used with all pET vectors. This engineered strain provides t7RNAs for rare codons AGG, AGA, AUA, CUA, CCC, GGA and enhances disulfide bond formation while expressing proteins in *E. coli*. These cells are compatible with ampicillin resistant vectors. The protein expression is carried out similar to section 5.15.2.

5.15.4. Protein expression using *E. coli* Rosetta2 cells

E. coli Rosetta 2 host strains are BL21 derivatives designed to enhance the protein expression that contain rare codons in *E. coli*. These strains supply tRNAs for rare codons AGA, AGG, AUA, CUA, GGA, CCC, and CGG on a compatible chloramphenicol-resistant plasmid. The protein expression is carried out similar to section 5.15.2.

5.15.5. Protein expression using *E. coli* ArcticExpress (DE3) RP cells

The high-level expression of proteins in *E. coli* results in incorrectly folded protein forming as aggregates/inclusion bodies. These proteins are usually inactive and refolding of this mis-folded protein is time consuming. It remains difficult to obtain the enzymes in active form. Another way of obtaining soluble protein, involves growing the cells at lower temperatures.

The *E. coli* cells have a high ability of protein expression at higher temperatures, while at lower temperatures, the cells growth are hampered and production of the soluble proteins is very minimum. The various chaperones involved inside the cells provide proper folding of these proteins in its native state. These chaperones have a reduced activity at lower temperature. Arctic express are the engineered strain that could adapt at lower temperatures of 4-12 °C. The strains are engineered to produce chaperones Cpn10 and Cpn60 (74% and 54% identity to the *E. coli* GroEL and GroES chaperones) which can function at cold temperatures and have high protein refolding activities. The Arctic Express strains are designed for expression of recombinant proteins using the T7 promoter using the DE3 lysogen strain and contain tRNAs for the rare arginine codons AGA, AGG and proline CCC, respectively.

E. coli ArcticExpress (DE3) RP cells were transformed with the construct for protein expression. An overnight culture was set up in LB using single colony along with 20 µg ml⁻¹ gentamycin and, antibiotic for selection of the respective plasmid construct and grown overnight at 37 °C. The overnight culture was diluted 1:100 in fresh LB without antibiotics and grown at 30 °C with shaking until the OD₆₀₀ 0.6 and IPTG (final concentration 0.1 mM) was added to the culture and grown at 15 °C or 12 °C overnight.

5.15.6. Isolation, purification and concentration of protein

1 M Di-Potassium hydrogen orthophosphate (K₂HPO₄)

172.4 g was dissolved in 1 litre of water to obtain 1M stock solution of K₂HPO₄.

1 M Potassium dihydrogen orthophosphate (KH₂PO₄)

136.08 g was dissolved in 1 litre of water to obtain 1 M stock solution of KH₂PO₄.

100 mM potassium phosphate buffer

100 mM phosphate buffer solution was prepared according to procedure mentioned in Sambrook *et al.*, 2003 by following the table below. The

solution was added as mentioned in the table 34 and made to 1 litre using dH₂O to obtain 100 mM phosphate buffer with desired pH.

Table 28. *Phosphate salts required for obtaining 100mM phosphate buffer of desired pH*

pH	Volume of 1 M	Volume of 1 M
	K ₂ HPO ₄	KH ₂ PO ₄
7.2	71.7	28.3
7.4	80.2	19.8

Cells were then collected by centrifugation at 5000 *xg* for 10 min and the pellets were re-suspended in 100 mM phosphate buffer. The cells lysed using sonicator. Before sonication the resuspended cells were added with 1 mg ml⁻¹ lysozyme, EDTA free protease inhibitor 1 tablet/10 ml and incubated at 30 °C for 1hr. Sonication was performed in Bandelin Sonoplus sonicator 2000 series which has a microtip MS 74 with 3 mm diameter and maximum output of 70 W_{eff}. Sonication was carried out in 70% amplitude at 30 second intervals to break the cell wall. The sonicated cells were centrifuged at 10,000 *xg* for 30 min at 4 °C. The supernatant was used for purification using an AKTA purifier 10 UPC system (GE Healthcare).

5.15.6.1. Protein purification by HiTrap™ Chelating HP column using AKTA FPLC system

1 M Nickel sulphate

1.877g of NiSO₄ was dissolved in 10 ml of dH₂O.

Buffers

20 mM Imidazole solution per litre pH 7.4(lysis buffer)

2.72 g KH₂PO₄, 29.22 g NaCl and 1.36 g of imidazole were dissolved in 1.0 litre of dH₂O; pH was adjusted to 7.4 and stored at 4 °C.

60 mM Imidazole solution per litre pH 7.4

2.72 g KH₂PO₄, 29.22 g NaCl and 4.08 g of imidazole were dissolved in 1.0 litre of dH₂O; pH was adjusted to 7.4 and stored at 4 °C.

90 mM Imidazole solution per litre pH 7.4

2.72 g KH₂PO₄, 29.22 g NaCl and 6.12 g of imidazole were dissolved in 1.0 litre of dH₂O; pH was adjusted to 7.4 and stored at 4 °C.

150 mM Imidazole solution per litre pH 7.4

2.72 g KH₂PO₄, 29.22 g NaCl and 10.2 g of imidazole were dissolved in 1.0 litre of dH₂O; pH was adjusted to 7.4 and stored at 4 °C.

200 mM Imidazole solution per litre pH 7.4

2.72 g KH₂PO₄, 29.22 g NaCl and 13.6 g of imidazole were dissolved in 1.0 litre of dH₂O; pH was adjusted to 7.4 and stored at 4 °C.

250 mM Imidazole solution per litre pH 7.4

2.72 g KH₂PO₄, 29.22 g NaCl and 17.02 g of imidazole were dissolved in 1.0 litre of dH₂O; pH was adjusted to 7.4 and stored at 4 °C.

500 mM Imidazole solution per litre pH 7.4

2.72 g KH₂PO₄, 29.22 g NaCl and 34.04 g of imidazole were dissolved in 1.0 litre of dH₂O; pH was adjusted to 7.4 and stored at 4 °C.

Wash buffer

2.72 g KH₂PO₄, 29.22 g NaCl and 14.62 g of EDTA were dissolved in 1.0 litre of dH₂O and stored at 4 °C.

Protein purification was performed by using Ni-NTA FPLC system (AKTA prime, Amersham) in conjunction with a HiTrap™ Chelating HP 5 ml column (GE Healthcare, UK). Before loading the cell lysate, a 0.1M Ni²⁺ ion solution was applied into the HiTrap™ Chelating HP 5 ml column and was equilibrated in lysis buffer until the absorbance read approximately zero. The protein was loaded, washed with lysis buffer for 6 column volumes followed by increased concentration. The recombinant protein was eluted with optimised imidazole concentration of 6 column volumes. The

column was washed and nickel ion was removed using wash buffer, followed by wash with ethanol and storing the column in 4°C.

5.15.6.2. Protein purification by Gravity flow method

The protein lysate was allowed to bind manually to the nickel column. A bed volume of 2 ml nickel was prepared and subjected to manual column. The column was washed with 3 volumes of water followed by binding with the binding buffer (phosphate buffer with 20 mM imidazole). The washed nickel was added to the protein in the centrifuge tube and kept in cold room for overnight with slow shaking. The overnight sample was directly loaded onto the manual column and allowed to further bind the nickel by gravity flow. The protein was washed with 60 mM imidazole to remove the nonspecifically bound proteins following which SttH was eluted in different fractions using 500 mM imidazole.

5.15.6.3. Removal of chaperones by ATP Buffer method

10 mM ATP, 50 mM KCl and 10 mM MgCl₂ in water, pH 7.2 was prepared. The protein was expressed and isolated (Section 5.15). The lysate was loaded onto a HiTrap TM Chelating column and washed with 500 mM phosphate buffer containing 60 mM imidazole, and 90 mM imidazole. Further wash and elutions were made using the ATP buffer before eluting the protein at a concentration of 500 mM imidazole using an AKTA purifier 10 UPC system (GE Healthcare).

5.15.6.4. Chromatographic techniques

5.15.6.4.1. Anion exchange chromatography

For carrying out anion exchange chromatography two different columns were used

HiTrap Q HP

HiTrap Q HP column are packed with 34 µm sized Q Sepharose bead that acts as strong anion exchange medium.

MonoQ 4.6/100 PE

MonoQ column is packed with 10 µm sized monobeads in Tricorn column that acts as strong anion exchange medium.

20 mM Tris with 20 mM, 60 mM, 90 mM, 150 mM, 250 mM and 500 mM NaCl buffer, pH7.4

2.42 g of Tris base and the required amount of NaCl (Table 29) was dissolved in 1litre of dH₂O, adjusted to pH 7.4 and stored at 4 °C.

Table 29. Amount of imidazole required for preparation of desired concentration of buffers

Concentration of NaCl	Amount (g)
20 mM	1.16
60 mM	3.50
90 mM	5.25
150 mM	8.76
250 mM	14.61
500 mM	29.22

The anion exchange chromatography involves the separation of negatively charged proteins. The anion exchange column was equilibrated with a lower salt concentration of 20 mM NaCl on AKTA 1.0 purifier. The protein devoid of salt concentration was allowed to bind the column followed by washes with increasing NaCl concentration. The protein elution was made using highest salt concentration of 500 mM NaCl. The concentration of the salt varies based on the enzymes needed to be purified.

5.15.6.4.2. Size exclusion chromatography

Superdex 200 10/300 GL

Superdex 200 10/300 GL is a prepacked Tricorn gel filtration column having separation range for molecules with molecular sizes between 10000 and 600000 daltons.

100 mM HEPES buffer

23.83 g of HEPES was dissolved in 1 litre of dH₂O; pH was adjusted to 7.4 and stored at 4°C.

20 mM Tris buffer

2.43 g of Tris base was dissolved in 1 litre of dH₂O; pH was adjusted to 7.4 and stored at 4°C.

Gel filtration chromatography separates proteins based on their sizes. The column is usually packed with Superdex or Sephadex beads which acts as a stationary phase of the column. Based on the sizes of the stationary phase, different sizes of proteins can be separated. The smaller molecule moves in the beads slowly due to their diffusion in the pore space while larger molecules does not diffuses, which allows them to move quickly through the pore space. Hence bigger sized proteins are eluted first followed by smaller sized proteins. Depending upon the proteins solvent environment, optimised buffers were used for eluting desired proteins.

5.15.6.5. Purification of protein using HRV 3C protease

10X HRV 3C protease cleavage buffer

HRV 3C protease

The concentration of the protein was obtained by Bradford assay (5.17). The components were prepared in a 1.5 ml microfuge tubes

Table 30. Reaction mixture for HRV 3C protease cleavage

Component	Amount
10X HRV 3C protease cleavage buffer	4 µl
Target protein	10, 50, 100 µg
HRV 3C protease	1 µl
dH ₂ O	Make it up

A HRV 3C Protease: target protein ratio (unit/µg) of 1:100 was set up and incubated at 4°C for 24 and 48 hr with shaking. The cleaved samples were purified by FPLC and separated on 12% (v/v) resolving gel.

5.15.6.6. Purification using hydrophobic column

The protein was expressed and isolated as per section 5.15. The lysate was loaded onto a HisTrap™ FF crude column and washed with 50 mM phosphate buffer containing 60 mM imidazole, before eluting the protein at a concentration of 500 mM imidazole using an AKTA purifier 10 UPC system (GE Healthcare). The eluted fraction was dialysed in 50 mM Tris-buffer (pH 8.0) by centrifugation (4 °C, 8,000 xg) using a Vivaspin 20 Centricon with a 10,000 MWCO (Sartorius, Germany), before loading onto a HiTrap™ Q HP anion exchange column. The column was washed with 50 and 100 mM NaCl before eluting halogenase proteins at a concentration of 1 M NaCl. The buffer in the protein containing sample was exchanged by centrifugation as described previously into phosphate buffer containing 1.5 M ammonium sulphate, before binding the protein to a HiTrap™ FF Butyl-sepharose column. The protein of interest eluted whilst running a gradient of 2 M to 0.5 M ammonium sulphate. Exchange of sample into buffer containing 20 mM HEPES, 100 mM NaCl (pH 7.2) preceded a final step of gel filtration chromatography (Superdex™ 200 10/300 column).

5.15.7. Buffer exchange and concentration

Storage buffer

10 mM phosphate buffer was diluted to 10 times and 10% glycerol was added to the solution.

Eluted fractions were pooled and preceded to buffer exchange by using Vivaspin centrifugal concentrator (Sartorius, UK) for proteins that forms precipitate on dialysis. The stable protein was subjected to dialysis in 14,000 daltons cut off dialysis membrane against 100 mM phosphate buffer. The protein was further concentrated using Vivaspin centrifugal concentrator (Sartorius, UK) and exchanged with storage buffer. Then protein was stored at -20 °C and analysed by resolving gel (SDS PAGE).

5.16. Sodium-Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)¹¹²

Acrylamide-bis solution (29:1%)

1.5 M Tris HCl pH 8.3 with 0.4% SDS

0.5 M Tris HCl pH 6.8 with 0.4% SDS

10% Ammonium persulphate

TEMED (N-tetramethylethylene diamine)

1 X Tank buffer

10.82 g glycine, 2.27g Tris base and 0.75g SDS was dissolved in 750 ml of dH₂O.

Staining solution

0.25 g Coomassie brilliant blue, 40 ml methanol and 10 ml acetic acid were dissolved in 50 ml of dH₂O.

De-staining Solution

40 ml methanol and 10 ml of acetic acid were dissolved in 50 ml of dH₂O.

4X Denaturing sample buffer

0.312 M Tris-Cl pH 6.8, 0.05% bromophenol blue, 50% glycerol, 10% SDS, 25% β-mercaptoethanol.

Separating and stacking gels were prepared by following the table

Table 31. SDS-PAGE recipe

Stock solution	Separating gel (12%)	Separating gel (20%)	Stacking gel (5%)
1.5 M Tris HCl pH 8.3+ 0.4% SDS	2.8 ml	2.5 ml	-
0.5M Tris-HCl pH= 6.8+0.4%SDS	-	-	2.5 ml
30% Acrylamide Bis acrylamide)	4.52 ml	7 ml	1.0 ml
H ₂ O	3.72 ml	0.5 ml	6.4 ml
10% (w/v) APS	40 µl	50 µl	100 µl
TEMED	9.2 µl	5 µl	10 µl

SDS-PAGE was used to separate proteins. Protein samples were mixed with denaturing sample buffer and were heated at 99 °C for 5 min. The denatured protein sample was then loaded onto the gel and run at 200 V until the dye front reached the bottom of the gel. The gel was then removed from the plate and the stacking gel was removed from the separating and the gel was stained for 2-4 h with the staining solution. The gel was then sequentially transferred to destaining solution overnight till the blue colour in the background disappeared. The Fermentas page ruler prestained marker protein ladder was used as a protein marker.

5.17. Determination of protein concentration by Bradford Assay

The assay dye reagent was prepared by diluting 1 part of BioRad Dye reagent concentrate with 4 parts of double distilled water. A dilution series of the protein standard BSA was made to produce a standard curve ranging from 0.2 to 0.9 mg ml⁻¹ for comparison of the samples of unknown concentration. Dye reagent was added to each protein standard and the sample solution and vortexed. Samples were left to incubate at room temperature for a minimum of 5 min and the absorbance was measured for each sample $\lambda = 600$ nm using a Varian UV-Vis Spectrophotometer.

5.18. Mass spectrometry analysis by trypsin cleavage

The mass spectrometry of the protein was carried out as service by Martin Read in Manchester interdisciplinary Biocenter, University of Manchester, UK. Briefly, the purified protein was cut out from the gel in about 1mm and the stain was removed using acetonitrile solution. The acetonitrile was removed and trypsin was added to digest the protein. The digestion was carried out at 37 °C overnight. The digested supernatant was pooled and dried in a speed vacuum. The dried sample was dissolved in 15 μ l of water, dried again and resuspended in 10% acetonitrile and 0.1% formic acid. From this 1 μ l of sample was mixed with 1 μ l of Matrix and loaded onto the machine for Mass spectrometry analysis. The matrix usually used is CCA (Cyanocinnamic acid).

The machine used was a Bruker Esquire 3000+ ions trap MS which is connected to a HPLC from LC Packings ultimate system. The HPLC takes maximum of 20 µl of sample and consists of 18 pepmapper 100 column. The flow rate was set at 250 µl/ min and the splitter is usually 1000/1. The buffer A consist 0.1% formic acid and buffer B consists 90% acetonitrile with 0.1% formic acid. The Mass spectrometry is used in positive ion mode with Pico tip fused silica needle emitter with 360 µm diameter and 10 µl of tip diameter. The mono spray flow rate is 200 nl/ min and uses either MS2 or MS3.

5.19. Western blot analysis

Transfer buffer (10X) stock

Transfer buffer (working solution, mix just before use)

Tris base	- 7.58 g	Transfer buffer (10X)	- 10 ml
Glycine	- 36.0 g	Methanol	- 20 ml
Dis. Water	- 250 ml	Sterile water	- 70 ml

Phosphate Buffered Saline (PBS) 10X stock

KH ₂ PO ₄	- 2.0 g
Na ₂ HPO ₄	- 11.5 g
KCl	- 2.0 g
NaCl	- 80.0 g

Adjust the pH to 7.0 with conc. HCl and the make up the final volume to 1000 ml with distilled water.

Blocking solution (for 1 litre)

Milk powder	- 50 g (5%)
10X PBS	- 100 ml
Tween 20	- 10 ml (1%)

Antiserum buffer (for 1 litre)

Milk powder	- 50 g (5%)
10X PBS	- 100 ml
Tween 20	- 1.0 ml (0.1%)

The proteins KtzQ and PrnA were separated with a 12% (v/v) resolving gel. The proteins were transferred onto a nitrocellulose membrane on Trans Blot (Bio-Rad) semi dry transfer cell electroblotting apparatus using Whatman No.3 filter paper strips (wetted with transfer buffer). The instrument was run at 15 Volts for 1 hr using Powerpac[®] 300 (Bio-Rad laboratories). After blotting the membrane was stored in 1X PBS at 4 °C. The membrane was immersed in blocking solution and incubated in shaker at 25-30 rpm for 1 hr at room temp. The blocking solution was discarded and 10-15 ml of antiserum buffer with anti his mouse monoclonal antibody (1: dilution) and incubated at 25-30 rpm rpm for 1 hr at room temp. After 1 hr, the solution was discarded and the membrane rinsed in antiserum buffer for three times at 5 min interval to remove the traces of primary antibody. Then 10-15 ml of antiserum buffer with 2-3 µl of secondary antibody (Goat anti mouse chemiluminescent tagged antibody) and incubated at 25-30 rpm for 1 h at room temperature. After 1 hr, the membrane was incubated in the dark for the colour development. After colour development, air dry the membrane and store at 4 °C.

5.20. Substrate loading assay

An assay was set up using 200 µM PCP, 2 µM Sfp and 1 mM of substrates (CoA-Hpg and SNAC-Hpg). The loading reaction was incubated at 28 °C for 15 min to 1 h in 10 mM phosphate buffer pH 7.5.

5.21. FMN reductase assay

The reductase assay reaction was set up using 500 µM NADH and 3 µM of FMN in 10 mM phosphate buffer (pH 7.4) as control. The absorbance was monitored at 340 nm for 20 min. Later 50 µl of reductase was added and absorbance was monitored at 340 nm for 20 min.

5.22. Halogenase assays for chlorination and bromination

The assay reaction was set up using 5 µM halogenase, 1.5 µM Fre, 0.6 mM substrate, 1 µM FAD, 12.5 mM MgCl₂ or 12.5 mM NaBr in 10 mM

phosphate buffer at 200 μ l scale. The assay mixture was incubated at 30°C for 1 hr with shaking at 800 rpm and reaction was stopped by heating them at 95°C for 10 min. The reaction was centrifuged at 14,000 xg and the supernatant solution was loaded on to HPLC.

5.23. Enzyme stability assays in solvent environment

5% DMSO/EtOH was made up with 95% of dH₂O to obtain 5% (v/v) DMSO/EtOH solution

10% DMSO/EtOH was made up with 90% of dH₂O to obtain 10% (v/v) DMSO/EtOH solution

10 mM potassium phosphate buffer

5 μ M of purified PyrH was dispensed in 5% (v/v) and 10% (v/v) DMSO/EtOH solution and incubated for 5 hr. As control, the enzyme was dispensed in 10 mM phosphate buffer and incubated for 5 hr. Post incubation, the reaction was added with 1.5 μ M Fre, 0.6 mM substrate, 1 μ M FAD, 12.5 mM MgCl₂ and further incubated at 30°C for 4 hr. Samples were collected hourly and reactions were stopped by heating them at 95°C for 10 min. The reaction was centrifuged at 14,000 xg and the supernatant solution was loaded on to HPLC.

5.24. High pressure liquid chromatography and method

All the assays were run on the Agilent Technologies 1260 Infinity Binary LC system with fraction collector using Zorbax Eclipse Plus C18 4.6x100 mm, 3.5 μ m column. The samples for NMR purification were run on Varian ProStar 320 UV/Vis detector using Varian Pursuit C18 10 μ semiprep column with following method (Figure 104) at various UV absorbance based on compounds.

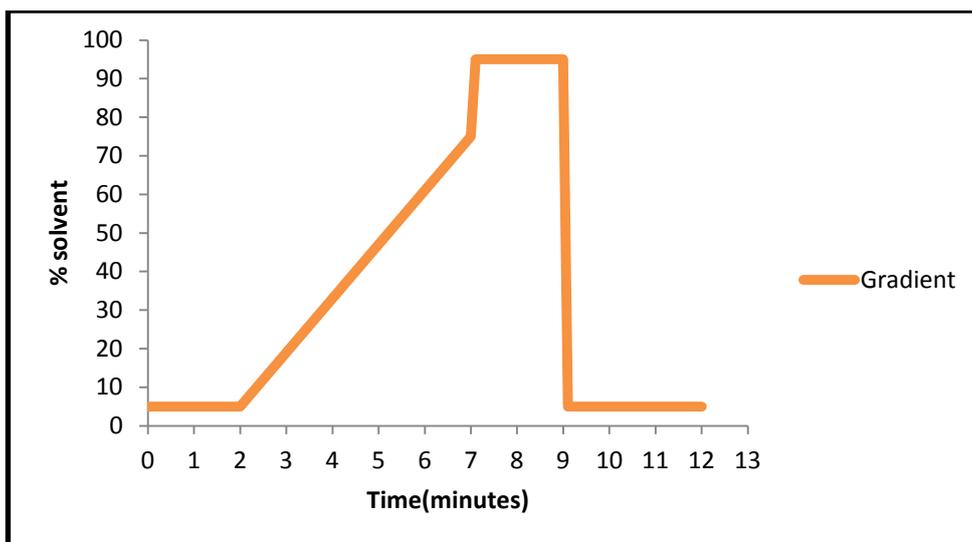


Figure 104. HPLC method used to separate halogenated compounds

5.25. Mass spectrometry of the HPLC samples

All the Mass spectrometry assays were run on the Agilent Technologies 1100 series quaternary pump HPLC with Agilent LC/MSD Trap XCP Mass Spectrometry Detector using C18 column. The samples were run using method mentioned at section 5.24 with detection at various UV absorbance based on compounds (Figure 104).

5.26. Preparation of samples for NMR

For scale up for NMR assay reactions were set up using 20 μM halogenase, 5 μM Fre, 0.6 mM substrate, 1 μM FAD, 12.5 mM MgCl_2 or 12.5 mM NaBr in 10 mM phosphate buffer at 10 ml scale. The assay mixture was incubated at 30 $^\circ\text{C}$ for 5 hr with shaking at 800rpm and reaction was stopped by heating them at 95 $^\circ\text{C}$ for 10 min. The reactions were centrifuged at 14,000 $\times g$ and the supernatant solution was loaded on to C18 bond elute column. The C18 bond elute column was activated with 5 ml methanol and washed with 5 ml of water for equilibration. The supernatant was loaded by gravity flow. After the sample was bound to the column, it was washed with 10% methanol, and 15% methanol to remove FAD and NADH from the bond elute. The pure product is eluted with 95% methanol. Methanol was

removed from the product using a Genevac EZ-2 personal evaporator. The dry product was resuspended in 0.75 ml of deuterated water (D₂O) and subjected to NMR.

5.27. NMR results

5-chloro-tryptophan

¹H NMR (400 MHz, D₂O) δ 7.58 (d, *J* = 1.9 Hz, 1H), 7.32 (d, *J* = 8.7 Hz, 1H), 7.18 (s, 1H), 7.08 (dd, *J* = 8.7, 1.9 Hz, 1H), 3.88 (dd, *J* = 7.9, 4.9 Hz, 1H), 3.33 – 3.04 (m, 3H).

5-chloro-kynurenine

¹H NMR (400 MHz, D₂O) δ 7.69 (d, *J* = 2.4 Hz, 1H), 7.22 (dd, *J* = 8.9, 2.4 Hz, 1H), 6.70 (d, *J* = 9.0 Hz, 1H), 3.99 (t, *J* = 5.5 Hz, 1H), 3.52 (d, *J* = 5.4 Hz, 2H).

5-chloro-anthranilamide

¹H NMR (400 MHz, D₂O) δ 7.55 (d, *J* = 2.2 Hz, 1H), 7.33 (dd, *J* = 8.8, 2.3 Hz, 1H), 6.87 (d, *J* = 8.7 Hz, 1H).

5-chloro-2-amino-4-methylbenzamide

¹H NMR (400 MHz, D₂O) δ 7.43 (s, 1H), 6.70 (s, 1H), 1.9 (s, 3H).

5-bromo-tryptophan

¹H NMR (400 MHz, D₂O) δ 7.91 (d, *J* = 1.6 Hz, 1H), 7.31 (d, *J* = 8.5 Hz, 1H), 7.15 (s, 1H), 7.02 (dd, *J* = 8.6, 1.9 Hz, 1H), 3.96 (dd, *J* = 7.8, 4.7 Hz, 1H), 3.59 – 3.12 (m, 3H).

5-bromo-Kynurenine

¹H NMR (400 MHz, D₂O) δ 7.99 (d, *J* = 2.1 Hz, 1H), 7.50 (dd, *J* = 8.9, 2.0 Hz, 1H), 6.82 (d, *J* = 9.0 Hz, 1H), 4.14 (t, *J* = 5.4 Hz, 1H), 3.67 (d, *J* = 5.4 Hz, 2H).

5-bromo-Anthranilamide

^1H NMR (400 MHz, D_2O) δ 7.69 (s, 1H), 7.46 (d, $J = 10.9$ Hz, 1H), 6.82 (d, $J = 8.8$ Hz, 1H).

5.28. Crystal soaking and crystal trial conditions

The concentrated protein was incubated with 5 mM FAD, 5 mM DTT and 20 mM substrate prior to crystallography trials. The crystals were obtained using 10% (v/v) PEG 8000, 0.1 M imidazole buffer pH 8.0 and 0.2 M Ca $(\text{CH}_3\text{CO}_2)_2$.

Chapter 6. References

1. The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrodinger, LLC.
2. Kandepi, V. and Narender, N. Ecofriendly oxidative nuclear halogenation of aromatic compounds using potassium and ammonium halides. *Synthesis (Stuttg)*. **44**, 15–26 (2011).
3. Gribble, G. Naturally occurring organohalogen compounds--A survey. *J. Nat. Prod.* **55**, 1353–1395 (1992).
4. Zopf, W. Zur Kenntnis der Flechtenstoffe. *Liebigs Ann. der Chemie* **336**, 46–85 (1904).
5. Harris, C. & Kannan, R. The role of the chlorine substituents in the antibiotic vancomycin: preparation and characterization of mono- and didechlorovancomycin. *J. Am. chem.Soc.* **20**, 6652–6658 (1985).
6. Groll, M., Huber, R. and Potts, B. C. M. Crystal structures of Salinosporamide A (NPI-0052) and B (NPI-0047) in complex with the 20S proteasome reveal important consequences of beta-lactone ring opening and a mechanism for irreversible binding. *J. Am. Chem. Soc.* **128**, 5136–5141 (2006).
7. Pereira, E. R., Belin, L, Sancelme, M, Prudhomme, M, Ollier, M, Rapp, M, Severe, D, Riou, J.F, Fabbro, D. & Meyer, T. Structure-activity relationships in a series of substituted indolocarbazoles: topoisomerase I and protein kinase C inhibition and antitumoral and antimicrobial properties. *J. Med. Chem.* **39**, 4471–4477 (1996).
8. Sirimulla, S., Bailey, J. B., Vegesna, R. & Narayan, M. Halogen interactions in protein-ligand complexes: implications of halogen bonding for rational drug design. *J. Chem. Inf. Model.* **53**, 2781–2791 (2013).
9. Herrera-Rodriguez, L., Meyer, H, Robins, K. & Khan, F. Perspectives on biotechnological halogenation. *Chim. oggi/Chemistry Today* **29**, 31–33 (2011).
10. Radzicka, A & Wolfenden, R. A proficient enzyme. *Science* **267**, 90–93 (1995).
11. Zhao, H. Highlights of biocatalysis and biomimetic catalysis. *ACS Catal.* **1**, 1119–1120 (2011).

12. Schoemaker, H. E., Mink, D. & Wubbolts, M. G. Dispelling the myths--biocatalysis in industrial synthesis. *Science* **299**, 1694–1697 (2003).
13. Straathof, A. J. J., Panke, S. & Schmid, A. The production of fine chemicals by biotransformations. *Curr. Opin. Biotechnol.* **13**, 548–556 (2002).
14. Bornscheuer, U. T., Huisman, G.W, Kazlauskas, R.J, Lutz, S, Moore, J.C. & Robins, K. Engineering the third wave of biocatalysis. *Nature*. **485**, 185–194 (2012).
15. Kintses, B.,Hein, C, Mohamed, M.F, Fischlechner, M, Courtois, F, Laine, C. & Hollfelder, F. Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution. *Chem. Biol.* **19**, 1001–1009 (2012).
16. Yoo, T., Pogson, M., Iverson, B. & Georgiou, G. Directed Evolution of Highly selective proteases by using a novel FACS based screen that capitalizes on the p53 regulator MDM2. *ChemBioChem.* **13**, 649–653 (2012).
17. Fernandez-Alvaro, E., Snajdrova, R, Jochens, H, Davids, T, Bottcher, D. & Bornscheuer, U.T. A combination of *in vivo* selection and cell sorting for the identification of enantioselective biocatalysts. *Angew. Chem. Int. Ed. Engl.* **50**, 8584–8587 (2011).
18. Marcheschi, R.J., Gronenberg, L.S. & Liao, J.C. Protein engineering for metabolic engineering: current and next-generation tools. *Biotechnol. J.* **8**, 545–555 (2013).
19. Huang, C.J., Lin, H. & Yang, X. Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *J. Ind. Microbiol. Biotechnol.* **39**, 383–399 (2012).
20. Retallack, D. M., Jin, H. & Chew, L. Reliable protein production in a *Pseudomonas fluorescens* expression system. *Protein Expr. Purif.* **81**, 157–165 (2012).
21. Sieber, S.A. & Marahiel, M.A. Molecular mechanisms underlying nonribosomal peptide synthesis: approaches to new antibiotics. *Chem. Rev.* **105**, 715–738 (2005).
22. Strieker, M., Tanovic, A. & Marahiel, M.A. Nonribosomal peptide synthetases: structures and dynamics. *Curr. Opin. Struct. Biol.* **20**, 234–240 (2010).

23. Chiu, H.T., Hubbard, B.K., Shah, A.N., Eide, J., Fredenburg, R.A., Walsh, C.T. & Khosla, C. Molecular cloning and sequence analysis of the complestatin biosynthetic gene cluster. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8548–8553 (2001).
24. Sandercock, A. M., Charles, E.H., Scaife, W., Kirkpatrick, P. N., O'Brien, S.W., Papageorgiou, E.A., Spencer, J.B. & Williams, D.H. Biosynthesis of the di-meta-hydroxyphenylglycine constituent of the vancomycin-group antibiotic chloroeremomycin. *Chem. Commun.* **14**, 1252–1253 (2001).
25. Hubbard, B.K. & Walsh, C.T. Biosynthesis of Antibiotics Vancomycin Assembly: Nature's Way *Angewandte Chemie.* **42**, 730–765 (2003).
26. Drauz, K., Groger, H. & May, O. *Enzyme catalysis in organic synthesis. Wiley-VCH.* **1**, 21 (2012).
27. Shin, D., Rew, Y. & Boger, D. L. Total synthesis and structure of the ramoplanin A1 and A3 aglycons: two minor components of the ramoplanin complex. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11977–11979 (2004).
28. Cavalleri, B., Pagani, H, Vople, G, Selva, E. & Parenti, F. A16686, A new antibiotic from Actinoplanes I. Fermentation, isolation and preliminary physico-chemical characteristics. *J. Antibiot.* **37**, 309–317 (1983).
29. Parenti, F., Ciabatti, R, Cavalleri, B. & Ketterning, J. Ramoplanin: a review of discovery and its chemistry. *Drugs Exp Clin Res.* **16**, 451–455 (1990).
30. McCafferty, D.G., Cudic, P, Frankel, B.A, Barkallah, S, Kruger, R.G. & Li, W. Chemistry and biology of the ramoplanin family of peptide antibiotics. *Biopolymers.* **66**, 261–284 (2002).
31. Fang, X., Tiyanont, K, Zhang, Y, Wanner, J, Boger, D. & Walker, S. The mechanism of action of ramoplanin and enduracidin. *Mol. Biosyst.* **2**, 69–76 (2006)
32. Hamburger, J.B., Hoertz, A.J, Lee, A, Senturia, R.J, McCafferty, D.G. & Loll, P.J. A crystal structure of a dimer of the antibiotic ramoplanin illustrates membrane positioning and a potential Lipid II docking interface. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 13759–13764 (2009).

33. Wu, M.C. Bio-engineering of Antibiotic Enduracidin Biosynthetic Pathways and PreQ1 Riboswitch. Unpublished PhD thesis. University of Manchester. 120–134 (2011).
34. Yin, X., Chen, Y. & Zhang, L. Enduracidin analogues with altered halogenation patterns produced by genetically engineered strains of *Streptomyces fungicidicus*. *J. Nat. Prod.* **73**, 583–589 (2010).
35. Hager, L., Morris, D., Brown, F. & Eberwein, H. Chloroperoxidase II. Utilization of halogen anions. *J. Biol.* **241**, 1769–1777 (1966).
36. Rantwijk, F. V. & Sheldon, R. Selective oxygen transfer catalysed by heme peroxidases: synthetic and mechanistic aspects. *Curr. Opin. Biotechnol.* **11**, 554–564 (2000).
37. Hager, L., Morris, D., Brown, F. & Eberwein, H. Chloroperoxidase II. Utilization of halogen anions. *J. Biol. Chem.* **241**, 1769–1777 (1966).
38. Smith, D.R.M., Gruschow, S. & Goss, R.J.M. Scope and potential of halogenases in biosynthetic applications. *Curr. Opin. Chem. Biol.* **17**, 276–283 (2013).
39. Kühnel, K., Blankenfeldt, W., Ternier, J. & Schlichting, I. Crystal structures of chloroperoxidase with its bound substrates and complexed with formate, acetate, and nitrate. *J. Biol. Chem.* **281**, 23990–23998 (2006).
40. Vilter, H. Peroxidases from Phaeophyceae. III: Catalysis of Halogenation by Peroxidases from *Ascophyllum nodosum* (L.) Le Jol. *Bot. Mar.* **26**, 429–436 (2009)
41. Winter, J.M., Moffitt, M.C., Zazopoulos, E., McAlpine, J.B., Dorrestein, P.C. & Moore, B.S. Molecular basis for chloronium-mediated meroterpene cyclization: cloning, sequencing, and heterologous expression of the napyradiomycin biosynthetic gene cluster. *J. Biol. Chem.* **282**, 16362–16368 (2007).
42. Butler, A. & Carter-Franklin, J.N. The role of vanadium bromoperoxidase in the biosynthesis of halogenated marine natural products. *Nat. Prod. Rep.* **21**, 180–188 (2004).
43. Messerschmidt, A. & Wever, R. X-ray structure of a vanadium-containing enzyme: chloroperoxidase from the fungus *Curvularia inaequalis*. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 392–396 (1996).
44. Weyand, M., Hecht, H. J., Kieb, M., Liaud, M.F., Vilter, H. & Schomburg, D. X-ray structure determination of a vanadium-

- dependent haloperoxidase from *Ascophyllum nodosum* at 2.0 Å Resolution. *J. mol. biol.* **293**, 595–611 (1999).
45. Isupov, M.N., Dalby, A. R, Brindley, A.A, Izumi, Y, Tanabe, T, Murshudov, G.N. & Littlechild, J.A. Crystal structure of dodecameric vanadium-dependent bromoperoxidase from the red algae *Corallina officinalis*. *J. Mol. Biol.* **299**, 1035–1049 (2000).
 46. Messerschmidt, A. & Wever, R. X-ray structure of a vanadium-containing enzyme: chloroperoxidase from the fungus *Curvularia inaequalis*. *Proc. Natl. Acad. Sci. U. S. A.* **93**, 392–396 (1996).
 47. Itoh, N., Hasan, A.K, Izumi, Y. & Yamada, H. Substrate specificity, regiospecificity and stereospecificity of halogenation reactions catalyzed by non-heme-type bromoperoxidase of *Corallina pilulifera*. *Eur. J. Biochem.* **172**, 477–484 (1988).
 48. Krenn, B. E., Izumi, Y, Yamada, H. & Wever, R. A comparison of different (vanadium) bromoperoxidases; the bromoperoxidase from *Corallina pilulifera* is also a vanadium enzyme. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **998**, 63–68 (1989).
 49. Itoh, N., Izumi, Y. & Yamada, H. Haloperoxidase-catalyzed halogenation of nitrogen-containing aromatic heterocycles represented by nucleic bases. *Biochemistry.* **26**, 282–289 (1987).
 50. Tschirret-Guth, R.A. & Butler, A. Evidence for organic substrate binding to vanadium bromoperoxidase. *J. Am. Chem. Soc.* **116**, 411–412 (1994).
 51. Martinez, J.S., Carroll, G.L, Tschirret-Guth, R.A, Altenhoff, G, Little, R.D. & Butler, A. On the regiospecificity of vanadium bromoperoxidase. *J. Am. Chem. Soc.* **123**, 3289–3294 (2001).
 52. Coughlin, P., Roberts, S, Rush, C. & Willetts, A. Biotransformation of alkenes by haloperoxidases: Regiospecific bromohydrin formation from cinnamyl substrates. *Biotechnol. Lett.* **15**, 907–912 (1993).
 53. Orjala, J. & Gerwick, W.H. Barbamide, a chlorinated metabolite with molluscicidal activity from the Caribbean cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* **59**, 427–430 (1996).
 54. Hartung, J. The Biosynthesis of Barbamide—A Radical Pathway for “Biohalogenation”? *Angew. Chemie Int. Ed.* **38**, 1209–1211 (1999).
 55. Sitachitta, N., Rossi, J, Roberts, M.A, Gerwick, W.H, Fletcher, M.D. & Willis, C.L. Biosynthesis of the Marine Cyanobacterial Metabolite

- Barbamide . 1 . Origin of the Trichloromethyl Group. *J. Am Chem. Soc.* **2**, 7131–7132 (1998).
56. Vaillancourt, F. H., Yin, J. & Walsh, C.T. SyrB2 in syringomycin E biosynthesis is a nonheme Fe(II) alpha-ketoglutarate- and O₂-dependent halogenase. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10111–10116 (2005).
 57. Kelly, W., Boyne, M. & Yeh, E. Characterization of the aminocarboxycyclopropane-forming enzyme CmaC. *Biochemistry.* **46**, 359–368 (2007).
 58. Ueki, M., Galonic, D.P, Vaillancourt, F.H, Garneau-Tsodikova, S, Yeh, E, Vosburg, D. A, Schroeder, F.C, Osada, H. & Walsh, C.T. Enzymatic generation of the antimetabolite gamma, gamma-dichloroaminobutyrate by NRPS and mononuclear iron halogenase action in a streptomycete. *Chem. Biol.* **13**, 1183–1191 (2006).
 59. Butler, A. & Sandy, M. Mechanistic considerations of halogenating enzymes. *Nature.* **460**, 848–854 (2009).
 60. Murphy, C. D and Grant, R. "Chemicals production-biohalogenations" In: Timmis KN (eds). *Handbook of Hydrocarbon and Lipid Microbiology.* 2903-2910.
 61. Blasiak, L.C., Vaillancourt, F.H., Walsh, C.T. & Drennan, C.L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature.* **440**, 368–371 (2006).
 62. Hammer, P. E., Hill, D.S, Lam, S.T, Van Pee, K.H. & Ligon, J.M. Four genes from *Pseudomonas fluorescens* that encode the biosynthesis of pyrrolnitrin. *Appl. Environ. Microbiol.* **63**, 2147–2154 (1997).
 63. Dairi, T., Nakano, T. & Aisaka, K. Cloning and nucleotide sequence of the gene responsible for chlorination of tetracycline. *Bioscience.* **59**, 1099–1106 (1995).
 64. Dong, C., Flecks, S, Unversucht, S, Haupt, C, Van Pee, K.H. & Naismith, J.H. Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science.* **309**, 2216–2219 (2005).
 65. Whitteck, J.T. *Enzymatic halogenation in natural product biosynthesis.* Report presented at University of Illinois. 49–56 (2006).

66. Lee, J.K. & Zhao, H. Identification and characterization of the flavin:NADH reductase (PrnF) involved in a novel two-component arylamine oxygenase. *J. Bacteriol.* **189**, 8556–8563 (2007).
67. Yeh, E., Garneau, S. & Walsh, C.T. Robust *in vitro* activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 3960–3965 (2005).
68. Fujimori, D.G., Hrvatin, S, Neumann, C.S, Strieker, M, Marahiel, M. A. & Walsh, C.T. Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proc. Natl. Acad. Sci.U.S.A.* **104**, 16498–16503 (2007).
69. Keller, S., Wage, T. & Hohaus, K. Purification and partial characterization of tryptophan 7-halogenase (PrnA) from *Pseudomonas fluorescens*. *Angew. Chemie.***39**, 2300–2302 (2000).
70. Unversucht, S., Hollmann, F, Schmid, A. & van Pee, K.H. FADH₂-dependence of tryptophan 7-Halogenase. *Adv. Synth. Catal.* **347**, 1163–1167 (2005).
71. Fujimori, D.G., Hrvatin, S, Neumann, C.S, Strieker, M, Marahiel, M. A. & Walsh, C.T. Cloning and characterization of the biosynthetic gene cluster for kutznerides. *Proc. Natl. Acad. Sci.U.S.A.* **104**, 16498–16503 (2007).
72. Zehner, S., Kotzsch, A, Bister, B, Sussmuth, R.D, Mendez, C, Salas, J.A. & Van Pee, K.H. A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in *Streptomyces rugosporus* LL-42D005. *Chem. Biol.* **12**, 445–452 (2005).
73. Sheng, D., Ballou, D. P. & Massey, V. Mechanistic studies of cyclohexanone monooxygenase: chemical properties of intermediates involved in catalysis. *Biochemistry.* **40**, 11156–11167 (2001).
74. Van Pee, K.H. & Patallo, E.P. Flavin-dependent halogenases involved in secondary metabolism in bacteria. *Appl.Microbiol. Biotechnol.* **70**, 631–641 (2006).
75. Sanchez, C, Butovich, I.A, Brana, A.F, Rohr, J, Mendez, C. & Salas, J.A. The biosynthetic gene cluster for the antitumor rebeccamycin: characterization and generation of indolocarbazole derivatives. *Chem. Biol.* **9**, 519–531 (2002).

76. Glenn, W.S., Nims, E. & O'Connor, S.E. Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor. *J. Am. Chem. Soc.* **133**, 19346–19349 (2011).
77. Payne, J.T., Andorfer, M.C. & Lewis, J.C. Regioselective arene halogenation using the FAD-dependent halogenase RebH. *Angew. Chem. Int. Ed. Engl.* **52**, 5271–5274 (2013).
78. Heemstra, J.R. & Walsh, C.T. Tandem action of the O₂- and FADH₂-dependent halogenases KtzQ and KtzR produce 6, 7-dichlorotryptophan for kutzneride assembly. *J. Am. Chem. Soc.* **130**, 14024–14025 (2008).
79. Zeng, J. & Zhan, J. Characterization of a tryptophan 6-halogenase from *Streptomyces toxytricini*. *Biotechnol. Lett.* **33**, 1607–1613 (2011).
80. Van Pee, K.H. Microbial biosynthesis of halometabolites. *Arch. Microbiol.* **175**, 250–258 (2001).
81. Rachid, S., Krug, D., Kunze, B., Kochems, I., Scharfe, M., Zabriskie, T.M., Blocker, H. & Muller, R. Molecular and biochemical studies of chondramide formation—highly cytotoxic natural products from *Chondromyces crocatus* Cm c5. *Chem. Biol.* **13**, 667–681 (2006).
82. Buedenbender, S., Rachid, S., Muller, R. & Schulz, G.E. Structure and action of the myxobacterial chondrochloren halogenase CndH: a new variant of FAD-dependent halogenases. *J. Mol. Biol.* **385**, 520–530 (2009).
83. Lin, S., Van Lanen, S.G. & Shen, B. Regiospecific chlorination of (S)-beta-tyrosyl-S-carrier protein catalyzed by SgcC3 in the biosynthesis of the enediyne antitumor antibiotic C-1027. *J. Am. Chem. Soc.* **129**, 12432–12438 (2007).
84. Dorrestein, P. C., Yeh, E., Garneau-Tsodikova, S., Kelleher, N.L. & Walsh, C.T. Dichlorination of a pyrrolyl-S-carrier protein by FADH₂-dependent halogenase PltA during pyoluteorin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 13843–13848 (2005).
85. Zeng, J. & Zhan, J. A novel fungal flavin-dependent halogenase for natural product biosynthesis. *Chembiochem* **11**, 2119–2123 (2010).
86. Podzelinska, K., Latimer, R., Bhattacharya, A., Vining, L.C., Zechel, D.L. & Jia, Z. Chloramphenicol biosynthesis: the structure of CmlS, a flavin-dependent halogenase showing a covalent flavin-aspartate bond. *J. Mol. Biol.* **397**, 316–331 (2010).

87. Bitto, E., Huang, Y. & Bingman, C. The structure of flavin-dependent tryptophan 7-halogenase RebH. *Proteins Struct. Funct.* **70**, 289–293 (2008).
88. Zhu, X., Laurentis, W.D, Leang, K, Herrmann, J, Ihlefeld, K, Van Pee, K.H. & Naismith, J.H. Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *J. Mol. Biol.* **391**, 74–85 (2009).
89. Lang, A., Polnick, S, Nicke, T, William, P, Patallo, E.P, Naismith, J.H. & Van Pee, K.H. Changing the regioselectivity of the tryptophan 7-halogenase PrnA by site-directed mutagenesis. *Angew. Chemie.* **50**, 2951–2953 (2011)
90. Strieker, M., Nolan, E.M, Walsh, C.T. & Marahiel, M.A. Stereospecific synthesis of threo- and erythro-beta-hydroxyglutamic acid during kutzneride biosynthesis. *J. Am. Chem. Soc.* **131**, 13523–13530 (2009).
91. Quadri, L.E, Weinreb, P.H, Lei, M, Nakano, M.M, Zuber, P. & Walsh, C.T. Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry.* **37**, 1585–1595 (1998).
92. Altschul, S., Gish, W. & Miller, W. Basic local alignment search tool. *J. Mol.biol.* **215**, 403–410 (1990).
93. Van Pee, K.H. Enzymatic chlorination and bromination. *Methods Enzymol.* **516**, 237–257.
94. Gao, B. & Ellis, H.R. Altered mechanism of the alkanesulfonate FMN reductase with the monooxygenase enzyme. *Biochem. Biophys. Res. Commun.* **331**, 1137–1145 (2005).
95. Hudlicky, T. & Reed, J.W. Applications of biotransformations and biocatalysis to complexity generation in organic synthesis. *Chem. Soc. Rev.* **38**, 3117–3132 (2009).
96. Flecks, S., Patallo, E. & Zhu, X. New insights into the mechanism of enzymatic chlorination of tryptophan. *Angew. Chemie.* **47**, 9533–9536 (2008).
97. Rial, D.V. & Ceccarelli, E.A. Removal of DnaK contamination during fusion protein purifications. *Protein Expr. Purif.* **25**, 503–507 (2002).

98. Thain, A. & Gaston, K. A method for the separation of GST fusion proteins from co-purifying GroEL. *Trends Genet.* **12**, 209–210 (1996).
99. Clare, D.K., Vasishtan, D, Stagg, S, Quispe, J, Farr, G.W, Topf, M, Horwich, A.L. & Saibil, H.R. ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell* .**149**, 113–123 (2012).
100. Elad, N., Farr, G.W, Clare, D.K, Orlova, E.V, Horwich, A.L. & Saibil, H.R. Topologies of a substrate protein bound to the chaperonin GroEL. *Mol. Cell* . **26**, 415–426 (2007).
101. Sanctis, D.D., Rego, A.T, Marçal, D, McVey, C.E, Carrondo, M.A. & Enguita, F.J. Overexpression, purification and crystallization of the tetrameric form of SorC sorbitol operon regulator. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **64**, 22–24 (2008).
102. Hicks, L.D., Alattia, J.R, Ikura, M. & Kay, C.M. Multiangle laser light scattering and sedimentation equilibrium. *Calcium-Binding Protein Protoc.* **173**, 127–136 (2002).
103. McCoy, A.J., Grosse-Kunstleve, R.W, Adams, P.D, Winn, M.D, Storoni, L.C. & Read, R.J. Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
104. Winn, M.D., Ballard, C.C, Cowtan, K.D, Dodson, E.J, Emsley, Paul, E, Phil, R, Keegan, R.M, Krissinel, E.B, Leslie, A.G.W, McCoy, A, McNicholas, S.J, Murshudov, G.N, Pannu, N.S, Potterton, E.A, Powell, H.R, Read, R.J, Vagin, A. & Wilson, K.S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 235–242 (2011).
105. Ye, Q., Li, X, Yan, M, Cao, H, Xu, L, Zhang, Y, Chen, Y, Xiong, J, Ouyang, P. & Ying, H. High-level production of heterologous proteins using untreated cane molasses and corn steep liquor in *Escherichia coli* medium. *Appl. Microbiol. Biotechnol.* **87**, 517–525 (2010).
106. Tjernberg, A., Markova, N, Griffiths, W. J. & Hallen, D. DMSO-related effects in protein characterization. *J. Biomol. Screen.* **11**, 131–137 (2006).
107. Selling, G. W., Hamaker, S.A.H. & Sessa, D. J. Effect of solvent and temperature on secondary and tertiary structure of zein by circular dichroism. *Cereal Chem.* **84**, 265–270 (2007).

108. Wells, J.A., Powers, D.B, Bott, R.R, Graycar, T.P. & Estell, D.A. Designing substrate specificity by protein engineering of electrostatic interactions. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 1219–1223 (1987).
109. Wilson, C. & Agard, D. Engineering substrate specificity. *Curr. Opin. Struct. Biol.* **1**, 617–623 (1991).
110. Sambrook, J. Molecular cloning. A laboratory manual. (Cold Spring Harbor Laboratory Press, New York, 2001).
111. Inoue H., Nojima H. & O. H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene.* **96**, 23–28.
112. Laemmli, U. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature.* **227**, 680–685 (1970).

Appendix

List of Primers

Table 32. List of primers used for the halogenation study

Primer Name	Sequence	RE
Primer P1	aaaaaa catatg aaccagaaggtcgaccgg	<i>NdeI</i>
Primer P2	aaaaaa gcggccgc tcacacgtgccggccag	<i>NotI</i>
Primer P3	aaaaaa catatg accttgtgcgccacgca	<i>NdeI</i>
Primer P4	aaaaaa gcggccgc tactgcagaccgtagatcgggc	<i>NotI</i>
Primer P9	gggcatgccgggttgacggccgagcgctttgtcgccgaccggttcggtgg	-
Primer P6	cgtcagccaggtcgcggtcgtcgtccgcgaggacgcgccaggggat aagcggc	-
Primer P7	aaaaaa catatg atcggtgcatcacctgg	<i>NdeI</i>
Primer P8	aaaaaa ctcgag cgcattggcattacctg	<i>XhoI</i>
Primer P9	aaaaaa ggtacc atgacaaccttaagctgtaa	<i>KpnI</i>
Primer P10	aaaaaa ctcgag tcagataaatgcaaacgcatc	<i>XhoI</i>
Primer P11	aaaaaa catatg atggatgacaatcgaattcgg	<i>NdeI</i>
Primer P12	aaaaaa ctcgag cgtgccctggtgcaggc	<i>XhoI</i>
Primer P13	aaaaaa ggtacc atggatgacaatcgaattcgg	<i>KpnI</i>
Primer P14	aaaaaa gcggccgc ctacgtgccctggtgcaggc	<i>NotI</i>
Primer P15	gtgggtaccatggatgacaatcgattcggagcatccttgcctg	-
Primer P16	cacgtgttcgacgattccacgtactgggagacctt cgactacgaattcaaga	-
Primer P17	attcaagaatttctggtgaacggctgtactactgcatcttggcggcttgg	-
Primer P18	cacgtactacgagaccttcgactactggtcaagaatttctggtgaacggca	-
Primer P19	actacgagaccttcgactaccgttcaagaatttctggtgaa	
Primer P20	cgactacgaattcaagaataatggtgaacggcaactact	-
Primer P21	cgactacgaattcaagaatcgtggtgaacggcaactact	-
Primer P22	caccacctggtcagcattccacgaaatcagaccttcgactacgaattcaag	-
Primer P23	caccacctggtcagcattccacgcgctacgagaccttcgactacgaattcaag	
Primer P24	cacctggtcagcattccacgtacaagagaccttcgactacgaattcaagaac	

Primer P25	cacctcgttcgacgattccacgtaccgcgagaccttcgactacga attcaagaac	-
Primer P26	atggttaccgcaaaacaccggatcgtgtgttattgttggtggcaccgcaggt tggatgaccgcagcatatctgaa	
Primer P27	aaaaaa catatg atggttaccgcaaaacacc	<i>NdeI</i>
Primer P28	aaaaaa gcggccgc ttaggatcctcgagaagctt	<i>NotI</i>
Primer P29	tccgggtcatcattttatcatccggctgagcagatgcgtagcgttgatggtt	
Primer P30	cgaaaatgtgttcccgtattatcatggtctgcctccggagtcaaaaacttctggct gaac	
Primer P31	cccagcagaattgccatataactatagttcagccagaagttttgaacte	
Primer P32	gaagattactctaccatttggccggcagcgtgccgaactgcgac	
Primer P33	ggatcggcgtgggcgagggcaccctcccg agttgcagaaggtgttcttc	-
Primer P34	tcggcgtgggcgagggcaccatctcgagttgcagaaggtgttcttc gac	-
Primer P35	gaattcaagaacttctggttgccggcaactactactgcacatctt	-
Primer P36	ctccgggcccggcagctccgggcccggcagctggcgttctgagc	-
Primer P37	gggcccggagctgccgggcccggagatgcgtgcatcacctgg	-
Primer P38	aaaaaa gcggccgc ttacgatgggcattacct	
Primer P39	ccgcttcttcgccgccgcttcttcgccgccgctccgccagctggcgttctgagc	-
Primer P40	gcgaaagaagcggcggcgaaagaagcggcggcgaaagcggcggcgatgcgt gcatcacctggcg	-
Primer P41	gattccacgtactacgagaccttcgactacggcaactactactgcacatcttccggt tg	-
Primer P42	caagccggcaaagatgcagtagtagttgccgtgctgaaggtctcgtagtagt ggaatc	-
Primer P43	cattatccgtattatcatggctttgaaaccgagttcaaaaacttctggctgaac	-
Primer P44	ccagattcatggtaatccagctgtagttcagccagaagttttgaacte	-

List of constructs obtained during the period for studies

Table 32. List of constructs made for the halogenation studies

S. No	Plasmid Name	Characteristics/constructs	Marker/Antibiotic	Source/Reference
1	pET-KK1	420 bp <i>s-pcp</i> gene was constructed in pET28a using <i>NdeI/NotI</i> restriction enzymes	Kanamycin (50 µg/ml)	Actinoplanes ATCC30766
2	pET-KK2	975 bp <i>l-pcp</i> gene was constructed in pET28a using <i>NdeI/NotI</i> restriction enzymes	Kanamycin (50 µg/ml)	Actinoplanes ATCC30766
3	pET-KK3	pET28a:: <i>lcp</i> mutant obtained by mutation of 2 CGA residues in the sequence for expression of the protein.	Kanamycin (50 µg/ml)	-
4	pET-KK4	1476 bp <i>ram20</i> gene was constructed in pET28a using <i>NdeI/XhoI</i> restriction enzymes	Kanamycin (50 µg/ml)	Actinoplanes ATCC30766
5	pET-KK5	576 bp <i>ssuE</i> gene was constructed in pET28a using <i>NdeI/XhoI</i> restriction enzymes	Kanamycin (50 µg/ml)	<i>E. coli</i> BL21 DE3 cells
6	pET-KK6	702 bp <i>fre</i> gene was constructed in pET45b using <i>KpnI/XhoI</i> restriction enzymes	Ampicillin (50 µg/ml)	<i>E. coli</i> BL21 DE3 cells
7	pET-KK7	1602 bp <i>ktzQ</i> was constructed in pET28a using <i>NdeI/XhoI</i> restriction enzymes to produced N terminal His(6) tagged protein	Kanamycin (50 µg/ml)	<i>Kutzneria spp.744</i>
8	pET-KK8	1602 bp <i>ktzQ</i> was constructed in pET30a using <i>NdeI/XhoI</i> restriction enzymes to produced C terminal His(6) tagged protein	Kanamycin (50 µg/ml)	<i>Kutzneria spp.744</i>
9	pET-KK9	1602 bp <i>ktzQ</i> was constructed in pET45b using using <i>NdeI/XhoI</i> restriction enzymes to produced N terminal His(6) tagged protein	Ampicillin (50µg/ml)	<i>Kutzneria spp.744</i>
10	pT-KK10	pET45b:: <i>ktzQ</i> mutant obtained by mutation of CGA to CGC residue in the sequence for expression of the protein.	Ampicillin (50 µg/ml)	-
11	pET-SAS1	1614 bp <i>prnA</i> was constructed in pET28a using <i>NdeI/NotI</i> restriction enzymes.	Kanamycin (50 µg/ml)	<i>Pseudomonas fluorescens BL915</i>
12	pETKK-11	pET28a:: <i>prnA</i> Y444W mutant with mutation of Y residue to W residue at 444 position of <i>prnA</i>	Kanamycin (50 µg/ml)	-
13	pETKK-	pET28a:: <i>prnA</i> N459W mutant with	Kanamycin	-

	12	mutation of N residue to W residue at 459 position of <i>prnA</i>	(50 µg/ml)	
14	pETKK-13	pET28a:: <i>prnA</i> E450W mutant with mutation of E residue to W residue at 450 position of <i>prnA</i>	Kanamycin (50 µg/ml)	-
15	pET-KK14	pET28a:: <i>prnA</i> E450R mutant with mutation of E residue to R residue at 450 position of <i>prnA</i>	Kanamycin (50 µg/ml)	-
16	pET-KK15	pET28a:: <i>prnA</i> F454K mutant with mutation of F residue to K residue at 454 position of <i>prnA</i>	Kanamycin (50 µg/ml)	-
17	pET-KK16	pET28a:: <i>prnA</i> F454R mutant with mutation of R residue to K residue at 454 position of <i>prnA</i>	Kanamycin (50 µg/ml)	-
18	pET-KK17	pET28a:: <i>prnA</i> Y443K mutant with mutation of Y residue to K residue at 443 position of <i>prnA</i>	Kanamycin (50 µg/ml)	-
19	pET-KK18	pET28a:: <i>prnA</i> Y443R mutant with mutation of Y residue to R residue at 443 position of <i>prnA</i>	Kanamycin (50 µg/ml)	-
20	pET-KK19	pET28a:: <i>prnA</i> Y444K mutant with mutation of Y residue to K residue at 444 position of <i>prnA</i>	Kanamycin (50 µg/ml)	-
21	pET-KK20	pET28a:: <i>prnA</i> Y444R mutant with mutation of Y residue to R residue at 444 position of <i>prnA</i>	Kanamycin (50 µg/ml)	-
22	pET-KK21	<i>prnA</i> constructed into pET-YSBLIC vector from pET28a:: <i>prn</i> (pET-SAS1)	Kanamycin (50 µg/ml)	-
23	pET-SAS2	1544 bp <i>ktzR</i> gene constructed in pET28a using <i>NdeI/XhoI</i> restriction enzymes	Kanamycin (50 µg/ml)	-
24	pET-SAS3	1544 bp <i>ktzR</i> gene constructed in pET30a using <i>NdeI/XhoI</i> restriction enzymes	Kanamycin (50 µg/ml)	
25	pET-SAS4	1544 bp <i>ktzR</i> gene constructed in pVLT33 using <i>HindIII/XbaI</i> restriction enzymes	Ampicillin (50 µg/ml)	
26	pET-SAS5	1544 bp <i>ktzR</i> synthetic gene sub-cloned in in pET28a using <i>NdeI/XhoI</i> restriction enzymes	Kanamycin (50 µg/ml)	
27	pET-KK22	1604 bp <i>ktzR</i> constructed in pET28a using <i>NdeI</i> and <i>NotI</i> restriction enzymes	Kanamycin (50 µg/ml)	
27	pET-SAS6	1572 bp <i>SttH</i> was constructed in pET28a using <i>NdeI</i> and <i>XhoI</i>	Kanamycin (50 µg/ml)	

		restriction enzymes		
28	pET-KK23	pET28a:: <i>SttH</i> F98A mutant with single mutation of F to A residue at 98 position	Kanamycin (50 µg/ml)	
29	pET-KK24	pET28a:: <i>SttH</i> LI mutant with 8 amino acid insert between 462 and 463 position of <i>SttH</i>	Kanamycin (50 µg/ml)	-
30	pET-KK25	1536 bp <i>pyrH</i> constructed in pET28a using <i>NdeI</i> and <i>XhoI</i> restriction enzymes	Kanamycin (50 µg/ml)	
30	pET-KK26	pET28a:: <i>prnA</i> F103A mutant with mutation of F residue to A residue at 103 position of <i>prnA</i>	Kanamycin (50 µg/ml)	
31	pET-KK27	pET28a:: <i>prnA</i> I52F mutant with mutation of I residue to F residue at 52 position of <i>prnA</i>	Kanamycin (50 µg/ml)	
32	pET-KK28	pET28a:: <i>prnA</i> P53S mutant with mutation of P residue to S residue at 53 position of <i>prnA</i>	Kanamycin (50 µg/ml)	
33	pET-KK29	pET28a:: <i>prnA</i> N459S mutant with mutation of N residue to S residue at 459 position of <i>prnA</i>	Kanamycin (50 µg/ml)	
34	pET-KK30	pET28a:: <i>pyrH</i> F49I mutant with mutation of F residue to I residue at 49 position of <i>PyrH</i>	Kanamycin (50 µg/ml)	
35	pET-KK31	pET28a:: <i>pyrH</i> S50P mutant with mutation of S residue to P residue at 50 position of <i>PyrH</i>	Kanamycin (50µg/ml)	
46	pET-KK32	pET28a:: <i>pyrH</i> S455N mutant with mutation of S residue to N residue at 455 position of <i>PyrH</i>	Kanamycin (50 µg/ml)	
37	pET-KK33	pET28a:: <i>pyrH</i> F49I / S50P mutant with double mutation of F residue to I residue at 49 position and S to P residue at 50 position of <i>PyrH</i>	Kanamycin (50 µg/ml)	
38	pET-KK34	pET28a:: <i>pyrH</i> F49I/S50P/S455N mutant with triple mutation of F to I residue at 49 position, S to P residue at 50 position and S to N residue at 455 position of <i>PyrH</i>	Kanamycin (50 µg/ml)	
39	pET-KK35	pET28a:: <i>prnA</i> LD mutant with 7 amino acid deletion between 450 and 457 position of <i>PrnA</i>	Kanamycin (50 µg/ml)	
40	pET-KK36	pET28a:: <i>pyrH</i> LI mutant with 7 amino acid insert between 453 and	Kanamycin (50 µg/ml)	

		454 position of PyrH		
--	--	----------------------	--	--

Gene sequences

ssuE sequence

atgctgtcatcacctggcgggtagtcctcgtttccttctcgtccagctccttgctggaatatgcgcggga
aaaactaaatggcctggatgtagaggttatcactggaatctgcaaaacttcgccccggaagatctgcttatg
ctcgtttcgatagtcggcactcaagacctcaccgaacagctgcaacaggctgatggactgatagtcgctac
acctgtgataaagccgcctattccggcgcgctgaaaacctgctcgacctgctgccagagcgcgctttgca
aggcaagtgggtgctaccgctggcgacggcggtaccgtggccatctgctggcggctgattatgccctta
aaccagtttaagcgcactgaaagctcaggagatcctgcacggcgtgttggcagatgactacaagtaattga
ttaccatcacagaccccagttcacgccaaatctgcaaacccgtcttgataccgcgctagaaactttctggcag
gcattgcaccgccgcgatgttcaggttctgaccttctgtctctgcgaggaatgccatgcgtaa

Protein sequence

MGSSHHHHHHSSGLVPRGSHMMRVITLAGSPRFPSRSSSLLEYAREK
LNGLDVEVYHWNLQNFAPEDLLYARF DSPALKTFTEQLQQADGLIV
ATPVYKAAYSGALKTLLDLLPERALQGKVVLPATGGTVAHLLAV
DYALKPVLSALKAQEILHGVFADDSQVIDYHHRPQFTPNLQTRLDT
ALETFWQALHRRDVQVPDLLSLRGNHAH

Fre sequence

atgacaacctaaagctgtaaagtgacctcggtagaagctatcacggataccgtataccgtgccgaatcgta
cagacggccttttctttctgtgctgctcagtttgatgtagtgatggatgagcgtgacaaaacgtccgtct
caatggcttcgacccagatgaaaagggttcacgagctgcatattggcgttctgaaatcaacctttacgc
gaaagcagtcattgaccgcattctaaagatcatcaaatcgtggcgcgatattccgcacggtgaggcatggctg
cgcgatgatgaagaacgtccgatgatttgattgcgggtggcaccgggttctctatgcgcgctcattttgct
gacggcgtggcgcgtaacccaaaccgtgatataccattactggggcggcgtgaagagcagcatctgt
atgatctctgcgagcttgaggcgtttcgttaagcctcctggtctgcaagtgggtgccggtggtgaacaacc
ggaagcgggctggcgtggcgtactggcactgtgtaacggcgggtattgcaggatcatggtacgctagcag
agcatgatactatattgccggacgtttgagatggcgaanattgccctgacctgtttgcagtgagcgtaatg
cgcgggaagatgcctgtttggcgtatgcgtttgcattatctga

Protein sequence

MGSSHHHHHHSSGLVPRGSHMMTTLSCKVTSVEAITDTVYRVRIVP
DAAFSFRAGQYLMVVMDERDKRPFMSASTPDEKGFIELHIGASEINL
YAKAVMDRILKDHQIVVDIPHGEAWLRDDEERPMILIAGGTGFSYA
RSILLTALARNPNRDITIYWGGREEQHL YDLCELEALSLKHPGLQVV
PVVEQPEAGWRGRTGTVLTA VLQDHGTLAEHDIYIAGRFEMA?IAR
DLFCSEARNAREDR LFGDAFAFI

Sfp sequence

atgaagatttacggaatttatatggaccgccgctttcacaggaagaaaatgaacggttcatgtctttcatatca
cctgaaaaacgggagaaatgccggagatttatcataaagaagatgctcaccgcaccctgctgggagatgt
gctcgttcgctcagtcataaagcaggcagtatcagttggacaaatccgatatccgcttagcacgcaggaatac
gggaagccgtgcatccctgatcttcccagcgtcatttcaacatttctactccggacgctgggtcatttgcgc
gttgattcacagccgatcggcatagatatgaaaaaacgaaaccgatcagccttgagatcgccaagcgctt
ctttcaaaaacagagtacagcgaccttttagcaaaagacaaggacgagcagacagactattttatcatctat
ggccaatgaaagaaagctttatcaaacaggaaggcaaggcttatcgttccgcttgattcctttcagtcgcg
ctgcaccaggacggacaagtatccattgagcttccggacagccattccccatgctatatcaaacgtatgag
gtcgatccccggctacaaaatggctgtatcgccgtacaccctgattccccgaggatatcaaatggctcgt
acgaagagctttataa

Protein sequence

MKIYGIYMDRPLSQEENERFMSFISPEKREKCRRFYHKEDAHRTLLG
DVLVRSVISRQYQLDKSDIRFSTQEYKPCIPDLPAHFNISHSGRWV
ICAFDSQPIGIDIEKTKPISLEIAKRFFSKTEYSDLLAKDKDEQTDYFY
HLWSMKESFIKQEGKGLSLPLDSFSVRLHQDQVSIELPDSHSPCYIK
TYEVDPGYKMAVCAVHPDFPEDITMVSYEELL

s-pcp sequence

aaccagaaggctgaccgggccgccctgcccgcgcccgagcgggagacgacgacaccgggtaaggcac
ccgccccggaccgctcggcaacctcaggagtcgatgtccaggcgttcgccgaggtgctcggcctcg
acagcgtcggccggacgacgacttcttcgcctggcgccactcgtgctcggcgtcgcgctcgtgca
gcggctcaaggcacgcggtgctcggcgtcacggtgcaggacatcatggcgcgccacggctcggagct
gatgggctcgtgagcatgctgctgatccgggactccctcggcacgctcctgccgatccggcgcaccggc
gagctcccgcgctgttctgcgtccatccggccggcgggctcagctggtgctacctgccgctggccccgc
acgtg

Protein sequence

NQKVDRAALPAPERETTPGKAPAPGPLGNLEESMCQAFAEVLGLD
SVGPDDDDFFALGGHSLLAVALVQRLKARGVAVTVQDIMAAPTVSE
LMGSLSMSSIRDSLGTLLPIRRTGELPPLFCVHPAGGLSWCYLPLARH
V

l-pcp sequence

accttgcgccacgcagcatctgctggatgacgggggtgccgatcgggcccggcgttgacaacaccgcg
tctacgttctcagcaccctcctcggccgggtcccagcgggtgtggtgggggagctgtatgtggccgggtcg
ggctcggcgcgtggctatcggggcatccggggttgacggccgagcgatttgcgccaccggttctcgg
gggtggtcgcctctaccgaccgggtgatctggtccgggtggaccgacgacggtgtgctgcacttcgccgggc
ggccgatgatcagtgaaagatccgcggctatcgggtgagccgggagaggttgaagcgggttctggtca
acccccagctcagccaggtcgcggctcgtcgtccgagaggacgccaggggataagcggctggtcgc
cctatgctcggcggggatgctgagcgtatcgcagggagcgccttcccggctacatggttccgctcgcc
ttcgtccatctggaagcgtgccgctgaccgcgaaccagaaggtcgaccgggccgccctgcccgcgcc

gagcgggagacgacgacaccgggtaaggcaccggccccggaccgctcggcaacctcaggagtcgat
gtgccaggcggtcggcagggtgctcggcctcgacagcgtcggccccggacgacgacttctcgcctgggc
ggccactcgtgctcggcgtcgcgctcgtgcagcggctcaaggcacgcggtgctcggcgtcacggtgcagg
acatcatggccgcggccacggctcggagctgatgggctcgtgagcatgctgctgatccgggactccctc
ggcacgctcctgccgatccggcgcaccggcgagctgccgccgctgttctgctccatccggccggcggg
ctcagctggtgctacctgccgctggccccggcacgtgccggccgaccgcccgatctacggtctgcag

Amino acids

MGSSHHHHHSSGLVPRGSHM
TLCATQHLLDDGVPIGRPLDNTRVYVLDDLLRPVPTGVVVGELYVAG
SGLARGYAGMPGLTAERFVADPFSVGGRLYRTGDLVRWTDDGV LH
FAGRADDQVKIRGYRVEPGEVEAVLAQHPDVSQVA VVVREDAPGD
KRLVAYVVGDDVEAYAQERLPGYMVPSAFVHLEALPLTANQKVDR
AALPAPERETTTPGKAPAPGPLGNLEESMCQAF AEVLGLDSVGPDD
DFFALGGHSLLAVALVQRLKARGVA VTVQDIMAAPTVSELMGSL S
MSSIRDSLGTLLPIRRTGELPPLFCVHPAGGLSWCYLPLARHVPADRP
IYGLQ

ram 20 sequence

gcagcacaccgggaagaattgatgttattgttgggtggtggtcgggtgtagcaccgcagcagcactg
accgcaaacagggtgcaaaagtctgctgctggaacctgaaaaattccgcgttatcagattggtgaaagc
ctgctgccgagcaccgttcatggtgttgaatctgctgggtgttggtgatgaaattgcaaaagcaggtttatg
cgtaaacatggtggcacctttaaatggggcaccagcaccgaacctggacctttaccttgaaccagtcgg
cgtatggcagggtccgaccagccatgcatttcaggttgaacctcgtcgttttgatcagattctgctgaaaatgc
acgtcgtctgggtgtggtggtcgtgaaaatcatccggttaccgaagcaattgcagatgatgaacgtgttcgtg
gtgttcgtttaccaggatggtcagaccgtaccgcactggcacgtttgttggatgcaagcggtaacgta
gcacctgcataaccaggtggtgtacacgtgaatatagcccgttttctgtaactggcactgtttggctat
gaaaatggtcgtcgcctgcctgcaccgaatagcggtaaatattctgtgttgcattggtagcgggtgtttgg
tatattccgctgagcgaaccctgaccagcgttgggtcagttgtcgtcgtgaaatggcacataaagttcagg
gtgatcaagaaaaagccctgttgaactgattcagaatgtccgatgattgccgatttctgggtgatgcaacc
cgtgtgaccgaaggtgattatggtcagattcgtgttcgtaaagattattcctatagcagcaccagctattggcgt
ccgggtatgtgctgtaggtgatgcagcatgtttattgatccggttttagcagcgggtgtcatctggcaacct
atagcggctgctggcagcacgtagcattaatagcgttctggcagcaccgttgatgaagatcgtgcatttac
cgaattgaacagcgttatcgtcgcgaattgggtgtgttcatgatttctggtgagcttctatgatatgcatgttga
tgaagcagctattttgggcagcccgtaaagtaccgaaagcagcgcaccggcaatggaaagctttaccga
actggtgggtggtattgccagcgggaagatgccctgaccggtagcacagaactggttcgtcgtcatagccc
tcagaccgcagaactgggtcaggcagttgcaggtctggaagaaggtggcaccggtttctgctggttagca
gcgttgttcacaggcaatgttgaagtagccagattcaagccggtgcaattctgggtccggaaggcacc
aagaacagccgctgtttgaggggtggtctgacccgagcggtaatggcctgacctgggtgacagccgattaa

Protein sequence

VAAQPEEFDVIVVGGGPGGSTAAALTAKQGAKVLLLEREFPRYQI
GESLLPSTVHGV CNLLGVGDEIAKAGFMRKHGGTFKWGTSTEPWTF
TFATSPRMAGPTSHAFQVERRRFDQILLENARRLGVDVRENHPVTE
AIADDERVRGVRFTQDGQTRTALARFVVDASGNRSTLHTTVGGTRE
YSPFFRNALFGYFENGRRLPAPNSGNILCVAFSGSGFWYIPLSETLT
SVGAVVRREMAHKVQGDQEKALFELIAECPMIADFLGDATRVTG
DYGQIRVRKDYSYSSTSYWRPGMCLVGDAACFIDPVFSSGVHLATY
SGLLAARSINSVLAGTVDEEDRAFTEFEQR YRREFGVFHDFLVSFYDM
HVDESSYFWAARKVTESSAPAMESFTEL VGGIASGEDAL TGSTELV
RRHSRQTAELGQAVAGLEEGGTGFLRGSSVVAQAMFEGSQIQAGAI
LGPEGTQEQLFEGGLTPSGNGLTWVAAD

kztQ sequence

atggatgacaatcgaattcggagcaccctgtccttggcggcggcagccggctggatgtccgctgctac
ctgagcaaggcgctcgggcccggcgtcagggtcaccgtgctcaggcgcctccatctcgcgcatccgg
gtcggcaggccaccattccaacctgcacaaggtcttctc gacttctgggcatcggcaggacgagt
gatgcgggagtgcaaccagctacaaggccgggtccggctcgtcaactggcggacggcggcgacg
gccaggccacggcggcggcgtccggacggccggccgaccactc gaccactgtcggccagctgc
ccgagcacgagaacctgccgtgtcgcagtactgggcgcaccggcgcctcaacggcctgaccgacgaac
ccttcgaccgctctgctacgtgcagcccagctgctggaccgcaagctctcggcagggtgatggacggc
acgaaactggccagctacgcctggcacttcgacggcagctggtggccgacttctctgccggtcggcgt
gcagaagctgaacgtgaccacgtccaggacgtgttcacgcatccgacctc gaccagcggccacatc
acggccgtcaacaccagtgccggccgacgctggccggcagctgttcacgactgcagcggcttccgca
gcgtgctcatgggcaaggtcatgcaggagccgttctggacatgagcaagcactgtcaacgaccgcgc
ggtggcgtgatgctcccgcacgacgacgagaaggtgggcatc gaggcgtacacctcgtcgtggccatg
cggtcgggctggtcgtggaagatcccgtgctcggccgggtcggctccggctacgtctactccagccagttc
acctcccaggacgaggcggccgaggagctctgccgcatgtgggacgtc gaccggcggagcagacggt
caacaactccgggtccgggtcggccgacggccggcggcctgggtgcgcaactgcgtcggcatcggcgt
gtccgccatgttcgtggagccgtggagtcgaccggcctgtactcagttacgcctcgtctaccagctggtg
aagcacttcccggacaagcgggtccggccgacatcctggccgaccggtcaaccgcgaggtggcgacatgt
acgacgacaccggcacttctccaggcgcacttcagcctgtcggcgcgtgacgactccgagttctggcgg
gcctgcaaggagctgccgttcggcggacgggttcggcgagaaggtc gaggatgtacaggccgggctgccg
gtcgaactgccggtcaccatcgcgacgggcactactacggcaatttcgaggccgagttccgcaactctg
gaccaactcgaactactactgcatcttcggcgggctcggttcctgcccgagcatccgctgccggtgctcga
attccggccggaggccgtcgtcgcgcggagccggtgttcggcgcggtgcggcggcgcacggaggagc
tggtcggcaccggccggaccatgcaggcctacctgcggcgcctgcaccagggcacgtag

Protein sequence

MDDNRIRSILVLGGGTAGWMSACYLSKALGPGVEVTVLEAPSRIR
VGEATIPNLHKVFFDFLGLIAEDEWMRECNASYKAAVRFVNWRTPG
DGQATPRRRPDGRPDHFDHLFGQLPEHENLPLSQYWAHRRNLGLTD
EPFDRSCYVQPELLDRKLSRPLMDGTKLASYAWHFDADLVADFLCR
FAVQKLVNTHVQDVFTHADLDQRGHITAVNTESGRTLAAADLFIDCS
GFRSVLMGKVMQEPFLDMSKHLNDRAVALMLPHDDEKVGIEPYT
SSLAMRSGWSWKIPLLGRFGSGYVYSSQFTSQDEAAEELCRMWDV
DPAEQTFNNVRFVGRSRRAWVRNCVAIGVSAMFVEPLESTGLYFS
YASLYQLVKHFDPKRFPIADRFRNREVATMYDDTRDFLQAHFSL
PRDDSEFWRACKELPFADGFAEKVEMYRAGLPVELPVTIDDGHYYG
NFEAEFRNFWTNSNYCYCIFAGLGLPEHPLPVLEFRPEAVDRAEPVF
AAVRRRTEELVATAPTMQAYLRRLHQGT

prnA sequence

atgaacaagccaatcaagaatatcgtcatcgtggcgccggcaccgcgggctggatggccgcttctgacct
cgtccggcgctccaacagcaggtaaacatcacgctcatcagctctcggcgatccccggatcggcgtg
ggcgagggcaccatcccagtttcagaaaggtgtcttcgacttctcgggataccggagcgggagtgat
gccccaaagtgaacggcgcttcaaggccgcgatcaagttcgtgaactggagaaaatctcccaccatcgc
gcgaagattacttctaccatttctcggcagcgtgccgaactgcgacggcgtgccgcttaccactactggct
gcgcaagcgcgaacagggcttccagcagccgatggcgtacgcgtgctatccgcagccggggcctcga
cggcaagtggcaccctgcctggccgacggcaccgccagatgtccacgcgtggcacttcgacgcgca
cctgtgcccgacttctgaagcgtggccgtcgcagcgcggggtgaatcgcgtggtcgcgaggtcgtg
gaggttcaactgaacgaccggctacatctccacctgtaaccaaggaagggcggacgctggagggcg
acctgttcactgactgctccggcatcgcgagggctcctgatcaatcaggccctgaaggaaccttcacgacat
gtccgactacctgctgtgcgacagcgcggtcggcagcggcgtgcccacgacgacgcgcgaggggg
tcgagccttacacctccgcgattgcatgaactcgggatggacctggaagattccgatgctgggcccgttcg
gcagcggctactgttctcgcgcaagttcacctcgcgcgaccaggccaccgccgacttctcaactctgg
ggcctctcggacaatcagcagctcaaccagatcaagttccgggtcgggcgcaacaagcgggctgggtca
acaactgcgtctcgcgctgctcgtcgtcttctggagcccctggaatcgcgggaatctacttcatctac
gcggcgctttaccaactcgtgaagcacttccccgacacctcgttcgaccgcggttcgcgacgcattcaac
gccgagatcgtctacatgttcgacgactgccgagacttccagcgcactatttcaactcgcgcgaa
gacacgccgttctggctcgcgaaccggcacgaactcggctctcggacgccatccaggagaaggttgagc
gtacaagggcgggctgccactgaccaccacctcgttcgacgattccacgtactacgagaccttcgactacg
aattcaagaacttctggttgaacggcaactactactgcacttctccggcctgggcatgctgccgaccggtc
gctgccgctcctgcagcaccgaccggagtcgatccagaaggccgaagcgcgatgttcgccagcatccggcg
cgaagccgagcgcctcgcgacgagcctcccgacgaactacgactacctcgggtcactcgtgacggcgc
gcagctgtcgcgcaaccagcacgggcccgcgctcgcggctcaggaaccagtag

Protein Sequence

MNKPIKNIVIVGGGTAGWMAASYLVRALQQQVNITLIESAAIPRIGV
GEATIPSLQKVFFDFLGIPEREWMPQVNGAFKAAIKFVNWRKSPDPS
REDYFYHLFGSVPNCDGVPLTHYWLKREQGFQQPMA YACYPPQG
ALDGKLAPCLADGTRQMSHA WHFDAHLVADFLKRWA VERGVNRV
VDEVVEVQLNDRGYISTLLTKEGRTLEADLFIDCSGMRGLLINQALK
EPFIDMSDYLLCDSAVASAVPNDDAREGVEPYTSAIAMNSGWTWKI
PMLGRFGSGYVFSSKFTSRDQATADFLKLWGLSDNQQLNQIKFRVG
RNKRAWVNNCVSIGLSSCFLEPLESTGIYFIYAALYQLVKHFPDTSFD
PRLRDAFNAEIVYMFDDCRDFVQAHYFTTSREDTPFWLANRHELRL
SDAIQEKVERYKAGLPLTTTSFDDSTYYETFDYEFKNFWLNGNYICI
FAGLGMLPDRSLPLLQHRPESIQKAEAMFASIRREAERLRTSLPTNY
DYLRSLRDGAQLSRNQHGPTLAAQERQ

ktzR sequence

ATGGTTACCGCAAAAACACCGGATCGTGTTGTTATTGTTGGTGGT
GGCACCGCAGGTTGGATGACCGCAGCATATCTGAAAACCGCATT
TGGTGATCGTCTGAGCATTACCGTTGTTGAAAGCAGCCGTATTGG
CACCATTGGTGTGGTGAAGCAACCTTTAGCGATATCCAGCACTT
TTTTCAGTTTCTGAATCTGCGTGAACAGGATTGGATGCCTGCATG
TAATGCAACCTATAAACTGGGTATCCGCTTTGAAAATTGGCGTCA
TGTTGGCCATCATTTTTATCAGCCGTTTGAGCAGATTTCGTCCGGTT
TATGGTTTTCCGCTGACCGATTGGTGGCTGCATGATGCACCGACC
GATCGTTTTGATACCGATTGTTTTGTTATGCCGAATCTGTGTGAA
GCAGGTCGTAGTCCGCGTCATCTGGATGGCACCCCTGGCAGATGA
AGATTTTGTGTAAGAGGGTGATGAACTGGCAAATCGTACCATGA
GCGAACATCAGGGTAAAAGCCAGTTTCCGTATGCCTATCATTTTG
AAGCAGCACTGCTGGCAAATTTCTGACCGGTTATGCAGTTGATC
GTGGTGTGAAACATGTTGTTGATGATGTTCTGGATGTTCTGTCTGG
ATCAGCGTGGTTGGATTGAACATGTGGTTACCGCAGAACATGGT
GAAATTCATGGTGACCTGTTTGTTGATTGTACCGGTTTTTCGTGGT
CTGCTGCTGAATAAAGCACTGGGTGTTCCGTTTGTTAGCTATCAG
GATACCCTGCCGAATGATAGCGCAGTTGCACTGCAGGTTCCGCTG
GATATGCAGCGTCGTGGTATTGTTCCGAATAACCACCGCAACCGCA
CGTGAAGCAGGTTGGATTTGGACCATTCCGCTGTTTGGTCGTGTT
GGCACCGGTTATGTTTATGCAAAGATTATCTGAGTCCGGAAGA
AGCAGAACGTACCCTGCGTGAATTTGTTGGTCCGGCAGCAGCAG
ATGTTGAAGCAAATCATATTCGTATGCGTATTGGTCGTAGCCAAG
AAAGCTGGCGTAATAATTGTGTTGCAATTGGTCTGAGCAGCGGTT
TTGTTGAACCGCTGGAAAGCACCGGTATCTTTTTATCCATCATG
CAATTGAACAGCTGGTGAAACATTTTCCGGCAGCAGATTGGAAT
CCGAAAAGCCGTGATATGTATAATTCAGCAGTTGCCCATGTGATG

GATGGTATTCGTGAATTTCTGGTGATTCATTATCGTGGTGCAGCA
CGTGCAGATAATCAGTATTGGCGTGATACCAAACCCGTCCGCT
GCCGGATGGTCTGGCAGAACGTATTGAATGTTGGCAGACCCAGC
TGCCGGATACCGAAACCATTTATCCGTATTATCATGGTCTGCCTC
CGTATAGCTATATGTGTATTCTGATGGGTGGTGGTGCAATTCGTA
CACCGGAAGCGCAGCACTGGCACTGACCGATCAGGGTGCAGCA
CAGAAAGAATTTGCAGCAGTTCGTGATCGTGCAGCACAGCTGCG
TGATACTGCCGAGCCATTATGAATATCTGGCACGTATGCGTGG
TCTGGATGTTAACTCGAGAAGCTTCTCGAGGATCCTAA

ktzR synthetic gene sequence

atggttaccgaaaaacaccggatcgtggttattggtgggtggcaccgcaggtggatgaccgcagcat
atctgaaaaccgcaattggtgatcgtctgagcattaccggttgaagcagccgattggcaccattggtgtg
gtgaagcaacctttagcgatataccagcactttttcagtttctgaatctgcgtgaacaggattggatgcctgcat
gtaatgcaacctataaactgggtatccgcttgaaaattggcgtcatgtggccatcattttatcagccgttga
gcagattcgtccggtttatggtttccgctgaccgattggtggctgcatgatgcaccgaccgatcgtttgatac
cgattgtttggtatccgaatctgtgtgaagcaggtcgtatccgcgtcatctggatggcaccctggcagatg
aagattttgtgaagagggtgatgaactggcaaatcgtaccatgagcgaacatcagggtaaaagccagttcc
gtatgcctatcatttgaagcagcactgctggcaaaatttctgaccggtatgcagttgatcgtggtgttgaacat
gttgtgatgatgttctggatgttcgtctggatcagcgtggttgattgaacatgtggtaccgcagaacatggt
gaaattcatggtgacctgtttgttattgtaccggtttctggtctgctgctgaataaagcactgggtgtccgtt
tgttagctatcagataccctgccgaatgatagcgcagttgactgcagttccgctggatgacagcgtcgt
ggtattgtccgaataaccaccgcaaccgcagtggaagcaggtggattggaccattccgctgtttggtcgtgt
tggcaccggttatgtttatgcaaaagattatctgagtcgggaagaagcagaacgtaccctgcgtgaattgtt
gtccggcagcagcagatgttgaagcaaatcatattcgtatgcgtattggtcgtagccaagaaagctggcgt
ataattgtttgcaattggtctgagcagcggttttgtgaaccgctggaaagcaccggtatcattttatccatcat
gcaattgaacagctggtgaaacattttccggcagcagattggaatccgaaaagccgtgatatgataattcag
cagttgcccatgtgatggatggtattcgtgaatttctggtgattcattatcgtggtgcagcacgtgcagataac
agtattggcgtgataccaaaaccgtccgctgccggatggtctggcagaacgtattgaatgttggcagacc
agctgccggataaccgaaaccatttatccgtattatcatggtctgcctccgtatagctatatgtgtattctgatggg
tgggtggtgcaattcgtacaccggcaagcgcagcactggcactgaccgatcagggtgcagcacagaaagaa
tttcagcagttcgtgatcgtgcagcacagctgcgtgatacactgccgagccattatgaatatctggcacgat
cgctggtctggatgtttaaactcgagaagcttctcgaggatcc taa

Full protein sequence

MVTAKTPDRVVIVGGGTAGWMTAAYLKTAFGDRLSITVVESSRIGT
IGVGEATFSDIQHFFQFLNLREQDWMPACNATYKLGIRFENWRHVG
HHFYQPFEQIRPVYGFPLTDWWLHDAPTDRFDTDVMPNLCCEAGR
SPRHLDGTLADEFVEEGDELANRTMSEHQKGSQFPYAYHFEEALL
AKFLTGYAVDRGVEHVVDVLDVRLDQRGWIEHVVTAEHGEIHGD

LFVDCTGFRGLLLNKALGVPFVSYQDTLPNDSAVALQVPLDMQRR
GIVPNTTATAREAGWIWTIPLFGRVGTGYVYAKDYLSPEEAERTLRE
FVGPAADVEANHIRMGRIGRSQESWRNNCVAIGLSSGFVEPLESTGI
FFIHHAIEQLVKHFPAADWNPKSRDMYNSAVAHVMDGIREFLVIHY
RGAARADNQYWRDTKTRPLPDGLAERIECWQTQLPDTETIYPYYHG
LPPYSYMCILMGGGAIRTPASAALALTDQGAAQKEFAAVRDRAAQL
RDTLPSHYEYLARMRGLDV

SttH sequence

atgaatacccgcaatccggataaagttgttattgttggtggcaccgcaggttgatgaccgcaagctatct
gaaaaaagcatttggtaacgtgttagcgttacctggtgaaagcggcaccattggcaccgttggtgttgg
gaagcaaccttagcgatattgccactttttgaattctggatctgcgtgaagaggaatggatgcctgcatgt
aatgcaacctataaactggcagttcgtttcaggattggcagcgtccgggtcatcattttatcatccgttgagc
agatgcgtagcgtttaggtttccgctgaccgattggtggctgcagaatggtccgaccgatcgtttgatcgt
gattgtttgttatggcaagcctgtgtgatgcaggtcgtagtcgcgttatctgaatggtagcctgctgcaaca
gaattgatgaacgcgcagaagaaccggcaggtctgacatgagcgaacatcagggtaaaaccagtttcc
gtatgcctatcatttgaagcagcactgctggcagaattctgagcggttatagcaaagatcgtggtgtaaaaca
tgttgggatgaagtctggaagtgaaactggatgctggttggttagccatggttaccaaagaacatg
gtgatattggtggtgacctgtttgttattgaccggtttcgtggtgtgctgctgaatcaggcactgggtgtcc
gtttgttagctatcaggataccctgccgaatgatagcgcagttgactgcaggtccgctggatggaagca
cgtggtattccgccttataccctgcaaccgaaaagaagccggttgattggaccattccgctgattggtc
gtattggtacaggttatgtttatgccaaagattactgtagtcgggaagaagcagaacgtacctgcgtgaatt
gttggccggaaagcagccgatgttgaagcaaatcatattcgtatgcgtattggtcgttcagaacagagctgga
aaaataactgtgttgaattggtctgagcagcgggtttgtgaaccgctggaaagcaccggtattttctttatcca
tcatgcaattgagcagctggtgaaacattttccggcagggcattggcactcctcagctgcgtgcaggttataatt
cagcagttgcaaatgttatggatggcgttcgtgaattcctggttctgcattatctgggtgcagcagtaataata
cccgttattggaagataccaaaaccgtgcagttccggatgactggccgaacgtattgaacgttggaaag
ttcagctccggatagcgaatgtttcccgattatcatggtctgcctccgtatagttatggaattctgct
gggtacaggtgccattggtctgcgtccgagtcggcactggcctggcagatcctgcagcagcagaaaaa
gaatttaccgcaattcgtgatcgtgcacgtttctggtgatacactgccgagccagtatgaatattttgcagcaa
tgggtcagcgtgtgtaa

Protein sequence

MNTRNPDKVVIVGGGTAGWMTASYLKKAFGERVSVTLVESGTIGT
VGVGGEATFSDIRHFFEFLLDREEEWMPACNATYKLAVRFQDWQRP
GHHFYHPFEQMRSVDGFPLTDWWLQNGPTDRFDRDCFVMASLCDA
GRSPRYLNGSLLQQEFDERAEPPAGLTMSEHQGKTQFPYAYHFEAA
LLAEFLSGYSKDRGVKHVVDEVLEVKLDDRGWISHVVTKEHGDIG
GDLFVDCTGFRGVLLNQALGVPFVSYQDTLPNDSAVALQVPLDME
ARGIPPYTRATAKEAGWIWTIPLIGRIGTGYVYAKDYCSPEEAERTL

REFVGPAAADVEANHIRMGRSEQSWKNNCVAIGLSSGFVEPLEST
GIFFIHAIEQLVKHFPA GDWHPQLRAGYNSAVANVMDGVREFLVL
HYLGAARNDTRYWKDTRAVPDALAERIERWKVQLPDSENVFPY
YHGLPPYSYMAILLGTGAIGLRPSPALALADPAAAEKEFTAIRDRAR
FLVDTLPSQYEYFAAMGQRV

pyrH sequence

atgattcgtagcgttgattgttggtggcaccgcaggtggatgaccgcaagctatctgaaagcagcatt
tgatgatcgtattgatgtaccctgggtgaaagcggtaatgttcgtcgtattggttggtgaagcaacctttagc
accgttcgtcactttttgattatctgggtctggatgaacgtgaatggctgcctcgttgccgggtggtataaact
gggtattcgtttgaaaattggagcgaaccgggtgagtattctatcatccgtttgaacgtctgcgtggtgat
ggctttaatatggcagaatggggctggcagttggtgacgtcgtaccagctttagcgaagcatgtatctgac
ccatcgtctgtggaagcaaacgtgcaccgcgtatgctggatggtagcctggtgcaagccaggtgatgaa
agcctgggtcgtagcaccctggcagaacagcgtgcacagttccgtatgcatatcattttgatgcagatgaag
ttcccgttatctgagcgaatatgcaattgccgtggttcgtcatgttggtgatgatgttcagcatggtggcca
ggatgaacgcgggttgattagcgggtgtcataccaaacagcatggtgaaattagtggtagcctggttgatt
gtaccggtttcgtggtctgctgattaaccagacctgggtggtcgtttcagagcttttagtgatgtctgccgaa
taatcgtgcagttgactgcgtgtccgcgtgaaaatgatgaagatatgctccgtataccaccgaaccgca
atgagtgccgggttgatgtggaccattccgctgtttaaacgtgatggtaatggctatgtttatagcgacgaatt
atcagtcgggaagaagcagaacgcgaactgcgtagtagcaccgggtcgtgatgatctggaagcaa
atcacattcagatgcgtattggtcgtaatgaacgtacctggattaataactgtgttcagttggtctgagcgcag
cctttgtgaaaccgtggaagcaccggtattttctttatcagcatgcaattgagcagctggtgaaacatttcc
gggtgaaactggtggatccgggtctgattagcgcataatgaacgcattggcacacatggtgatggcgtgaa
agaatttctggtgctgattataaaggtgcacagcgtgaagataccccgtattgaaagcagccaaaaccg
tgcaatgccggatggtctggcacgtaaacggaactgagcgaagccatctgctggatgagcagaccattta
tccgtattatcatggctttgaaacctacagctggattaccatgaatctgggctgggtattgtccggaactcc
gcgtccggcactgctgcacatggatcctgcaccggcactggcagaattgagcgcctgcgtcgtgaaggtg
atgaaactgattgcagcactgccgagctgttatgaatatctggcaagcattcagtaa

Protein sequence

MIRSVVIVGGGTAGWMTASYLKAADFDRIDVTLVESGNVRRIGVGE
ATFSTVRHFFDYLGDEREWLPRCAGGYKLGIRFENWSEPGEYFYH
PFERLRVVDGFNMAEWWLAVGDRRTSFSEACYLTHRLCEAKRAPR
MLDGSLFASQVDESLGRSTLAEQRAQFPYAYHFDADEVARYLSEYA
IARGVRHVDDVQHVQDERGWISGVHTKQHGEISGDLFVDCTGF
RGLLINQTLGGRFQSFSVDLPNNRAVALRVPRENDEDMRPYTTATA
MSAGWMWTIPLFKRDGNGYVYSDEFISPEEAERELRSTVAPGRDDL
EANHIQMRIGRNERTWINNCVAVGLSAAFVEPLESTGIFFIQHAIEQL
VKHFPGERWDPVLISAYNERMAHMVDGVKEFLVLHYKGAQREDTP
YWKAAKTRAMPDGLARKLELSASHLLDEQTIYPYHGFETYSWIT

MNLGLGIVPERPRPALLHMDPAPALAEFERLRREGDELIAALPSCYE
YLASIQ