STRUCTURAL AND BIOCHEMICAL STUDIES ON THE BIOSYNTHETIC PATHWAYS OF CYANOBACTINS

Andrew F. Bent

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews

2016

Structural and Biochemical Studies on the Biosynthetic Pathways of Cyanobactins

A thesis submitted in fulfilment for the degree of

Doctor of Philosophy

March 2016

Andrew F. Bent

Supervisor - Prof. James H. Naismith

II. Abstract

Cyclic peptides have potential as scaffolds for novel pharmaceuticals, however their chemical synthesis can be challenging and as such natural sources are often explored. Several species of cyanobacteria produce a family of cyclic peptides, the cyanobactins, through the ribosomal synthesis of precursor peptides and post-translational tailoring. The patellamides, a member of the cyanobactin family, are cyclic octapeptides containing D-stereo centres and heterocyclised amino acids. A single gene cluster, *patA - patG,* contains the genes for the expression of the precursor peptide and the enzymes responsible for post-translational modifications including a heterocyclase, protease, macrocyclase and oxidase. Biochemical and structural analysis on the patellamide and related cyanobactin pathways has been carried out.

The crystal structure of PatF, a proposed prenyl transferase, has been determined, highlighting that it is likely evolutionary inactive due to changes to key residues when compared to active homologues. This is in agreement with the knowledge that no naturally prenylated patellamides have been discovered to date.

The crystal structure of the macrocyclase domain of PatG has been determined in complex with a substrate analogue peptide. The structure, together with biochemical analysis has allowed a mechanism of macrocyclisation to be proposed, confirming the requirement of a specific substrate conformation to enable macrocyclisation.

Using isolated enzymes from the patellamide and related pathways, a small scale library of macrocycles made up of diverse sequences has been created *in vitro* and characterised by mass spectrometry and in certain cases NMR. In order to further enhance diversity, macrocycles containing unnatural amino acids have been created using three approaches; SeCys derived precursor peptides, intein-mediated peptide ligation and pEVOL amber codon technology.

Finally, two oxidase enzymes from cyanobactin pathways have been purified, characterised and confirmed active for thiazoline oxidation. Native X-ray datasets on crystals of the oxidase CyaGox have been collected and phasing trials are on-going.

III. Declaration

I, Andrew F. Bent, hereby certify that this thesis, which is approximately 50,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in September, 2011 and as a candidate for the degree of Doctor of Philosophy in August, 2015; the higher study for which this is a record was carried out in the University of St Andrews between 2011 and 2015.

Date …….. Signature of candidate ……..

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Doctor of Philosophy in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date …….. Signature of supervisor ……..

IV. Copyright Declaration

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Access to all of printed copy but embargo of all electronic publication of thesis for a period of one year on the following grounds:

Publication would be commercially damaging to the researcher, or to the supervisor, or the University

Date ..... Signature of candidate .....

Date ..... Signature of supervisor .....

V. Contents Page

VI. List of Figures and Tables

VI.1 List of Figures

VI.1 List of Tables

VII. Abbreviations

ADP	Adenosine diphosphate
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
AU	Asymmetric unit
BME	β-mercaptoethanol
β-OG	n-octyl-β-D-glucoside
CCD	Charge coupled device
CCP4	Collaborative Computing Project Number 4
DMAPP	Dimethylallyl pyrophosphate
DMATS	Dimethylallyl tryptophan synthase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease I
DTT	Dithiothreitol
FMN	Flavin mononucleotide
HPLC	High-performance liquid chromatography
IPTG	Isopropylthio-β-D-galactoside
LAPs	Linear azol(in)e-containing peptides
LB	Luria-Bertani (growth medium)
LC	Liquid chromatography
MALDI	Matrix-assisted laser desorption ionisation
MBP	Maltose-binding protein
MME	Monomethyl ether
MR	Molecular replacement
MS	Mass spectrometry
MW	Molecular weight
NMR	Nuclear magnetic resonance

OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
PPi	Pyrophosphate
RiPPs	Ribosomally synthesized and post-translationally modified peptides
RNA	Ribonucleic acid
SAD	Single-wavelength anomalous dispersion
SAR	Structure-activity relationship
SDM	Site-directed mutagenesis
SDS	Sodium dodecylsulfate
SEC	Size-exclusion chromatography
SUMO	Small ubiquitin-like modifier
TCEP	Tris(2-carboxyethyl)phosphine
TEV	Tobacco Etch Virus
TLS	Translation, libration and screw-rotation
TOF	Time of flight
TOMMS	Thiazole/oxazole-modified microcins
tRNA	Transfer ribonucleic acid
UV	Ultra-violet

Standard Amino Acids

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

Non-Standard Amino Acids

Benzophenylalanine	Bpa	Selenocysteine	SeCys
Selenomethionine	SeMet	Hydroxyproline	Hyp
Thienylalanine	Thi		

Heterocyclised Amino Acids

Oxazolines (Serine)	S^{Oxn}	Oxazoles (Serine)	S^{Oxz}
Oxazolines (Threonine)	T^{Oxn}	Oxazoles (Threonine)	T^{Oxz}
Thiazolines (Cysteine)	C^{Thn}	Thiazoles (Cystine)	C^{Thz}
Selenazolines	$SeCys^{Sen}$	Selenazoles	$SeCys^{Sez}$

Nucleic Acids

A- Adenosine C- Cytosine G- Guanine T- Thymidine

VIII. Acknowledgements

First and foremost my thanks go to Prof Jim Naismith for allowing me the fantastic opportunity to study for my PhD and for his guidance and support throughout. I would like to thank Dr Jesko Koehnke, Mr Greg Mann and Dr Emilia Oueis and the numerous project students for making the entire patellamide project viable by collaborative discussion. Thanks to Prof. Marcel Jaspars and his laboratory (University of Aberdeen) for continued collaboration, in particular the transfer of DNA constructs and synthetic peptides. My thanks go to the entire Naismith laboratory in particular Dr Huanting Liu, Dr Judith Reeks and Mr Fraser Duthie for their help in getting me started with molecular biology and protein purification. A massive thank you must go to BSRC Mass Spectrometry team of Dr Catherine Botting, Dr Sally Shirran and Dr Matt Fuszard for their fast and efficient service which has been crucial in many of my studies. Thanks to Prof Nick Westwood for useful discussion, Dr Uli Schwarz-Linek for NMR analysis, Prof Tom Muir (Princeton University) and his group for helpful tips with intein ligation and to Prof Satish Nair (University of Illinois) for the kind sharing of the unpublished PagF structure.

Finally, my thanks go to my parents Frank and Irene, my brothers Colin and David and my wonderful fiancée Emma for their support and understanding through my PhD.

IX. Publications

IX.1 Journal Articles Published

J. Koehnke*, **A. Bent***, W.E. Houssen*, D. Zollman, F. Morawitz, S. Shirran, J. Vendome, A.F. Nneoyiegbe, L. Trembleau, C.H. Botting, M.C.M. Smith, M. Jaspars & J. H Naismith, *The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain*, Nature Structural & Molecular Biology, **19**, 767-773, doi: 10.1038/nsmb.2340 (2012).

J. Koehnke, F. Morawitz, **A.F. Bent**, W.E. Houssen, S.L. Shirran, M.A. Fuszard, I.A. Smellie, C.H. Botting, M.C.M. Smith, M. Jaspars & J. H Naismith, *An enzymatic route to selenazolines*, ChemBioChem, **14**, 564-567, doi: 10.1002/cbic.201300037 (2013).

A.F. Bent, J. Koehnke, W.E. Houssen, M.C.M. Smith, M. Jaspars & J. H Naismith, *Structure of PatF from Prochloron didemni*, Acta. Cryst. F, **69**, 618-623, doi: 10.1107/S1744309113012931 (2013).

J. Koehnke*, **A.F. Bent***, D. Zollman, K. Smith, W.E. Houssen, X. Zhu, G. Mann, T. Lebl, R. Scharff, S. Shirran, C.H. Botting, M. Jaspars, U. Schwarz-Linek & J. H Naismith, *The cyanobactin heterocyclase enzyme operates via an adenylation mechanism and is processive with a defined order of reaction.* Ange. Chem. Int. Edt. **52**, 13991-13996 doi: 10.1002/anie.201306302 (2013)

W.E. Houssen*, **A.F. Bent***, A.R. McEwan, N. Pieiller, J. Tabudravu, A. Raab, J. Koehnke, G. Mann, R.I. Adaba, L. Thomas, U. Hawas, M.C.M. Smith, J.H. Naismith and M. Jaspars - *An Efficient Method for the in vitro Production of Azol(in)e-based Cyclic Peptides.* Ange. Chem. Int. Edt. **53**, 14171-14174 doi: 10.1002/anie.201408082 (2014)

G. Mann, J. Koehnke, **A.F. Bent**, R. Graham, W.E. Houssen, M. Jaspars and J.H. Naismith, *Structural studies of the cyanobactin DUF domains*. Acta. Cryst. F **70**, 1597-1603, doi: 10.1107/S2053230X1402425X (2014).

J. Koehnke*, **A.F. Bent***, W.E. Houssen*, G. Mann*, M. Jaspars* and J.H. Naismith*, *The structural biology of cyanobactin synthesis.* Curr. Opin. Struc. Biol. **29**, 112-121, doi: 10.1016/j.sbi.2014.10.006 (2014).

J. Koehnke*, G. Mann*, **A.F. Bent***, H. Ludewig, S. Shirran, C. Botting, T. Lebl, W. Houssen, M. Jaspars, and J.H. Naismith, *Structural analysis of leader peptide binding enables leader-free cyanobactin processing*. Nature Chem. Biol. **11**, 558-563, doi: 10.1038/nchembio.1841 (2015)

* Authors contributed equally.

Copies of all published papers are enclosed at the rear of this thesis.

IX.2 Patents Filed

WO Application No: PCT/GB2013/051735 – Naismith *et al.* - Production of Cyclic Peptides

1. Introduction

1.1 Cyclic Peptides, RiPPs and Cyanobactins

1.1.1 Cyclic Peptides

Cyclic peptides have long been of interest to the biotechnology and pharmaceutical industries as novel therapeutics. They already have diverse applications in biology including antibiotics (e.g. Daptomycin [1], [2] Figure 1.1 A), immunosuppressants (e.g. Cyclosporine A [3] Figure 1.1 B), anti-cancer drugs (e.g. Somatostatin [4] Figure 1.1 C), and also as various probes and tools in biotechnology [5] (Figure 1.1).

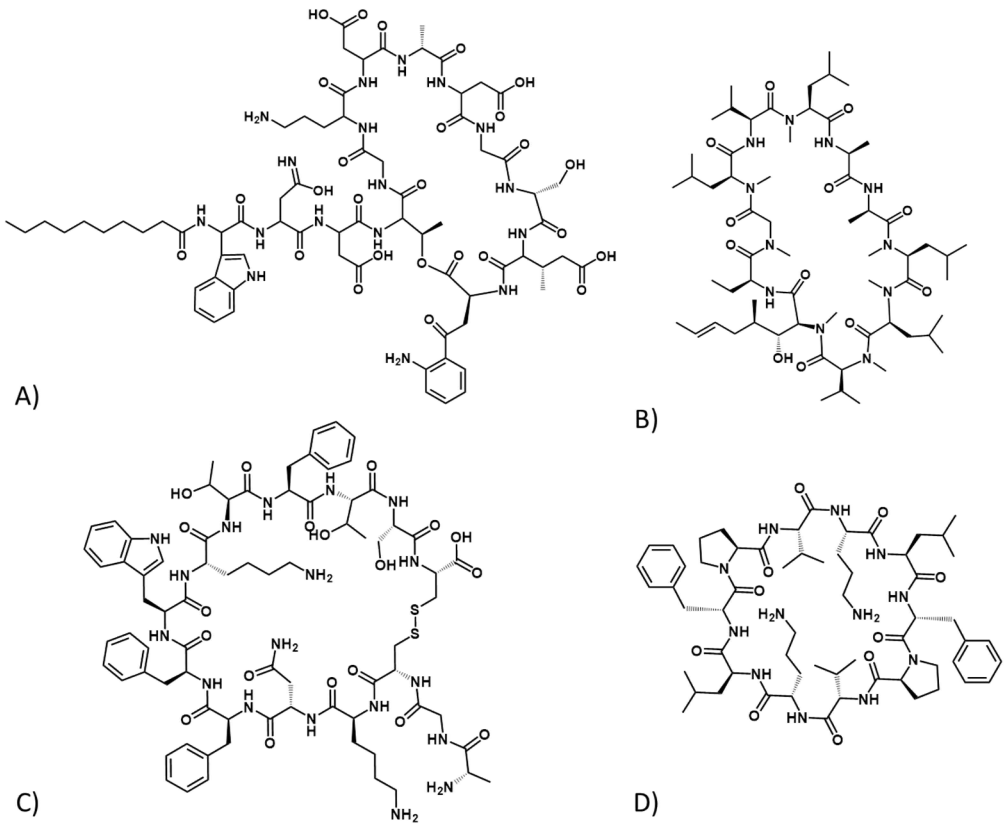


Figure 1.1: Chemical Structures of Cyclic Peptides. Chemical structures of (A) the antibiotic Daptomycin, (B) the immunosuppressant Cyclosporine A, (C) the anti-cancer drug Somatostatin and (D) the antibiotic Gramicidin S.

Cyclic peptides possess several properties that make them interesting pharmaceutical entities. They are considerably more stable (less prone to breakdown) than linear peptides due to their conformational rigidity and their lack of N- and C- termini means they are less prone to degradation by exo-proteases. They have increased bioavailability as the absence of the charged termini facilitates their movement across lipid membranes [6]. In some cases, stability is further increased by the presence of internal bonding between side chains e.g. disulfide bonds between cysteine residues [7]. As drug molecules, cyclic peptides often occur beyond the molecular weight limits defined by the Lipinski "rule of five" guidelines for small molecule compounds [8], however their size and increased chemical diversity can offer improvements in potency and specificity [9]. The use of peptides also bridges the chemical space gap between small molecules and biologics (e.g. antibodies).

The chemical synthesis of cyclic peptides is particularly challenging due to cost and low yields [10] and achieving diversity is difficult due to limitations in chemical composition [11]. As a result, there is interest in examining the biosynthesis of natural cyclic peptides and harnessing the potential for utilising these processes to generate novel diverse compounds [12] [13]

Many cyclic peptide molecules used in therapies today are derived from natural sources, e.g. Cyclosporine A from the fungi *Tolypocladium inflatum* [3], Daptomycin *(Streptomyces roseosporus*) [1] [2], or derivatised from a natural source e.g. the antibiotic Gramicidin S (Figure 1.1 D), a cyclic derivative of Gramicidin (*Bacillus brevis*) [14]. Marine organisms are one such natural source which provide a significant numbers of cyclic peptides [15].

1.1.2 Ribosomally Synthesised and Post-translationally Modified Peptides (RiPPS)

Ribosomally synthesised and post-translationally modified peptides (RiPPs) are a class of peptidic natural products which are derived from a ribosomally synthesised

precursor peptide and matured through a range of post translational modifications [16]. RiPPs cover a diverse range of natural products and can be sub-characterised into over twenty groups; exemplars include lanthipeptides [17], bottromycins [18], microcins [19], linear azol(in)e-containing peptide (LAPs) [20], amatoxins and phallotoxins [21], cyanobactins [22] and cyclotides [23] (Figure 1.2). Across the RiPP family, a range of biological activities have been established including antibiotics [18] [24], toxins [21] and anti-HIV activity [25].

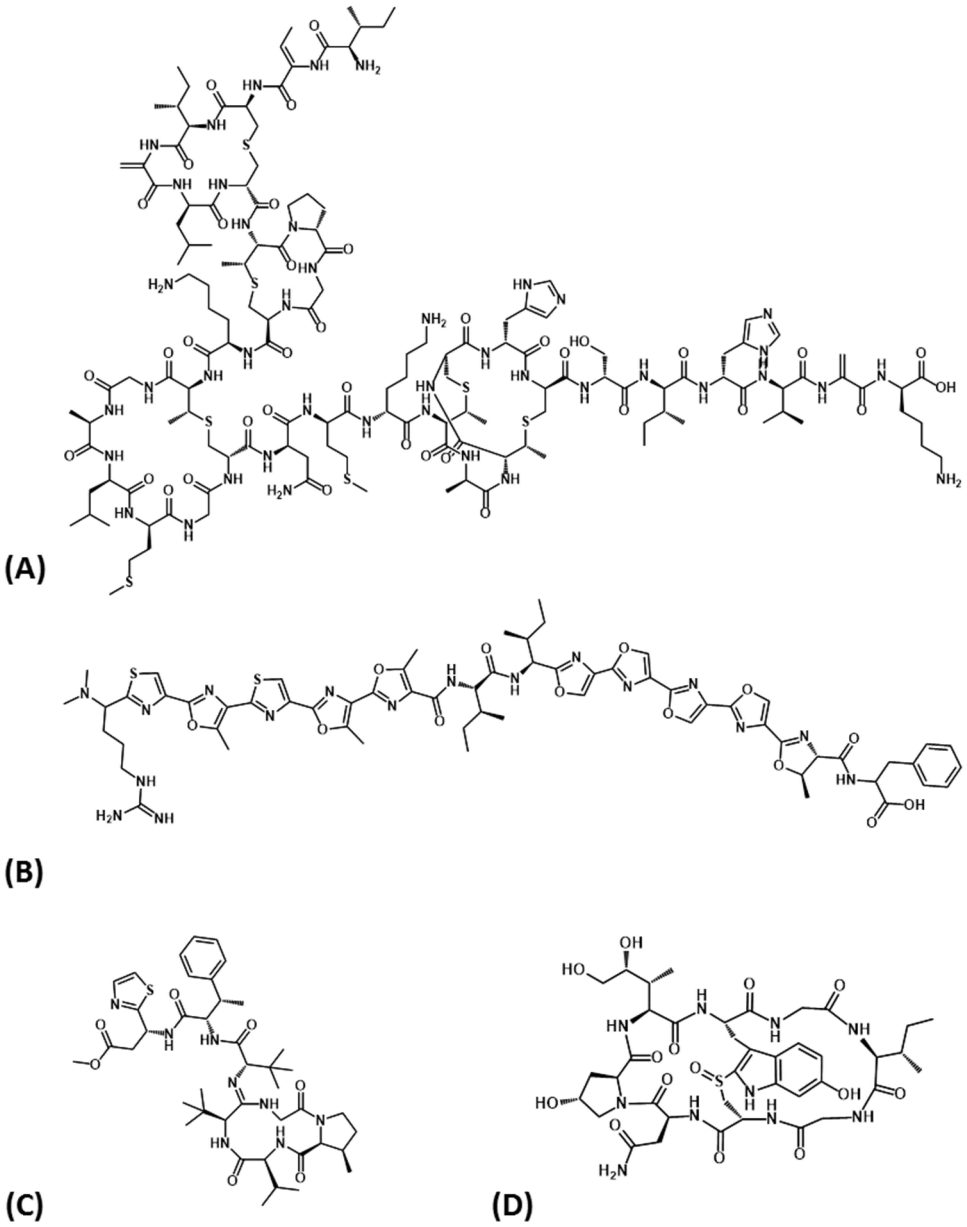


Figure 1.2 Chemical Structures of RiPPs. Chemical structures of (A) The lanthipeptide NisinA [24], (B) The LAP Plantazolicin [26] [27],(C) Bottromycin A2 [28] and (D) the amatoxin α-amanitin [29]).

All RiPPs have a common general biosynthesis with a precursor peptide ribosomally synthesised containing a leader peptide (or in rarer cases a follower peptide) prior to one or multiple core peptides and in some cases is also followed by a C-terminal enzyme recognition sequence (Figure 1.3) [16]. The full length peptide undergoes post translational modifications with enzymes that interact with the leader sequence and then subsequently modify the core peptide. This process therefore uncouples recognition from catalysis, potentially to allow hypervariable core peptide sequences to be modified by the same set of enzymes. Following modification, the core peptide is proteolytically excised to yield the mature RiPP [16]. In several subcategories of RiPPs the peptide is cyclised to form a macrocycle (e.g. cyanobactins [22] and cyclotides [23]).

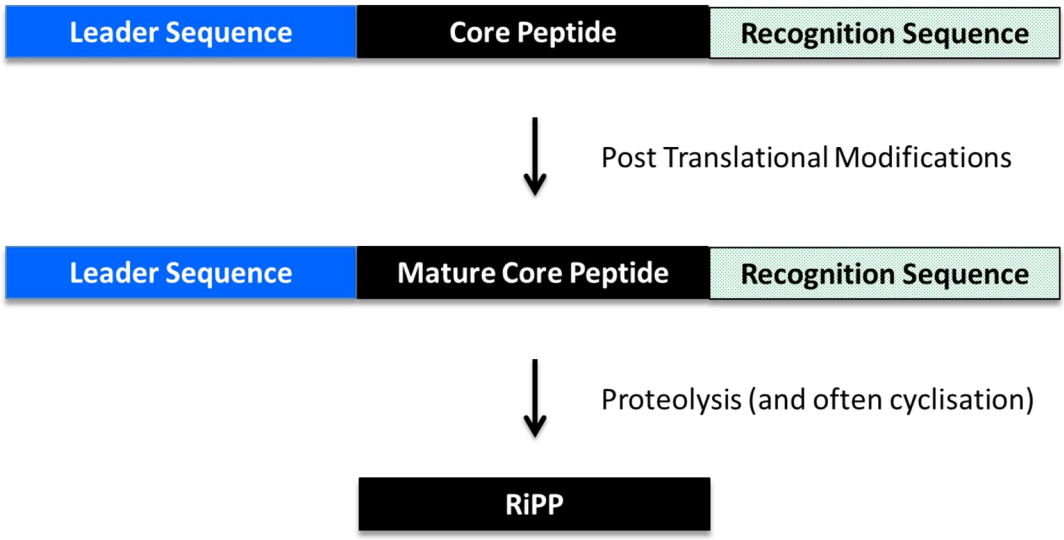


Figure 1.3: Schematic of RiPP Biosynthesis. General schematic of RiPP biosynthesis with a ribosomal precursor peptide tailored by enzymatic post-translational modifications. The enzymes act through recognition of the N-terminal leader and in some cases C-terminal recognition sequences. These sequences are subsequently cleaved leaving the mature compound. Figure adapted from Arnison et al (2013) [16]

One recent example of the RiPPs ability to decouple binding from catalysis is shown in the study of the NisB, a dehydrogenase from the nisin lanthipeptide pathway [30]. NisB catalyses the dehydration of serine and threonine residues to dehydroalanine and

dehydrobutyrine respectively in the precursor peptide NisA, one of the early steps in the biosynthesis of the complex molecule NisinA (Figure 1.2 A) [31]. The dehydration is known to occur via the glutamylation of the serine and threonine side chains followed by a glutamate elimination step [32]. The study determined the crystal structure of the NisB in complex with the precursor peptide NisA showing the distinct peptide binding and glutamylation sites (Figure 1.4) [30]. The distinct binding and catalytic sites is an uncommon feature of enzymes in general but is highly prevalent among the RiPP enzymes [16].

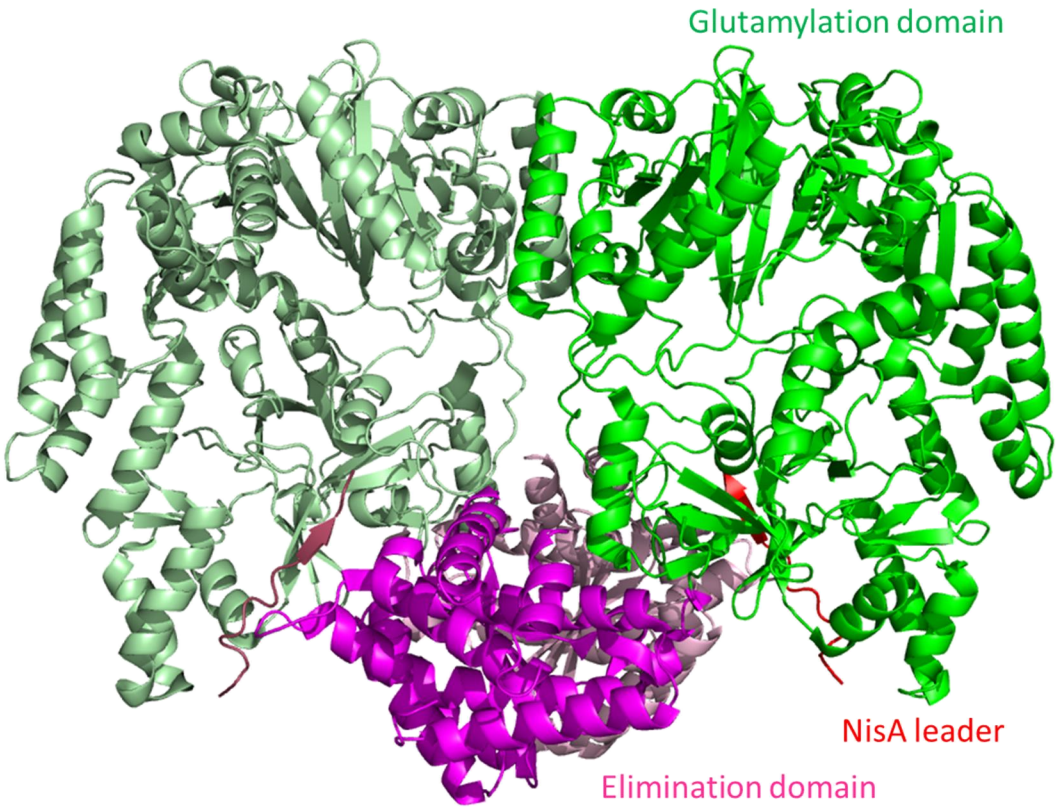


Figure 1.4 Structure of NisB in Complex with Substrate NisA. Dimer structure of NisB in complex with NisA with glutamylation domain represented in green, elimination domain in magenta and the NisA leader sequence in red (secondary molecule shown in paler equivalent colours) [30]. All protein structure figures in this thesis were created in Pymol unless otherwise stated [33].

This thesis is primarily concerned with the cyanobactin family of RiPPs, which will be discussed in detail.

1.1.3 Cyanobactins

The cyanobactins, a member of the RiPPs, are a superfamily of cyclic peptides (macrocycles) produced by both free-living and symbiotic cyanobacteria [22]. The cyanobactins are derived from a precursor peptide which is post-translationally tailored by a range of enzymatic reactions to form cyclic peptides containing modifications such as D-stereo centres, heterocyclised amino acids, disulfide bonds and prenylation. The size of these peptides ranges from six to at least twenty amino acids [22], [34]. The post-translation modifications are carried out by a range of enzymes including proteases, heterocyclases, oxidases and prenyl transferases. Bioinformatic comparisons of the related cyanobactin pathways show a high degree of homology between the related enzymes [35].

Figure 1.5 represents the diverse range of cyanobactins including tenuecyclamide (*N. spongiaeforme)* [36], anacyclamide (*Anabaena* sp.) [34], trichamide (*T. erythraeum)* [37] and trunkamide [38] , ulithiacyclamide [39] and patellamide [40] (all *Prochloron* sp.). The cyanobactins are of particular interest as they have been shown to have diverse biological effects including anticancer (e.g. Ulithiacyclamide A [41]' [42]) and trunkamide A [43]), anti-parasitics (e.g. Venturamide A [44]) as well as reversal of multi-drug resistance (e.g. Dendroamide A [45]). In addition, the cyanobactins are produced using a defined set of tailoring enzymes and so are attractive for bioengineering [46].

The high degree of homology between enzymes involved in cyanobactin production across the specific pathways potentially allows for enzyme exchange in order to lead to more diverse novel products. Additionally, modification of the gene encoding the precursor peptide in the core peptide region could lead to novel sequences and thus novel cyclic compounds.

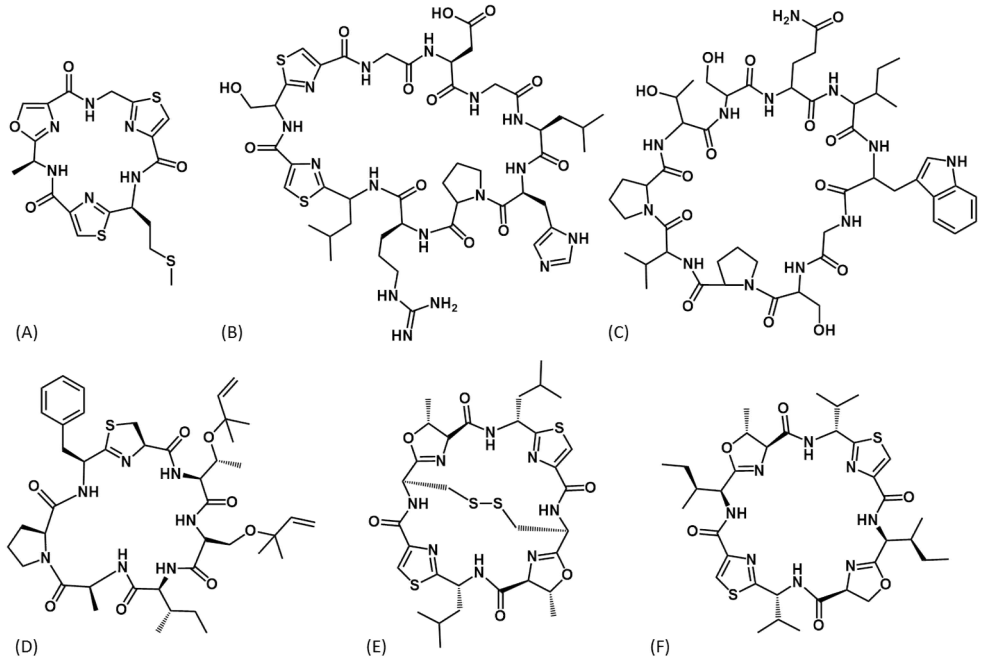


Figure 1.5: Chemical Structures of Cyanobactins. The chemical structures of cyanobactins from the following families (A) Tenuecyclamide, (B) Anacyclamide, (C) Trichamide, (D) Trunkamide, (E) Ulithiacyclamide and (F) Patellamide. (Adapted from Sivonen *et al*. [22])

The patellamides are one of the most studied cyanobactins and therefore represents an excellent model system for exploitation and bioengineering.

1.2 Patellamide Biosynthesis

The patellamides are a cyanobactin family produced by *Prochloron* sp., cyanobacteria that exist symbiotically with the marine sea squirt *Lissoclinum patella* [47]. *L. patella*, along with its obligate symbionts, is commonly found in tropical oceans, such as near the Palau Islands or at the Great Barrier Reef, both in the Pacific Ocean [40] [48].

The patellamides are macrocyclised octapeptides containing heterocyclised residues (Thr/Ser and Cys) giving rise to oxazolines and thiazolines, the latter of which can be further oxidised to thiazoles [49] (Figure 1.6). Additionally, D-stereocentres are found adjacent to the thiazoles.

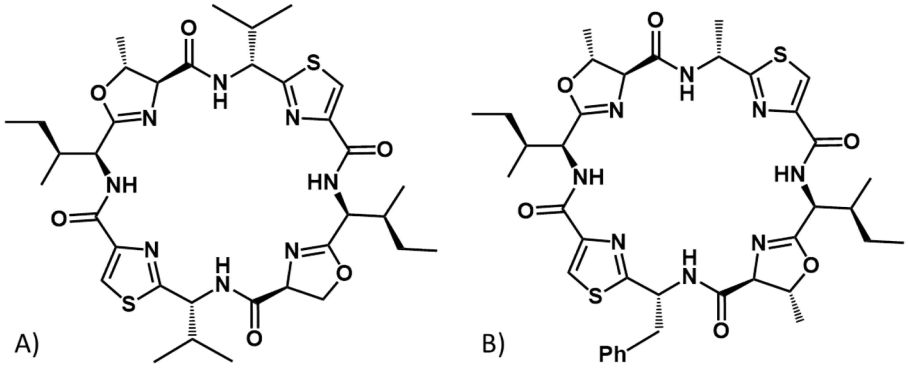


Figure 1.6: Chemical Structure of Patellamides. The chemical structures of (A) Patellamide A and (B) Patellamide D.

The first patellamides, patellamides A-C, were isolated in 1981 by Ireland *et al.* and originally believed to be produced directly from the sea squirt, hence the naming convention [40]. It has since been discovered that *Prochloron* sp. are the true source of production [47]. Since the first discovery, several new patellamides have been discovered (patellamides D-G) [42] [50] [51] [52] [53].

The patellamides show considerable chemical diversity and as such their activity is dependent on their cyclic nature, the heterocycles present, and the amino acids side chains which make up the macrocycle. Variation in any of these properties, even subtle changes, can have large consequences on their function [49]. Lissoclinamide 4, a related cyanobactin, shows an increase of two orders of magnitude in cytotoxicity against T24 bladder carcinoma cells compared to lissoclinamide 5, despite varying only by the presence of a thiazoline as opposed to a thiazole, its oxidised analogue. [42] [54].

Although the patellamides function for the organism is unknown, they could potentially be produced as a defence mechanism to protect its environment from other bacterial species. The patellamides have been tested for a range of biological activities and been found to be cytotoxic to leukaemia cells and also having the ability to reverse multiple drug resistance [40] [55]. To date the basis of patellamide mode of action is not well defined but there is data to suggest that proteins associated with DNA, RNA and protein synthesis are all targets [49]. Patellamide D is however known to target the ATP-binding cassette multidrug efflux pump P-glycoprotein (P-gp) [55] which regulates the distribution and bioavailability of drugs [56]. Inhibition of this protein could play a role in cancer therapies where chemotherapeutic resistance has occurred. Previous structural studies on the mouse P-gp protein have identified the binding sites for cyclic peptide inhibitors containing heterocycles and it is likely that Patellamide D interacts in a similar manner [57] [58]. The inhibitors are bound in the transmembrane portion of the protein (Figure 1.7).

Patellamides are produced *in vivo* through a biosynthetic pathway (discussed in detail in following pages), however it is possible to synthesise them chemically. Garcia-Reynaga and VanNieuwenhz (2008) [59] carried out the total synthesis of Patellamide A, an overall 18 step process (including the utilisation of the work of You *et al.* (2003) [60] to achieve the starting material) with an overall yield of approximately 25 % (Figure 1.8). This multiple step process gives a reasonably low yield, particularly when it can be considered that Patellamide A is one of the more symmetrical compounds

(mirror image with exception of serine/threonine) which allows for some of the initial starting materials to be used for both the initial Ser and Thr containing 4-mers. As a result of this, it is likely that when more diversity is required in the patellamide, the synthesis would require additional steps and reduce the efficiency of the yield further. As such, the chemical synthesis is a plausible yet lengthy strategy for producing patellamides and patellamide-derived novel compounds.

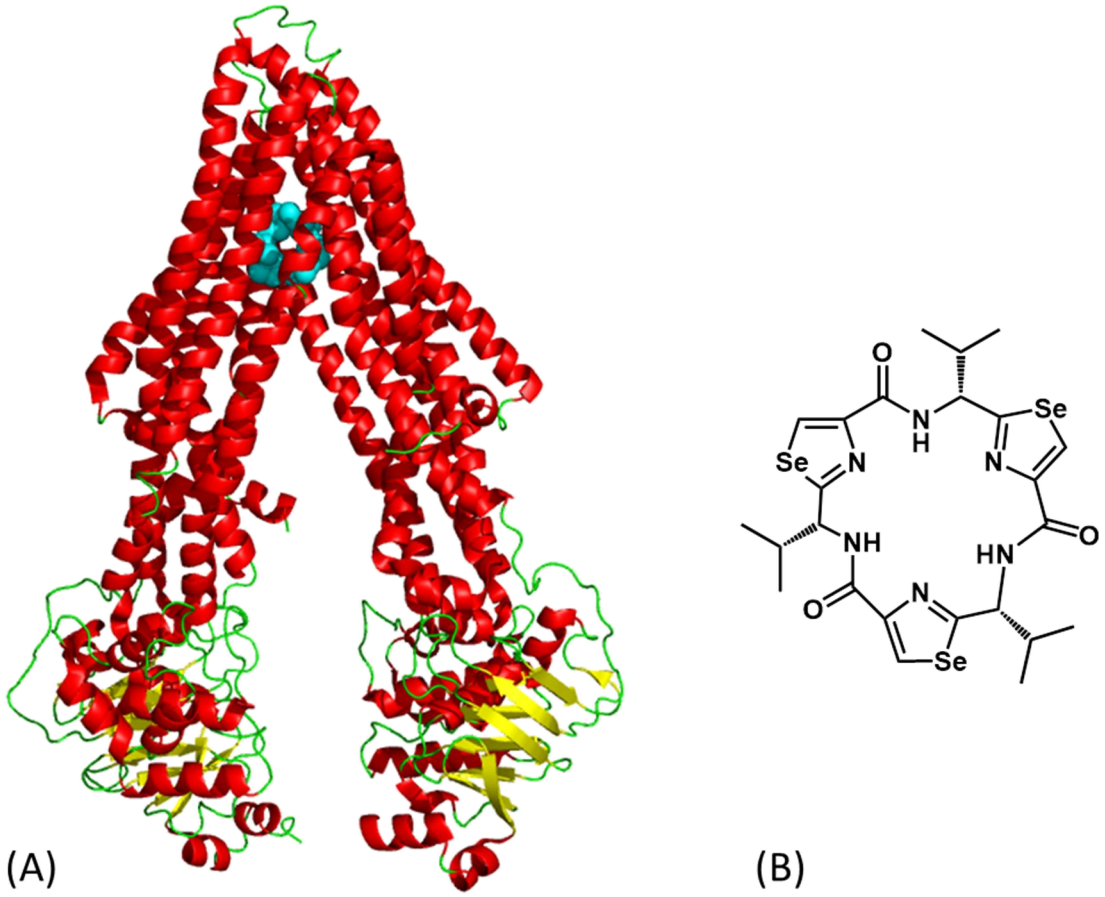


Figure 1.7: P-glycoprotein in Complex with Cyclic Peptide. (A) Crystal structure of mouse P-glycoprotein (alpha helices - red, beta sheet - yellow) in complex with the hexamer cyclic peptide QZ59-SSS containing heterocycles (spheres, cyan) PDB Code: 4M2S (B) Chemical structure of QZ59-SSS.

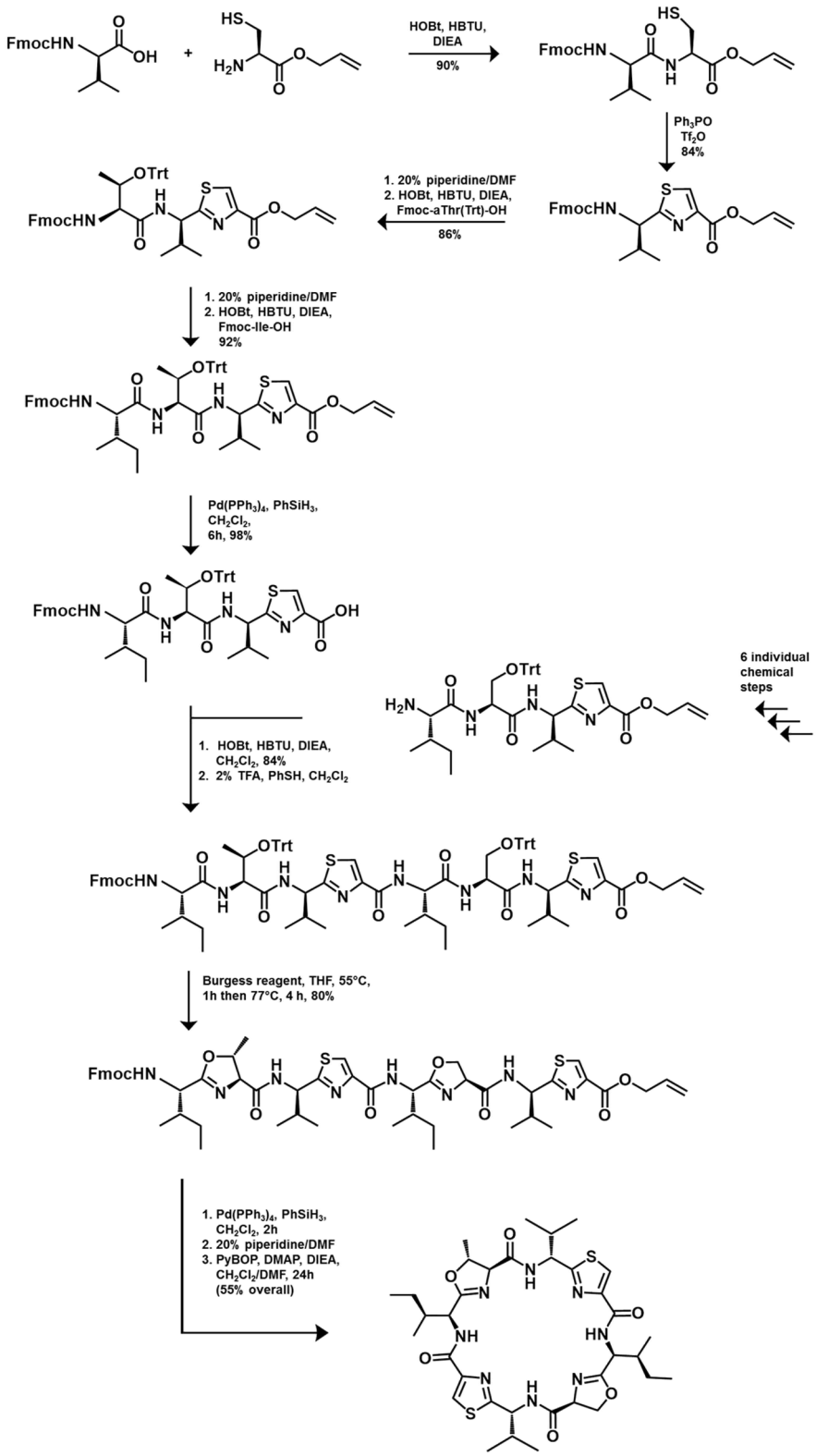


Figure 1.8: Chemical Synthesis of Patellamides. 18 step total chemical synthesis of Patellamide A as described by Garcia-Reynaga & VanNieuwenhz (2008) [59] including the initial steps of You *et al.* (2003) [60]. The overall reaction has a yield of ~25 %.

Patellamides are biosynthesised ribosomally as a precursor peptide which undergoes several post-translational modifications through a range of enzymatic (and potentially non-enzymatic) reactions. A single gene cluster (Figure 1.9) of *Prochloron sp*. has been identified which encodes both the precursor peptide (PatE) and the tailoring enzymes (PatA, B, C, D, F, G) [47]. Studies of the gene cluster when heterologously expressed in *E. coli* have shown that PatA, D, E, F and G are essential for patellamide production and their absence leads to no detectable products. The absence of PatB and PatC however still results in patellamide products [47] [61] [62].

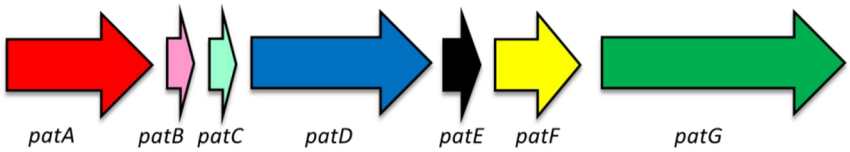


Figure 1.9: Schematic of Patellamide Gene Cluster. The gene cluster responsible for encoding the seven proteins of the patellamide biosynthetic pathway with seven genes *patA-G*.

PatE, the 71 amino acid precursor peptide, consists of a 37 amino acid leader sequence followed by two eight residue core peptides each flanked by N- and C-terminal cleavage/macrocyclisation recognition signals of five and three amino acids respectively (Figure 1.10). The composition of the core peptide can be hypervariable and it is this core peptide that goes on to constitute the patellamide [47]. In some cyanobactin families, the precursor peptide contains up to four core peptides (e.g. Tenuecyclamides [62]).

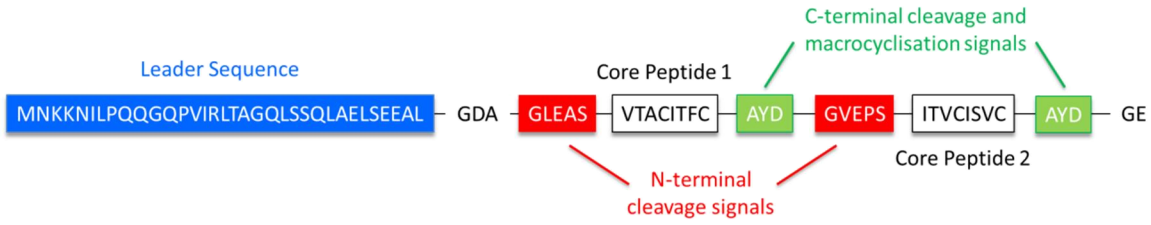


Figure 1.10: Sequence Schematic of the PatE Precursor Peptide. The PatE precursor peptide highlighting the leader sequence (blue), the N-terminal cleavage recognition signals (red), the variable core peptides (white) and the C-terminal cleavage signals (green).

PatA (77 kDa) consists of an N-terminal subtilisin-like protease domain and a C-terminal domain of unknown function (DUF). The protease domain is responsible for N-terminal cleavage of the core peptides [63]. PatB (9 kDa) and PatC (7 kDa) currently have no defined function [61]. PatD (89 kDa) is a three domain protein responsible for heterocyclisation of specific amino acids within the core peptides [64]. PatF (36 kDa) currently has no defined function but equivalent enzymes in related pathways have shown that the PatF family is responsible for the prenylation of specific amino acids within the core peptide [65]. Finally, PatG (131 kDa) consists of three domains, an N-terminal oxidoreductase domain, a subtilisin-like protease domain and a C-terminal DUF [63] [66]. The oxidoreductase domain is likely to catalyse the oxidation of thiazolines to thiazoles, however this has yet to be confirmed, while the protease domain carries out C-terminal cleavage and macrocyclisation of each core peptide [63]. In addition, the epimerisation of two C_α positions adjacent to thiazolines also occurs and may be an enzymatic or a non-enzymatic process.

The order in which these reactions occur is not yet fully defined, although it is clear that certain reactions probably occur before others, as will be discussed.

1.2.1 Heterocyclisation

Note: A selection of the work in this section has been carried out by the Naismith group including myself. This was achieved during the PhD research period but not within the scope of this thesis. All work has been published and references for this work are noted with an asterisk.

PatD is an ATP-dependent heterocyclase enzyme which catalyses the formation of oxazolines and thiazolines from threonine/serine and cysteine residues respectively (Figure 1.11, 1.12). Each heterocyclisation event results in the loss of one water molecule which can be observed by a reduction in mass of 18 Da [64]. The binding of the PatE precursor peptide to the PatD is mediated through the 37 amino acid leader sequence of the PatE and removal of this sequence abolishes the majority of the heterocyclase activity [64]. The C-terminal cysteine of the core peptide will still process without the leader sequence as long as the C-terminal recognition sequence remains present [67]* [68].

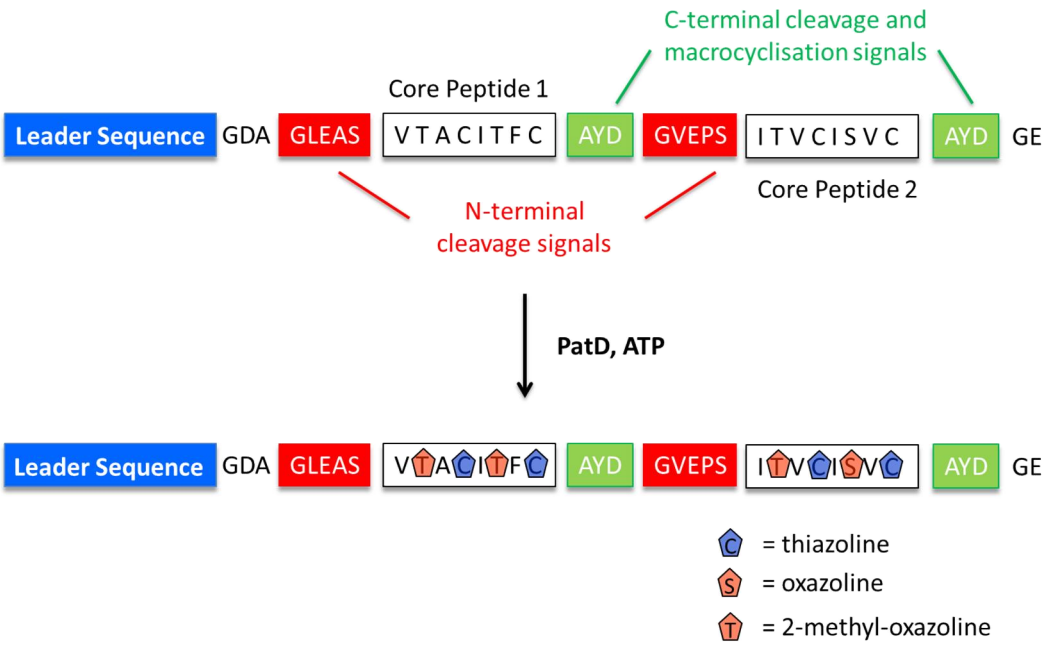


Figure 1.11: PatD Catalysed Heterocyclisation of PatE. Schematic of the heterocyclisation of the PatE precursor peptide by the heterocyclase enzymes PatD highlighting cysteine (C), serine (S) and threonine (T) heterocyclisation within the core peptides to thiazoline, oxazoline and 2-methyl-oxazoline respectively.

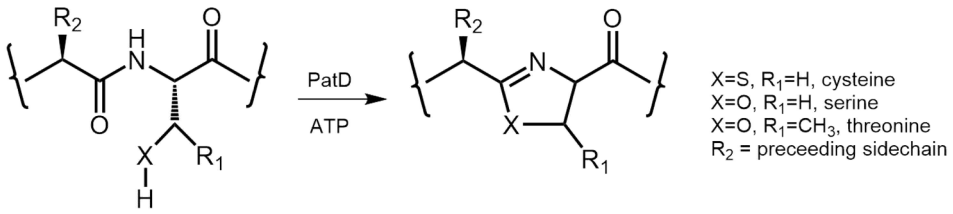


Figure 1.12: PatD Reaction at a Molecular Level. Chemical structure schematic of the substrate amino acids cysteine, serine and threonine and their heterocyclic products upon reaction with PatD and ATP.

Within the cyanobactin family there are variations in the type of activity each heterocyclase adopts. TruD, an analogue from the trunkamide pathway, and LynD, an analogue from the aestuaramide pathway, are reported to only heterocyclise cysteine residues and both leave threonine and serine residues in their native state [64] [69]*. This is intriguing considering the high homology of TruD and LynD to PatD; 88 % and 76 % identity respectively (Figure 1.13). The variation between the three enzymes may be a result of chemical selectivity, however one study on PatD and TruD has suggested that this is primarily down to regiospecificity [64]. Two positions within the core peptide (positions 1 and 5 from C-terminus) were found to be more liable to heterocyclisation by TruD, while PatD could act on at least five positions (positions 1, 3, 4, 5 and 7). The order in which the heterocyclisations within the precursor peptide occur has been determined using the heterocyclase TruD and was found to process from the C-terminus backwards [67]*.

The crystal structure of TruD (PDB entry: 4BS9) has been determined and the enzyme was found to consist of three domains; domain 1 (residues 2-85), domain 2 (residues 86-321) and the larger domain 3 (residues 323-781) (Figure 1.14 A) [67]*. Two molecules of the protein interact to form a biologically relevant anti-parallel dimer (Figure 1.14 B) [67]*. Domains 1 and 2 were found to show structural homology to MccB, an adenylase from a microcin pathway [70], and both enzymes were found to contain two CXXC motifs along with bound structural zinc ion. These features are common amongst one class of adenylases [71]. The ATP binding site of MccB was not conserved in TruD meaning it was not possible to locate the enzyme active site. It was also not possible to identify where the leader sequence may bind in the structure.

```
LynD   MQSTPLLQIQPHFHVEVIEPKQVYLLGEQANHALTGQLYCQI   42
PatD   .MQPTALQIKPHFHVEIIEPKQVYLLGEQGNHALTGQLYCQI   41
TruD   .MQPTALQIKPHFHVEIIEPKQVYLLGEQGNHALTGQLYCQI   41

LynD   LPLLNGQYTLEQIVEKLDGEVPPEYIDYVLERLAEKGYLTEA   84
PatD   LPFLNGEYTREQIVEKLDGQVPEEYIDFVLSRLVEKGYLTEV   83
TruD   LPFLNGEYTREQIVEKLDGQVPEEYIDFVLSRLVEKGYLTEV   83

LynD   APELSSEVAAFWSELGIAPPVAAEALRQPVTLTPVGN.ISEV   125
PatD   APELSLEVAAFWSELGIAPSVVAEGLKQPVTVTTAGKGIREG   125
TruD   APELSLEVAAFWSELGIAPSVVAEGLKQPVTVTTAGKGIREG   125

LynD   TVAALTTALRDIGISVQTPTEA......GSPTALNVVLTDDY   161
PatD   IVANLAAALEEAGIQVSDPKAPKAPKAGDSTAQLQVVLTDDY   167
TruD   IVANLAAALEEAGIQVSDPRDPKAPKAGDSTAQLQVVLTDDY   167

LynD   LQPELAKINKQALESQQTWLLVKPVGSVLWLGPVFVPGKTGC   203
PatD   LQPELAAINKEALERQQPWLLVKPVGSILWLGPLFVPGETGC   209
TruD   LQPELAAINKEALERQQPWLLVKPVGSILWLGPLFVPGETGC   209

LynD   WDCLAHRLRGNREVEASVLRQKQAQQQRNGQS.GSVIGCLPT   244
PatD   WHCLAQRLRGNREVEASVLQQKRALQERNGQNKNGAVSCLPT   251
TruD   WHCLAQRLQGNREVEASVLQQKRALQERNGQNKNGAVSCLPT   251

LynD   ARATLPSTLQTGLQFAATEIAKWIVKYHVNATAPGTVFFPTL   286
PatD   ARATLPSTLQTGLQWAATEIAKWMVKRHLNAIAPGTARFPTL   293
TruD   ARATLPSTLQTGLQWAATEIAKWMVKRHLNAIAPGTARFPTL   293

LynD   DGKIITLNHSILDLKSHILIKRSQCPTCGDPKILQHRGFEPL   328
PatD   AGKIFTFNQTTLELKAHPLSRRPQCPTCGDQEILQRRGFEPL   335
TruD   AGKIFTFNQTTLELKAHPLSRRPQCPTCGDRETLQRRGFEPL   335

LynD   KLESRPKQFTSDGGHRGTTPEQTVQKYQHLISPVTGVVTELV   370
PatD   KLESRPKHFTSDGGHRATTPEQTVQKYQHLIGPITGVVTELV   377
TruD   KLESRPKHFTSDGGHRAMTPEQTVQKYQHLIGPITGVVTELV   377

LynD   RITDPANPLVHTYRAGHSFGSA.TSLRGLRNTLKHKSSGKGK   411
PatD   RISDPANPLVHTYRAGHSFGSSAGSLRGLRNTLRYKSSGKGK   419
TruD   RISDPANPLVHTYRAGHSFGSA.TSLRGLRNVLRHKSSGKGK   418

LynD   TDSQSKASGLCEAVERYSGIFQGDEPRKRATLAELGDLAIHP   453
PatD   TDSQSRASGLCEAIERYSGIFLGDEPRKRATLAELGDLAIHP   461
TruD   TDSQSRASGLCEAIERYSGIFQGDEPRKRATLAELGDLAIHP   460

LynD   EQCLCFSDGQYANRETLNEQ.ATVAHDWIPQRFDASQAIEWT   494
PatD   EQCLHFSDRQYDNRDALNAEGSAAAYRWIPHRFAASQAIDWT   503
TruD   EQCLHFSDRQYDNRESSNER.ATVTHDWIPQRFDASKAHDWT   501

LynD   PVWSLTEQTHKYLPTALCYYHYPLPPEHRFARGDSNGNAAGN   536
PatD   PLWSLTEQKHKYVPTAICYYNYLLPPADRFCKADSNGNAAGN   545
TruD   PVWSLTEQTHKYLPTALCYYRYPFPPEHRFCRSDSNGNAAGN   543

LynD   TLEEAILQGFMELVERDGVALWWYNRLRRPAVDLGSFNEPYF   578
PatD   SLEEAILQGFMELVERDSVALWWYNRLRRPEVELSSFEEPYF   587
TruD   TLEEAILQGFMELVERDSVCLWWYNRVSRPAVDLSSFDEPYF   585

LynD   VQLQQFYRENDRDLWVLDLTADLGIPAFAGVSNRKTGSSERL   620
PatD   LQLQQFYRSQNRELWVLDLTADLGIPAFAGLSRRTVGSSERV   629
TruD   LQLQQFYQTQNRDLWVLDLTADLGIPAFVGVSNRKAGSSERI   627

LynD   ILGFGAHLDPTIAILRAVTEVNQIGLELDKVPDENLKSDATD   662
PatD   SIGFGAHLDPKIAILRALTEVSQVGLELDKVPDEKLDGESKD   671
TruD   ILGFGAHLDPTVAILRALTEVNQIGLELDKVSDESLKNDATD   669

LynD   WLITEKLADHPYLLPDTTQPLKTAQDYPKRWSDDIYTDVMTC   704
PatD   WMLEVTLETHPCLAPDPSQPRKTANDYPKRWSDDIYTDVMAC   713
TruD   WLVNATLAASPYLVADASQPLKTAKDYPRRWSDDIYTDVMTC   711

LynD   VNIAQQAGLETLVIDQTRPDIGLNVVKVTVPGMRHFWSRFGE   746
PatD   VEMAKVAGLETLVLDQTRPDIGLNVVKVMIPGMRTFWSRYGP   755
TruD   VEIAKQAGLETLVLDQTRPDIGLNVVKVIVPGMR.FWSRFGS   752

LynD   GRLYDVPVKLGWLDEPLTEAQMNPTPMPF   775
PatD   GRLYDVPVQLGWLKEPLAEAEMNPTNIPF   784
TruD   GRLYDVPVKLGWREQPLAEAQMNPTPMPF   781
```

Figure 1.13: PatD, TruD and LynD Sequence Alignment. Sequence alignment of the heterocyclases PatD and TruD from *Prochloron sp.* and LynD from *Lyngbya aestuarii* showing a high degree of homology. All sequence alignments were generated using ClustalW [72] or Clustal Omega [73] and the figures created with ALINE [74].

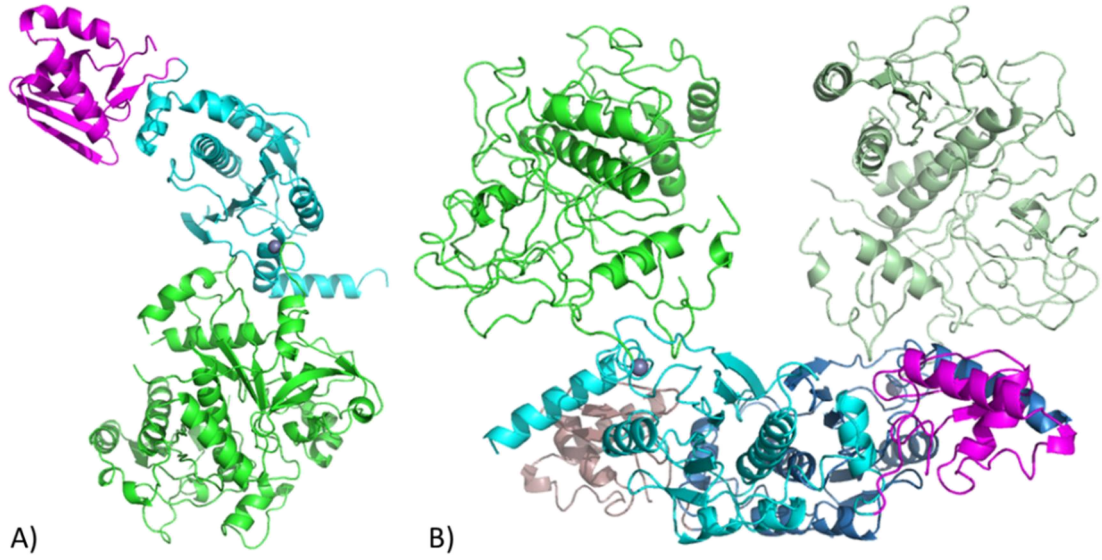


Figure 1.14: **Crystal Structure of TruD Heterocyclase Enzyme.** TruD crystal structure to 2.95 Å resolution (PDB entry: 4BS9). (A) The asymmetric unit of domains one, two and three coloured magenta, cyan and green respectively. The zinc ion is shown as a grey sphere (B) The dimer representation of TruD which is biologically relevant showing the anti-parallel dimer interface occurring between domains one and two (symmetry molecule in darker equivalent colours) [67]*.

The PatE leader sequence had been proposed to contain a helical structure from residues 13 - 26 [75] which would resemble that of the precursor peptide for microcin B17, where the helix plays a crucial role in peptide:protein binding [76], [77]. These original studies however were carried out in organic solvent and a more recent study has determined using NMR analysis of a ^{15}N labelled PatE in aqueous buffer that there is no degree of secondary structure in the leader sequence [67]*.

The recent determination of the crystal structure of LynD in complex with PatE and ATP (PDB entry: 4V1T) has allowed for the location of the peptide binding clamp and the ATP binding site to be determined [78]* (Figure 1.15). LynD is structurally very similar to TruD and forms the same dimer interface, however under binding of PatE the protein undergoes a conformational change. The PatE leader sequence binds and extends a beta sheet of domain 1 of LynD which in turn interacts with domain 3 of the

secondary molecule essentially bringing the two domains together and confirming the biological relevance of the dimer. The conformational change which occurs upon leader sequence binding results in the activation of the heterocyclase, subsequently allowing complete processing of the core peptide.

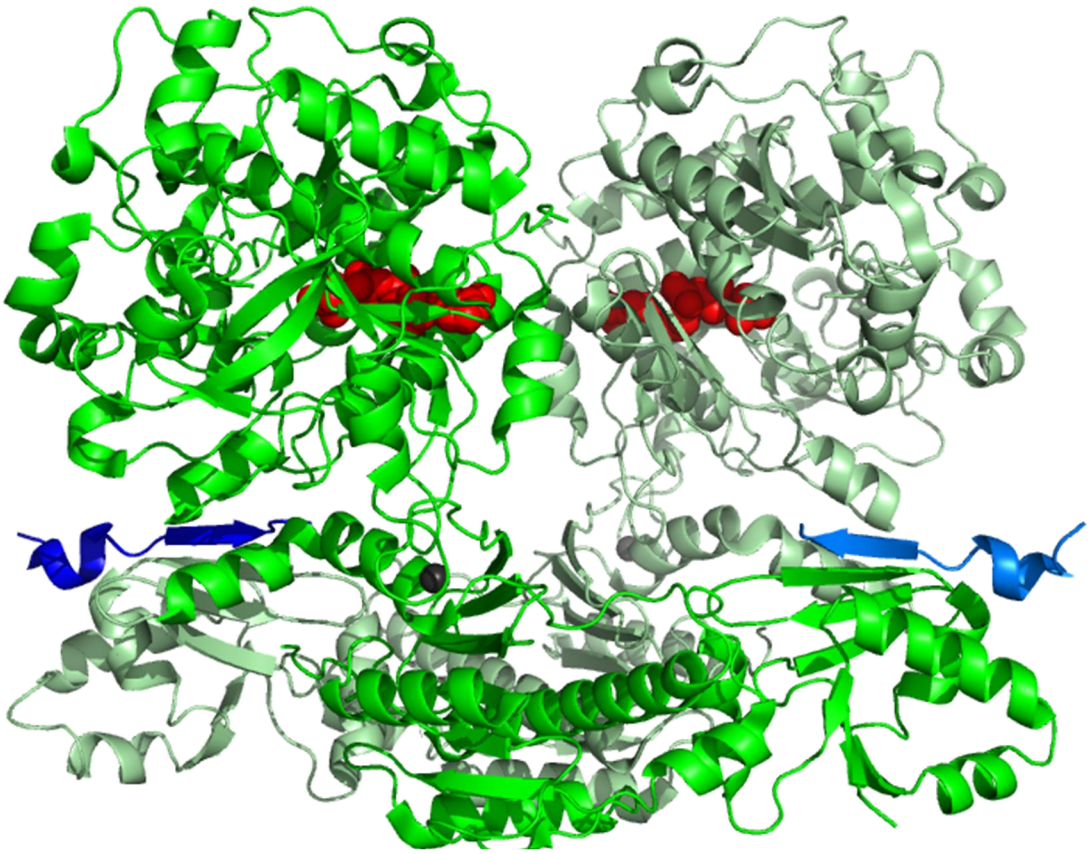


Figure 1.15: **Crystal Structure of LynD Heterocyclase Enzyme in Complex with PatE and ATP.** LynD crystal structure to 2.14 Å resolution (PDB entry: 4V1T) in complex with PatE and ATP (broken down to ADP and PO_4^-). The dimer representation of LynD (green/pale green) with two copies of the PatE leader sequence (blue/pale blue), two molecules of ATP (red spheres) bound and two zinc ions (grey spheres). Domain 1 of the first molecule is brought together with domain 3 of the second molecule and vice versa [78]*.

The catalytic reaction of PatD is known to be ATP-dependent although the mechanism of this reaction remains unclear. McIntosh *et al*. (2010) initially proposed that heterocyclisation is driven by ATP through a molecular machine basis where there is no direct interaction between the heterocyclisation chemistry and ATP breakdown [79] (Figure 1.16).

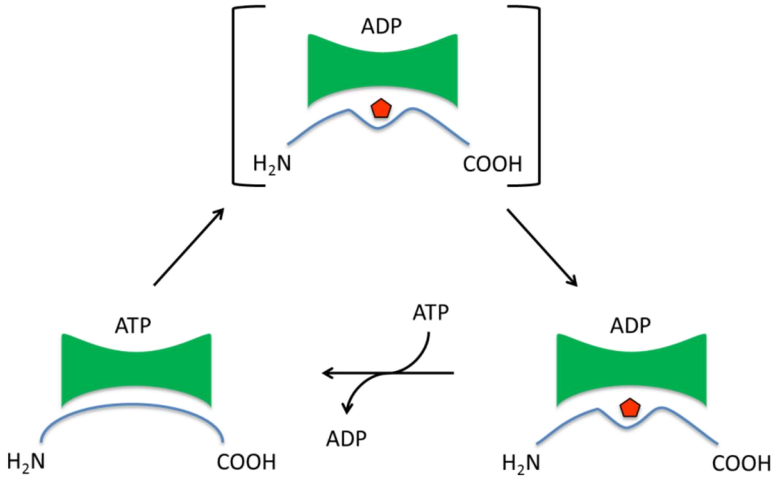


Figure 1.16: ATP Molecular Machine Mechanism for PatD: Proposed mechanism for PatD ATP depletion with ATP binding at a distinct site from the PatE causing turnover. Figure adapted from McIntosh *et al* (2010) [79].

Studies on the related RiPP biosynthetic pathway of thiazole/oxazole-modified microcins (TOMMs) then suggested that heterocyclisation occurs through a kinase mechanism where the peptide backbone carbonyl oxygen is phosphorylated by ATP [80] [81]. This then drives an O-elimination reaction which generates the final oxazoline/thiazoline (Figure 1.17). This releases H_2O, ADP and phosphate, the latter of which was directly observed in experimental analysis. Unlike the patellamide pathway where heterocyclisation is carried out by a single enzyme, PatD, in the TOMMS pathway heterocyclisation is carried out by three enzymes working together, BalhC (scaffold protein), BalhD (catalytic protein) and BalhF (peptide binding protein) [80] [81] [82] [83]. There is high sequence homology between the third domain of PatD and BalhD suggesting that they will perform the same reaction therefore it would be likely that PatD also process via a kinase mechanism.

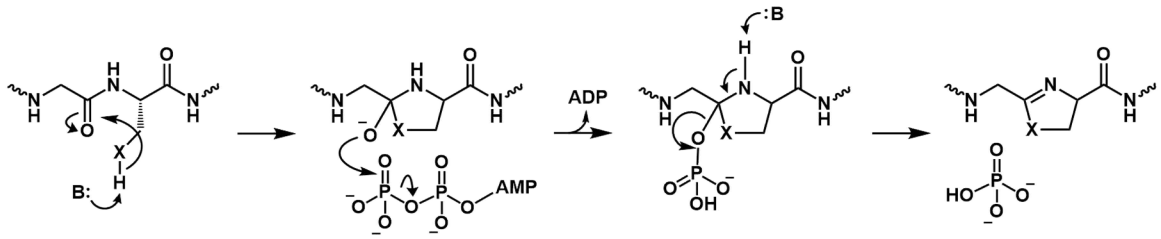


Figure 1.17 – Kinase Mechanism of Heterocyclisation. Proposed mechanism for heterocycle formation in the TOMMS pathway by BalhC/BalhD/BalhF with ATP phosphorylating the carbonyl oxygen followed by an O-elimination with the release of ADP, PO_4^- and H_2O. X = O or S. (Figured adapted from Dunbar *et al*., 2012, [80])

Next, a study on the TruD protein showed by an NMR approach that heterocyclisation potentially occurs via an adenylation mechanism through production of AMP and pyrophosphate from ATP during turnover and failed to observe ADP [67]* (Figure 1.18). In addition, the crystal structure identified an adenylase fold in the second domain correlating with the NMR data (Figure 1.14).

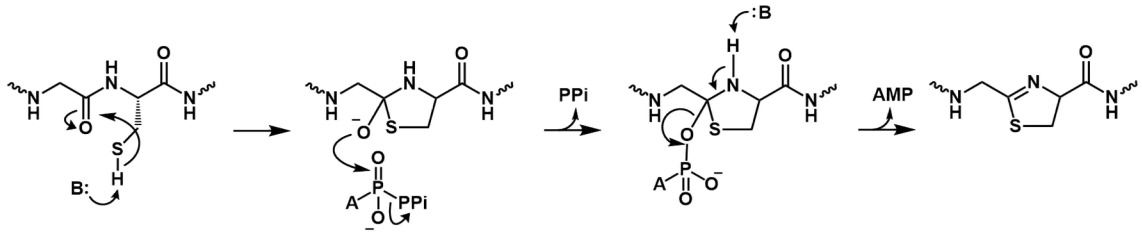


Figure 1.18: Adenylation Mechanism of Heterocyclisation. Proposed mechanism for thiazoline formation by TruD with ATP adenylating the carbonyl oxygen followed by an O-elimination with the release of AMP, PPi and H_2O. (Figure adapted from Koehnke *et al*, 2013, [67]*)

The adenylation mechanism appeared to be further backed up by a study on the *E.coli* YcaO domain, a close homologue of BalhD, which was, in the absence of substrate, shown to also hydrolyse ATP to AMP and PP_i [82]

Finally, like TruD, NMR studies on the LynD protein observed the production of AMP and pyrophosphate during heterocyclisation reactions with ATP [78]*. However, the crystal structure of LynD, PatE and ATP in complex cast doubt on to the proposed

adenylation mechanism [78]*. The binding of ATP in the active site (where in the structure ATP has actually hydrolysed to ADP and PO_4^-) found that the nucleotide and the alpha phosphate were buried in a pocket whilst the gamma phosphate was exposed (Figure 1.19) [78]*. This would therefore favour a kinase mechanism through nucleophilic attack of the beta-gamma bond. Thus the two sets of data are conflicting.

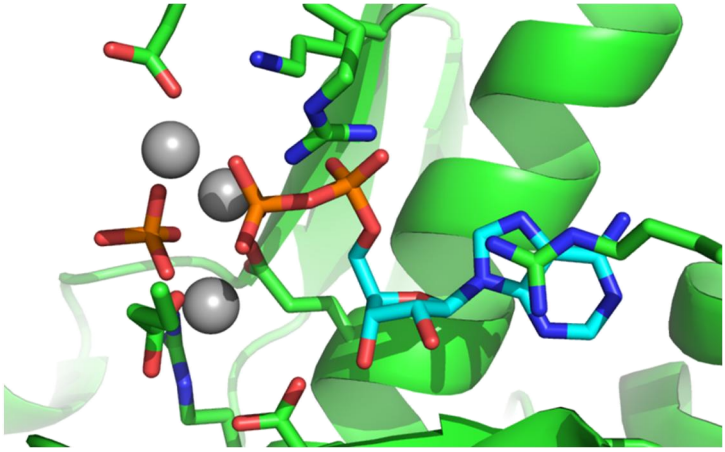


Figure 1.19: Nucleotide Binding to LynD. Structural binding of ADP and PO_4^- to LynD showing structural shielding of the adenosine ring and the alpha phosphate by side chains of the enzyme whilst the beta and gamma phosphates are exposed. The phosphates are co-ordinated by three Mg^{2+} ions (grey spheres). (Figure adapted from Koehnke et al, 2015, [78]*)

More studies will have to be carried out to fully elucidate the mechanism of heterocyclisation and the fate of ATP in this mechanism; however given current data, neither a kinase nor an adenylation mechanism appear to be entirely correct.

1.2.2 N-terminal Core Peptide Cleavage

The N-terminal subtilisin-like protease domain of PatA (PatApr) catalyses the N-terminal cleavage of the modified core peptide residue, resulting in the removal of the PatE precursor peptide leader sequence (Figure 1.20). This step must proceed after heterocyclisation due to the requirement of the leader sequence to bind PatD to enable heterocyclase activity [79].

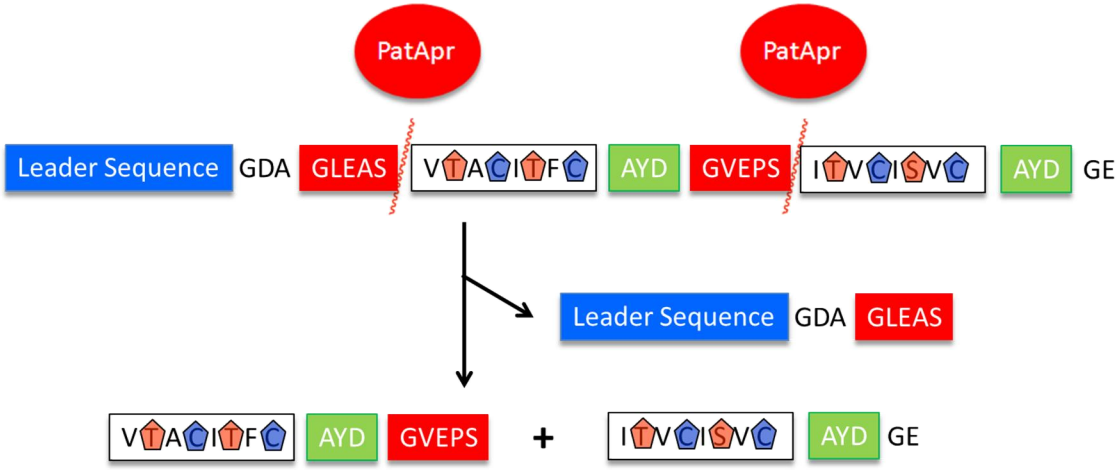


Figure 1.20: PatA Protease Activity. Schematic for the action of PatApr which cleaves at the N-terminal of the modified core peptide (white box) upon recognition of the sequence 'G (L/V) E (A/P) S' (red box).

PatApr contains the classic catalytic triad of subtilisin-like serine proteases [84], [85] with residues Asp23, His58 and Ser218 making up the active site. The cleavage site nomenclature uses P1-5 for residues N-terminal to cleavage site and P1'-P5' C-terminal. The N-terminal side of each core peptides contains the PatApr recognition sequence, consisting of a highly conserved five amino acids 'P5 (G) P4 (L/V) P3 (E) P2 (A/P) P1 (S)'. There does not appear to be any conservation at the P1' or subsequent positions for protease activity, i.e. no restrictions on the first residue in the core peptide [63]. Like all subtilisin proteases [86], cleavage occurs through nucleophilic attack of the peptide bond after the recognition site by the catalytic serine (S218) following proton extraction by the catalytic histidine (H58) forming a tetrahedral

intermediate. The tetrahedral intermediate collapses, breaking the peptide bond and the C-terminal peptide is then released and protonated by His58 thus leaving an acyl-enzyme intermediate. Finally, upon water addition a second tetrahedral intermediate is formed and its subsequent collapse releases the N-terminal peptide and returns the enzyme to its native state, ready to catalyse the next reaction (Figure 1.21).

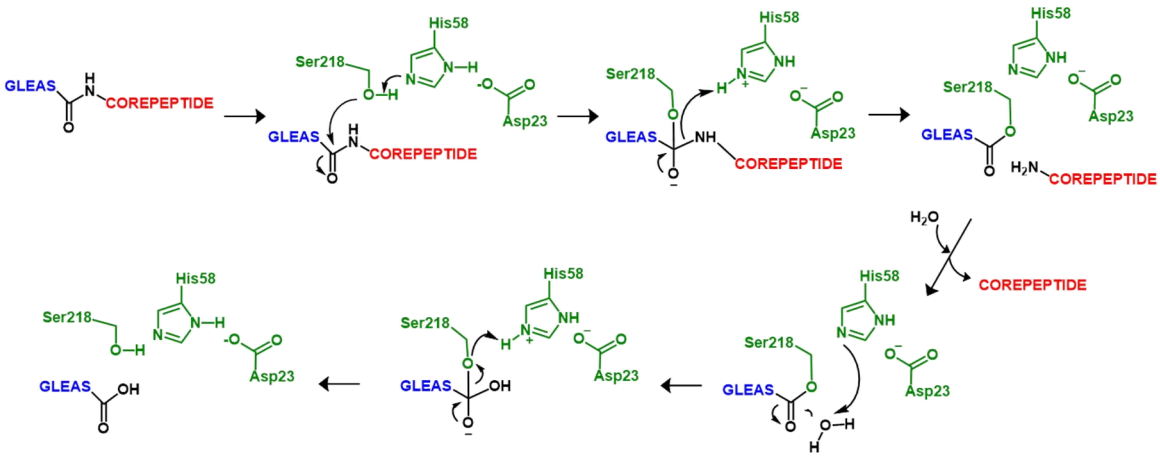


Figure 1.21: PatA Protease Mechanism. Schematic for the mechanism of action of PatApr showing classic serine protease characteristics.

The crystal structure of PatApr has been determined and is consistent with known subtilisin like protease domains (Figure 1.22, 1.23) [66] [87]

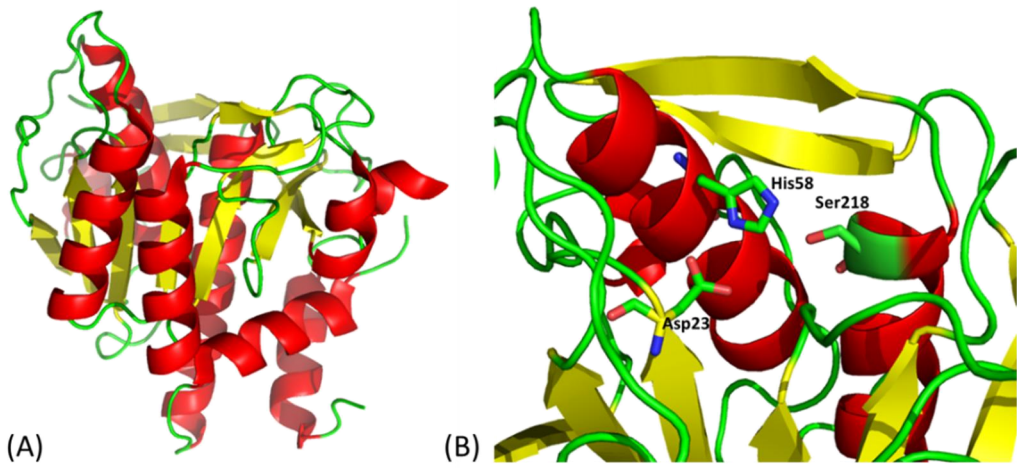


Figure 1.22: PatA Protease Domain Structure. A) Crystal structure of the PatA protease domain coloured by secondary structure elements (alpha helices in red, beta sheets in yellow) PDB Code: 3ZXX. B) Zoom view of the active site showing the classic serine protease triad with catalytic residues Asp23, His58 and Ser218 represented as sticks.

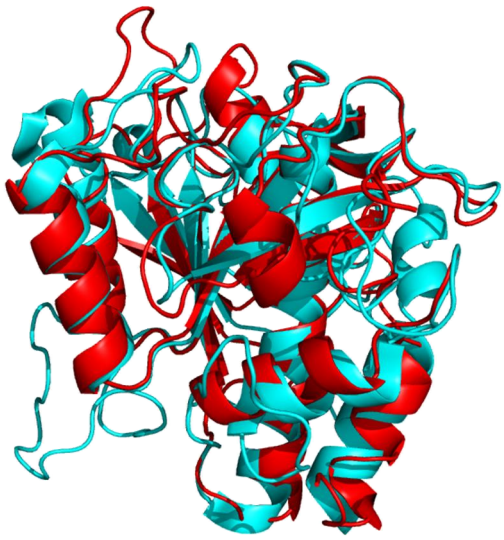


Figure 1.23: PatA Structural Alignment with Subtilisin. Structural alignment of the protease domain from PatA (red, PDB code: 3ZXX) with subtilisin (cyan, PDB code: 2ZRQ) shows a high degree of structural homology.

The cleavage by PatApr is a relatively slow reaction (up to 48 hours) when compared to heterocyclisation by PatD (two hours) [63], [79]. This may be intentional to ensure that heterocyclisation can be driven to completion before removal of the leader sequence which reduces heterocyclisation activity [67]*. In all related cyanobactin pathways, a subtilisin-like protease domain (homologous to PatApr) is always present (Figure 1.24) however the N-terminal cleavage recognition sites on the precursor peptide upon which they act can vary quite significantly (Table 1.1).

Protease	Organism	N-terminal Cleavage Site 1	N-terminal Cleavage Site 2
PatApr	*Prochloron sp.*	GLEAS	GVEPS
PagApr	*Planktothrix agardhii*	GLTPH	-
LynApr	*Lyngbya aestuarii*	GVDAS	-
TruApr	*Prochloron sp*	GVDAS	-
MicApr	*Microcystis aeruginosa*	GMDAS	GADAS
TenApr	*N. spongiaeforme var. tenue*	GVGAS	GAGAS

Table 1.1: PatA Protease and Homologue Recognition Sites. Variation of N-terminal cleavage sites within the cyanobactin precursor peptides which are recognised by the PatApr family of enzymes.

```
PatApr   ....MNRDILRTL..SLKGDHNIRVAILDGPVDIAHPCFQ   34
PagApr   MTKMTNIPGLKKLWSETRGDPKICVAVLDGIVDQNHPCFI   40
ArtApr   ....MEILGLSELQNRTQGSPEITIAVLDGPVDRTHPCFQ   36
TruApr   ....MNRDILRTL..SLKGDHNIRVAILDGPVDIAHPCFQ   34
TenApr   ........MLETLSTISQGNNSICVAVLDGPVDRTHPCFQ   32
LynApr   ........MLEKLSTISQGDSSICVAVLDGPVDQTHPCFQ   32
MicApr   ........MLDRLHTISQGDSSICVAVLDGPVDQAHPCFQ   32

PatApr   GADLTVLPTLAPTAARSDGFMSAHGTHVASIIFGQPETSV   74
PagApr   GADLTRLPSLVSGEANANGSMSAHGTHVASIIFGQHDSPV   80
ArtApr   GANLTTLPTLVPDAARSDGSMSGHGTHVTSIIFGQPGSDI   76
TruApr   GYDLTVLPTLAPTAARSDGFMSAHGTHVASIIFGQPETSV   74
TenApr   GADLKYLPSLVKDDAKINGNMSAHGTHVGSIIFGQPGSLV   72
LynApr   GADLTYLPTLVQDQAKADGNMSMHGTHVASIIFGQPGSPV   72
MicApr   GANLTYLPTLVQEAAKVDGSMSMHGTHVASILFGQPGSPV   72

PatApr   PGIAPQCRGLIVPIFSDDRRRITQLDLARGIERAVNAGAH   114
PagApr   TGIAPQCRGLIVPVFADESLKLSQLDLSRAIEQAVNNGAN   120
ArtApr   PGIAPQCRGLIVPVFADDRCSATQLDLARGIEQAANAGAQ   116
TruApr   PGIAPQCRGLIVPIFSDDRRRITQLDLARGIERAVNAGAH   114
TenApr   QGIAPQCKGIIVPVFSDDRLRTSQLDLARGIEQAVNAGAH   112
LynApr   EGIAPHCKGIIVPIFADDRRKTSQLDLARGIEQAVNAGAH   112
MicApr   QGIVPQCKGIIIPIFADDRRRTSQLDLARGIEQAVHAGAH   112

PatApr   IINISGGELTDFGEADGWLENAVSLCRQNNVLLVAAAGNN   154
PagApr   IINVSAGQLTDAGEADTWLEKAIQLCQEKNVLLIAATGND   160
ArtApr   IINLSGGQLSDFGEAEGWLENAVRLCQERNILLIAAAGND   156
TruApr   IINISGGELTDFGEADGWLENAVSLCRQNNVLLVAAAGNN   154
TenApr   IINLSGGQLTDYGESDGWLQNAVRLCRENNVLLVAAAGNN   152
LynApr   VINLSGGQLTDFGEADGWLENAVRLCRQNNVLLVAAAGNN   152
MicApr   IINLSGGQLTDFGESDGWLKNAVRLCRENNVLLVAAAGNN   152

PatApr   GCDCLHVPAALPAVLAVGAMDDHGHPLDFSNWGSTYEQQG   194
PagApr   GCECLHVPASLPTVLAVGAMDDQGKPVDFSNWGDAYQKQG   200
ArtApr   GCECLHVPAALPAVLAVGAMDARGQPLDFSNWGSKYQTQG   196
TruApr   GCDCLHVPAALPAVLAVGAMDERGHPLDFSNWGSTYEQQG   194
TenApr   GCDCLHVPAAMPAVLAVGAMDANGKPLDFSNWGEAYQSQG   192
LynApr   GCDCLHVPAALPSVLAVGAMDGNGKPLDFSNWGETYKNQG   192
MicApr   GCECLHVPAALPSVLAVGAMDAQGKPLDFSNWGEAYQTQG   192

PatApr   ILAPGEDILGAKPGGGTERLSGTSFATPIVSGVAALLLSE   234
PagApr   ILAPGKDILGAKPGGGTVKLSGTSFATPIVSGVAALLLSL   240
ArtApr   ILAPGEQILGAKPGGGTERLNGTSFAAPIVSGVAALLLSL   236
TruApr   ILAPGEDILGAKPGGGTERLSGTSFATPIVSGVAALLLSE   234
TenApr   ILAPGENILGAKVGGGTHRLSGTSFAAPIVSGVAALLLSL   232
LynApr   ILAPGKDILGAKPGGGTHRLSGTSLATPIVSGVAAVLLSL   232
MicApr   ILAPGEDILGAKAGGGTHRLSGTSFATPIVSGVAGVLLSL   232

PatApr   QVRRGETPDPQKVRQLLLQSALPCDDDAPEQARRCLAGRL   274
PagApr   QLKRGEKPDPQKVKNAILQSATPCNPKDTDDQSRCLMGKL   280
ArtApr   QRSQGETPDPARVRQILLQSALPCDAALPADAQRCLAGRI   276
TruApr   QVRRGETPDPQKVRQLLLQSALPCDDDAPEQARRCLAGRL   274
TenApr   QLERGETPDPHKIRQVLLQSALPCDADLPEDAKRCLAGKL   272
LynApr   QREQGETPDPQKIRQILLQSALPCDPDLPEETRRCLAGKL   272
MicApr   QREQGDSPDPQKIRQILLQSALPCDLDLPEESRRCLAGKL   272

PatApr   NVSGAFTLLKGGNMSEELATASFPSVE.ASCGC   306
PagApr   NILDAIEYLTGETMSEDLE...LESVEASGCGC   310
ArtApr   NIAGALTLLTGGNMSEAQEAVTPAGVEPS....   305
TruApr   NVSGAFTLLKGGNMSEELATASFPSVE.ASCGC   306
TenApr   NVSGAITLLKGGKMAEEFTSVSASEVEAAGCGC   305
LynApr   NVSGAITLLKGGKMAEEFPSVEVSEVQAAGCGC   305
MicApr   NVSGAITLLKGEKMSEEFSSVAASEVTAAGCGC   305
```

Figure 1.24: PatApr Homologue Alignment. Sequence alignment of the protease domain of PatA (PatApr) with its cyanobactin homologues showing a high degree of homology.

1.2.3 C-terminal Core Peptide Cleavage and Macrocyclisation

The C-terminal cleavage of the modified core peptide is catalysed by the PatG subtilisin-like protease domain (PatGmac). This process occurs after both heterocyclisation (PatD) and N-terminal cleavage (PatApr) because the leader sequence must be removed to allow the N-terminus of the core peptide to bond directly with its C-terminus [64]. PatGmac, like PatApr, contains a classic subtilisin triad and the active site is comprised of residues Asp548, His618 and Ser783 [84]′ [85]. However, unlike PatApr, PatGmac has macrocyclase activity. At a molecular level, it has been reported that PatGmac recognises the cleavage signal 'AYD' in the P1' – P3' positions (and possibly Gly in the P4' position), [88] whilst in many related cyanobactin pathways the C-terminal cleavage signal is 'SYD'. There is also the requirement for a heterocycle (thiazoline/thiazole, oxazoline/oxazole) or a proline in the final position of the modified core peptide, the residue which also constitutes the P1 position of the protease substrate [88].

Macrocyclisation is a common feature in both non-ribosomal and ribosomal peptide biosynthetic pathways. For non-ribosomal peptides macrocyclisation is generally carried out by thioesterase (TE) domains containing the same Asp-His-Ser catalytic triad as PatGmac and PatApr [89]. The substrate initially binds to a peptidyl carrier protein (PCP) as a thioester, followed by the transfer to the active site serine of a TE domain, forming an acyl-enzyme intermediate [90]. (Figure 1.25 A). The N-terminal amine of the peptide then attacks the ester carbonyl, displacing the serine and forming the macrocycle.

The molecular detail of how macrocyclisation occurs in ribosomally synthesised peptides was unclear. There is no homology between PatGmac and known thioesterase domains (< 4 %). PatGmac is proposed to proceed through cleavage of the peptide, leading to the formation of an acyl-enzyme intermediate with the catalytic serine, with no requirement for a PCP. Following removal of the 'AYD(G)'

cleavage recognition residues from the active site, the N-terminal amine of the modified core peptide attacks the acyl-enzyme intermediate, displacing the serine and yielding the macrocycle in a similar mechanism to the TE domain [63]' [88]. (Figure 1.25 B). The macrocyclisation process has been found to be particularly slow with only one reaction per enzyme per day [63]. This may be to allow time for the N-terminal of the peptide to orientate itself for the nucleophilic attack of the acyl-enzyme intermediate.

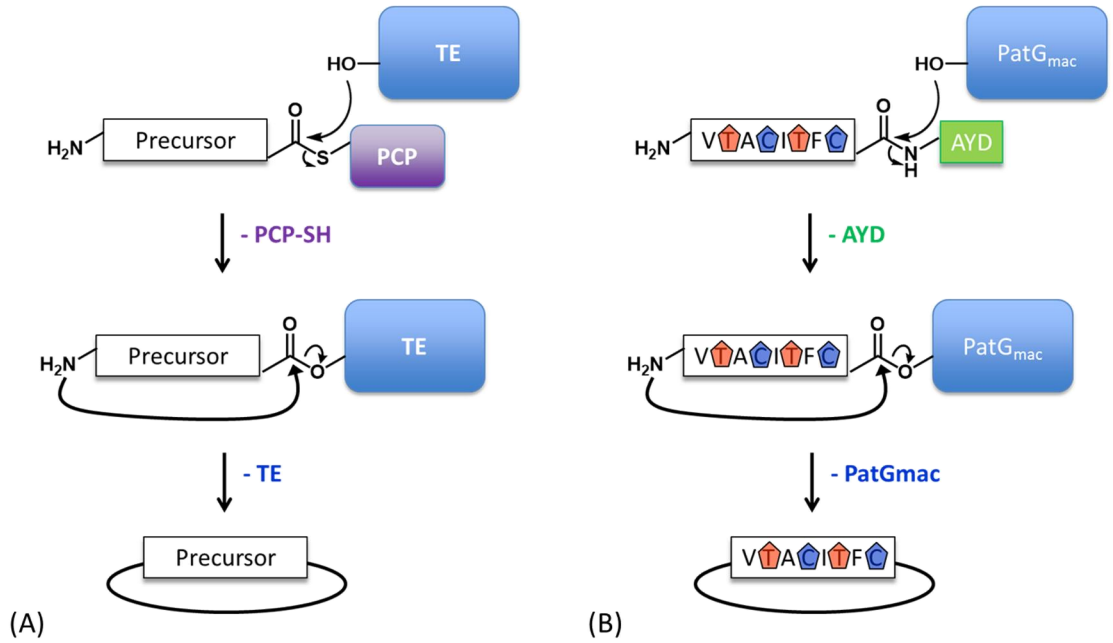


Figure 1.25: Peptide Macrocyclisation. Schematic of (A) peptide macrocyclisation via thioesterase intermediate formation and (B) the proposed peptide macrocyclisation by PatGmac showing N-terminal attack of the acyl-enzyme intermediate.

PatGmac is one of only a handful of ribosomal macrocyclases to have been isolated to date. These other macrocyclases are involved in pathways across the RiPP family including PCY1 of the caryophyllaceae cyclic peptide pathway [91], GmPOPB [92] in α-amanitin biosynthesis and butelase-1 in cyclotide production [93]. At present, no structural or mechanistic data has been determined on these macrocyclases however extensive kinetic analysis shows they have wide variation in their reaction rates. Butelase 1 (0.06 - 17 s^{-1}) and GmPOPB (5.6 s^{-1}) turnover at an order of magnitude of three to four higher than PCY1 (1 h^{-1}) and PatGmac (1 h^{-1}).

1.2.4 Epimerisation

In the patellamides, two of the amino acids, adjacent to the thiazoles, are epimerised to D-stereoisomers (Figure 1.26). The process of this epimerisation is likely to occur after heterocyclisation and before oxidation of the thiazoline as the α-carbon before a thiazoline is more liable to epimerisation due to the adjacent imine bond [94] (Figure 1.27). The pK_a of the α-carbon when adjacent to a thiazoline will be approximately 30 however when adjacent to a thiazole it will be higher at approximately 40 (estimated using JChem). Epimerisation post-oxidation is possible but disrupting the newly formed, highly stable, aromatic heterocycle makes this less likely.

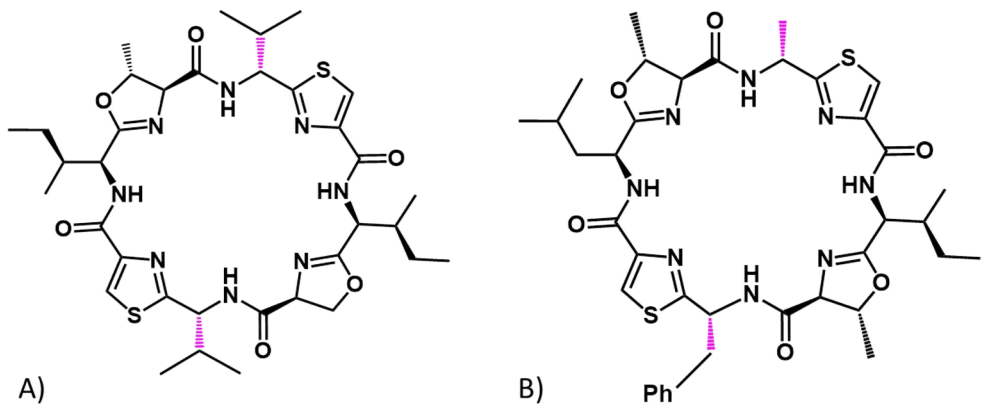


Figure 1.26: Epimerisation in Patellamides. Chemical structures of A) Patellamide A and B) Patellamide B with the epimerised amino acids adjacent to the thiazoles highlighted in magenta.

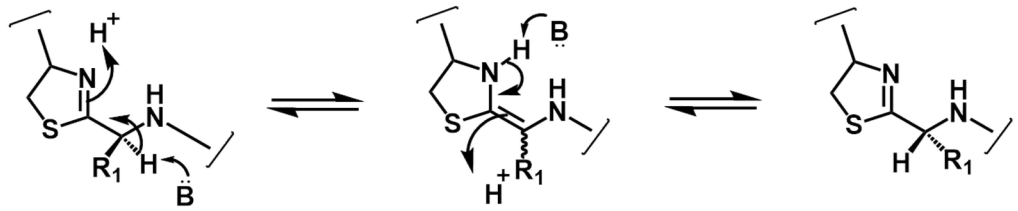


Figure 1.27: Epimerisation Mechanism. Mechanism schematic of the epimerisation of the amino acid adjacent to the thiazoline.

There are currently two hypotheses for how the epimerisation occurs; either it is controlled enzymatically or it can occur spontaneously [94]. Bioinformatics analysis of the patellamide gene cluster however shows no evidence of an epimerase enzyme [47]. As there are two epimerisation events, it is possible that these are catalysed individually by the related DUF domains of PatA and PatG (51 % homologous to each other, Figure 1.28), however as they show no sequence homology to any known epimerases, these would represent a novel class.

```
PatADUF   APSELANSQFAYVLGTLGYDFGTEARRDTFKQLMPPFDFAGN   458
PatGDUF   VEASTAFSGNVYALGTIGYDFGDEARRDTFKERM........   946

PatADUF   MVPANPYDARQMVDYLGNNISEARSLIWTVNIELTPVYAIDP   500
PatGDUF   ...ADPYDARQMVDYLDRNPDEARSLIWTLNLEGDVIYALDP   985

PatADUF   TGPFASSTYHALQELLSGQIQAEDNEEYVERVSIPGVLTNRS   542
PatGDUF   KGPFATNVYEIFLQMLAGQLEPETSADFIERLSVPARRTTRT  1027

PatADUF   VKLFSGQVVPVVEPQSTRGLYGWKVNGLVNAALEAVRAEGGD   584
PatGDUF   VELFSGEVMPVVNVRDPRGMYGWNVNALVDAALATVEYEE..  1067

PatADUF   AGEARIRQTLDGFLNRIYYDLRNLGTTSQDRALNFAVTNAFQ   626
PatGDUF   ADEDSLRQGLTAFLNRVYHDLHNLGQTSRDRALNFTVTNTFQ  1109

PatADUF   AAQTFSQSVAAGMELDSVTVEKSPFCRLDSDCWDIKLKFFDP   668
PatGDUF   AASTFAQAIASGRQLDTIEVNKSPYCRLNSDCWDVLLTFYDP  1151

PatADUF   ENNRRAKKIYRFTIDVSDLVPVTMGEVRSWSSSY......   702
PatGDUF   EHGRRSRRVFRFTLDVVYVLPVTVGSIKSWSLPGKGTVSK   1191
```

Figure 1.28: PatA and PatG DUF Alignment. Sequence alignment of the C-terminal domains of unknown function from PatA and PatG.

The structure of the PatG DUF domain (PatG residues 914-1191) has been determined by Greg Mann (University of St Andrews) [95] yet has given no further insight into its potential role in epimerisation (Figure 1.29) or indeed any other reactions however binding studies have ruled out any interaction with the leader peptide [95]. Alternatively, the epimerisation reaction could be non-enzymatic as studies have shown that lissoclinamide, a related cyanobactin, chemically synthesised in the wrong enantiomer can be epimerised by heating to 60 °C in mild base [94]. This may be a result of conformation change in the patellamide which influences epimerisation.

Although a spontaneous epimerisation remains an option, occurring under these harsh conditions would be unlikely in nature.

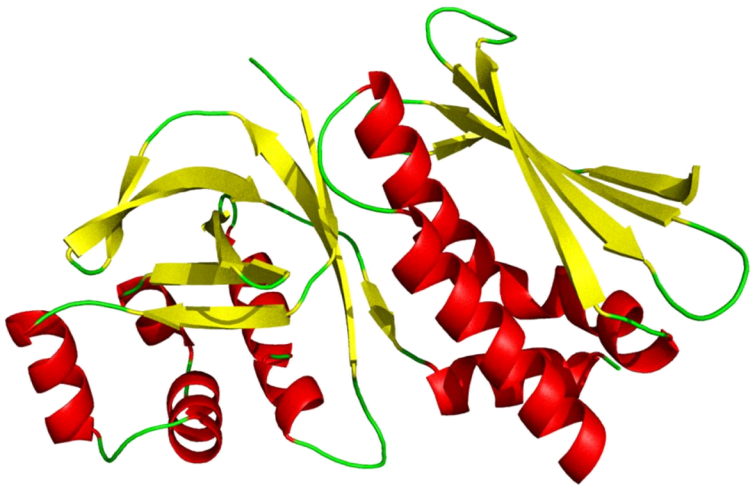


Figure 1.29: PatG DUF Structure. Secondary structure representation of the C-terminal DUF domain from PatG (alpha helices – red, beta sheets – yellow) (PDB code: 4UVQ, [95])

1.2.5 Oxidation

Heterocyclisation of threonine/serine and cysteine residues results in the formation of oxazolines and thiazolines respectively as described previously. In final natural products the thiazolines have been further oxidised to form thiazoles (Figure 1.30) [49]. Although at present there is no definitive evidence that the N-terminal domain of PatG carries out this oxidation, it would appear to be highly likely as sequence analysis of this domain shows a high degree of conservation with other oxidases [47], and most cyanobactin pathways also contain a homologous oxidase domain either within the PatG equivalent or as a stand-alone protein [34].

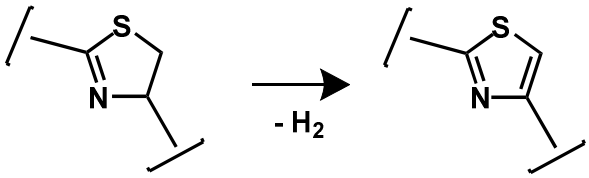


Figure 1.30: Thiazole Structure. Chemical structure of thiazoline and its oxidised analogue thiazole.

It is also currently unknown at what stage oxidation occurs, whether before or after macrocyclisation of the modified core peptide. An interesting observation is the lack of oxazoles within the patellamides [49], suggesting that the PatG oxidase domain is specific in oxidising only thiazolines. This is in contrast to other cyanobactins where oxazoles and thiazoles occur together (e.g. tenuecyclamides [36], Figure 1.5 A). This may be a result of either substrate recognition (only binds thiazolines) or down to chemistry (can only oxidise thiazolines). At this stage it is also unclear which co-factors are required for oxidase activity, although FMN is required for similar thiazoline oxidases such as in the TOMMS, bleomycin and epothilone biosynthetic pathways [96] [97].

1.2.6 Prenylation

Another post-translational modification found in many of the cyanobactins is the prenylation of specific residues within the core peptide (Figure 1.31). Prenylation, the addition of a prenyl group (3-methyl-but-2-en-1-yl) (Figure 1.32 A), can occur in the forward or reverse positions commonly on oxygen or carbon atoms in cyanobactins (Figure 1.32 B, C) [65]. In proteins, multiple prenyl groups can be linked together to from longer chains (e.g. farnesylation, geranylgeranylation [98] [65]) however in cyanobactins only prenylation and in rarer cases geranylation occur (e.g. piricyclamides [99]). Prenylation is generally associated with membrane targeting and cellular localisation or in protein:protein interactions, while they can also make compounds more lipophilic aiding in pharmacokinetics [100]. In the cyanobactins, the prenyl group is derived from dimethylallyl pyrophosphate (DMAPP) (Figure 1.32 D), an intermediate of the mevalonate pathway [65] [101].

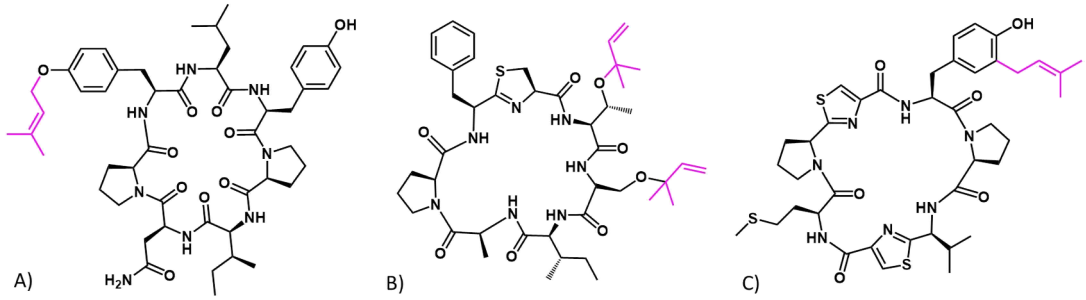


Figure 1.31: Cyanobactin Prenylation. Chemical structures of cyanobactins containing prenylated side chains (magenta). (A) Prenylagaramide B exhibiting forward O-prenylation on a tyrosine residue (B) Trunkamide A containing reverse O-prenylation on both threonine and serine residues [65] (C) an Aestuaramide showing forward C-prenylation on a tyrosine residue [102].

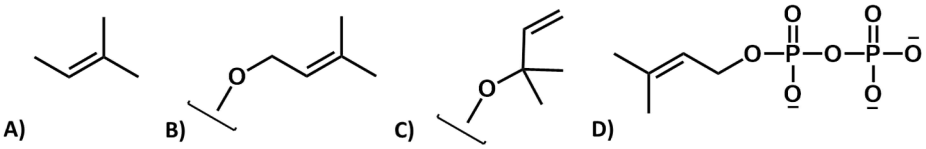


Figure 1.32: Prenylation. Chemical structures of (A) a prenyl group, (B) O-linked forward prenylation (C) O-linked reverse prenylation and (D) Dimethylallyl pyrophosphate (DMAPP).

To date the function of PatF has not been confirmed however studies on related cyanobactin pathways have shown that the PatF family of enzymes is responsible for the prenylation of specific amino acids within the core peptide [65]' [103]. TruF1 from the trunkamide pathway (41 % homologous to PatF [62]) catalyses the addition of prenyl groups to the hydroxyl group of threonine and serine residues in the macrocycle [103], while studies on LynF (44 % homologous to PatF) from the Aestuaramide pathway (*Lyngbya aestuarii)* show that it catalyses prenylation on the hydroxyl of tyrosine residues which is followed by a Claisen rearrangement to yield forward C-prenylation ortho to the hydroxyl group on the aromatic ring [65] [104] [102].

Of the patellamide natural products isolated and studied to date, there is no evidence of prenylation [49]. This could be explained by the lack of naturally occurring tyrosine

residues within the core peptide or that PatD, unlike TruD, heterocyclises threonine and serine residues early in the biosynthesis process and as a result these residues would no longer be amenable to prenylation.

Interestingly, although these proteins have some degree of sequence homology among themselves, they show no sequence homology to other known prenyl transferases. The PatF family of enzymes therefore represent a novel subclass of prenyl transferases unique to the cyanobactins.

1.2.7 Non-defined Patellamide Proteins

The PatB and PatC proteins have been determined to be non-essential in the production of patellamides [47], [61], yet these proteins are highly conserved across cyanobactin pathways (Figures 1.33, 1.34). Nevertheless, their protein sequences do not offer any insights into their potential function [34]. There can be several hypotheses for the function in which they may be involved from offering the host a form of resistance against the patellamide or to regulators of the pathway.

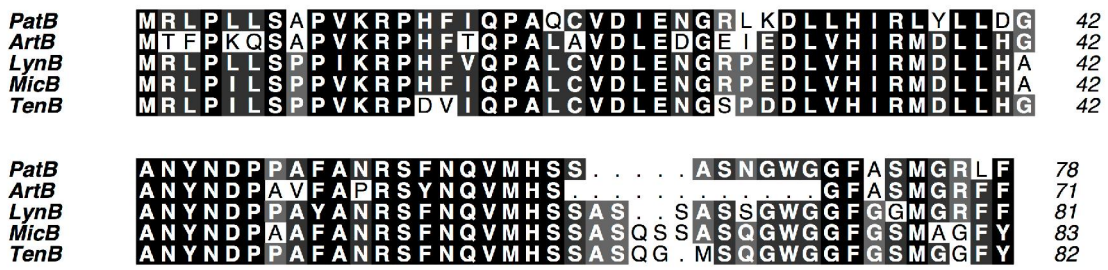


Figure 1.33: PatB Homologue Alignment. Sequence alignment of PatB with its cyanobactin homologues showing a high degree of homology.

```
PatC   ..................................................
ArtC   ..................................................
LynC   MPKPIANAQPSIDPSTLSVAQFDGKKDYIEVGFQEVLNPPQFT   43
MicC   ..................................................
TenC   ...................................MPQQT     5

PatC   ..................................................
ArtC   ..................................................
LynC   VECWAKVTGGQGSWRSPITSRGEGPLAGYIIYAGQNNKWQFWI   86
MicC   ..................................................
TenC   QTLWQIGQPGKGGQEFS....QEG........GWQGEFTYVV   35

PatC   ..................................................
ArtC   ..................................................
LynC   GSGGSWS.TYQGPAIKLNAWTHCAGSYDGEVMRFYVNGKEVGS  128
MicC   ..................................................
TenC   GSDSDPINQPQMPSLLV.....VPG........LKKKPRSKGK   65

PatC   ..................................................
ArtC   ..................................................
LynC   PVTAPLTLNTRFPLRIGTGCTEQSPQFFFNGQIAE.VRVWNCV  170
MicC   ..................................................
TenC   QLFSTEKLNIQFTLTQSYNPGEL..TLFYNFFGSEVDTLWVDG  106

PatC   ..................................................
ArtC   ..................................................
LynC   RTPEEIQD.......................YMNQRCTGKEP   189
MicC   ..................................................
TenC   EPVTEIVGVGSSKVTQAQIPLPALSAGEHTLTLTTASQSGDGA  149

PatC   ......................................MVTNNPTA    8
ArtC   ....................MMAKKNTPKN..PDDLPETAPAA   21
LynC   SLVAYWPLSDSAGESAKDQTRSVNHGVLQG....ATWNTTDD  227
MicC   ..............MSKNKTTTTSSEVPSNPEPTPPEPSKSEP   29
TenC   HRIDYLKLEAVVATQQP..QTLQTPQTPKETMAKEKTETTPPA  190

PatC   QSKTPPAEKRKSYVMTSMTGMQDYSYWCAYVRERYGEEAKKYE   51
ArtC   NPATPPAPAPKSYVPTAMTGLQDYSYWWEYAREH....ANKPK   60
LynC   LPLTTPEPPKQSYVPTSQTGLQDYSYWCQYVKDQ....AKKSD  266
MicC   TTI.TTPEPKKSYVPTSQTGLQDYSYWCQYVKDQ....AKKSD   67
TenC   VSPTPEPKKKKSYVPTSQTGLQDYSYWWEYARQH....AKAPE  229

PatC   RDPKPFRRGGVWA   64
ArtC   PNSKTFRRGRIWA   73
LynC   KDPKSFRRGRIWA  279
MicC   KDPKSFRRGRIWA   80
TenC   .EGKSFRRGRIWA  241
```

Figure 1.34: PatC Homologue Alignment. Sequence alignment of PatC with its cyanobactin homologues showing a highly conserved region. LynC and TenC however have significant additions to their N-terminus in comparison to PatC, ArtC and MicD.

1.3 Aims

This thesis is concerned with the ribosomal biosynthesis of the cyanobactin cyclic peptide superfamily, with particularly emphasis on the patellamides.

The study of the prenylase family of enzymes from cyanobactins (PatF family) will be explored through structural and biochemical methods to obtain an understanding of prenylation on the final macrocyclic peptides.

The PatG macrocyclase domain will be explored in structural and biochemical studies to determine substrate binding and rationalise the mechanism of macrocyclisation.

The oxidation step in patellamide biosynthesis has not been characterised to date, but is believed to be catalysed by the N-terminal oxidoreductase domain of PatG. This protein and its homologues will be explored both structurally and biochemically to confirm this hypothesis and determine its associated mechanism.

Through isolation of the proteins (or their individual domains) from the patellamide (or related) biosynthetic pathway(s) the reconstitution of patellamide production *in vitro* will be pursued the process used to create a diverse range of compounds including known and unknown amino acid sequences, and also in varying peptide length.

Once the pathway is established *in vitro,* identification of the flexibility of the process by using unnatural amino acids in the precursor peptide to incorporate them in the final patellamide product will be studied. These precursor peptides can be derived from cell growth on media containing unnatural amino acids or by semi-synthesis with techniques such as intein technology.

2. Structural and Biochemical Studies of PatF and Homologues

2.1 Introduction

PatF has previously been reported as essential in the production of patellamides on the basis of studies knocking out the gene encoding for it which results in no detectable patellamide products [47]. The biochemical role of PatF in patellamide production has yet to be confirmed; however related pathways have revealed the PatF family to be prenyl transferases. LynF, from *Lyngbya aestuarii* [65]' [104] and PagF from *Planktothrix agardhii* (unpublished, Nair *et al.*) have both been shown to prenylate tyrosine residues, while TruF1 from the trunkamide pathway (*Prochloron* sp.) prenylates serine and threonine residues [103] (Figure 2.1 A). At present no patellamide natural products have been discovered where prenylation is evident so it is unknown which residue(s) PatF acts on (if any).

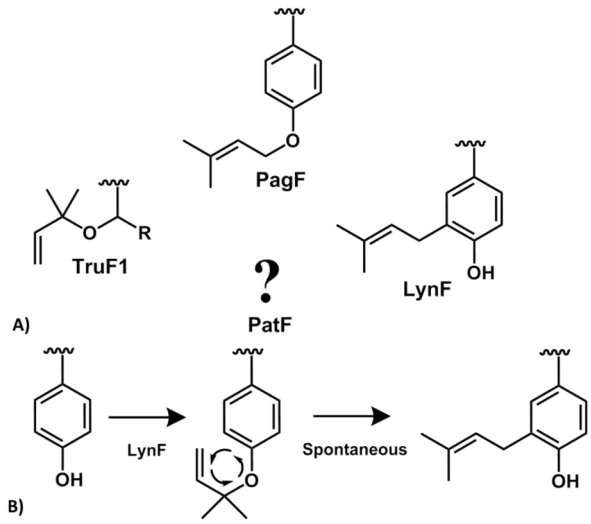


Figure 2.1: PatF Family Prenylation. (A) Products of cyanobactin prenylation, PagF – forward O-prenylation of tyrosine, LynF – Forward C-prenylation of tyrosine, TruF1 – reverse O-prenylation of Ser/Thr (R=Me or H) and PatF – unknown. (B) Mechanism for LynF catalysed prenylation; initial reverse-O prenylation on tyrosine followed by a spontaneous Claisen rearrangement to give forward C-prenylation. Figure adapted from McIntosh *et al.* (2011) [65]

It is clear that there are significant differences in the mechanisms associated with the closely related prenylases. LynF catalyses the reverse O-prenylation of tyrosine residues which is followed by spontaneous Claisen rearrangement resulting in forward C-prenylation of the tyrosine aromatic ring *ortho* to the hydroxyl group (Figure 2.1 B) [65] [102] [104]. PagF carries out forward prenylation directly on the hydroxyl group with no rearrangement (unpublished, personal communication, S. Nair - University of Illinois). Finally, TruF1 also reverse prenylates on oxygen, this time in threonine and serine residues.

The X-ray crystal structure of PatF was sought in order to explore its potential function and give insight into its catalytic mechanisms. Biochemical characterisation of both PatF and its homologues using *in vitro* assays with a range of different amino acid substrates was also pursued.

2.2 Materials and Methods

2.2.1 DNA Cloning

The gene encoding full length PatF was synthesised in the pJexpress 411 plasmid using optimised codons for *E. coli* (DNA 2.0) [105]. The plasmid consists of full length PatF with an N-terminal His_6-tag and additional Arg and Ser residues at the C-terminus (a cloning artefact) (Figure 2.2 A).

A Tobacco Etch Virus (TEV) protease recognition site was subsequently introduced between the His_6-tag and the PatF protein (Figure 2.2 B) using the site-directed mutagenesis (SDM) technique of Liu & Naismith [106] (Figure 2.3). Polymerase chain reactions (PCR) were carried out on the original plasmid using KOD polymerase (Novagen) with specifically designed mutagenesis primers (Life technologies, Table 2.1) to insert the nucleotide sequence for TEV recognition site. The template DNA (methylated) was digested using DPN-1 enzyme (20 units) leaving only the newly synthesised (mutated) plasmid. This plasmid was used to transform DH5α *E. coli* cells. Single colonies were picked and grown in 10 ml Luria-Bertani (LB) media at 37 °C, 200 rpm overnight. The cultures were harvested (4,000 x g) and the plasmid DNA was extracted using the Qiagen mini-prep kit. Purified plasmid DNA was sequenced (GATC Biotech) to confirm the presence of the desired mutation(s).

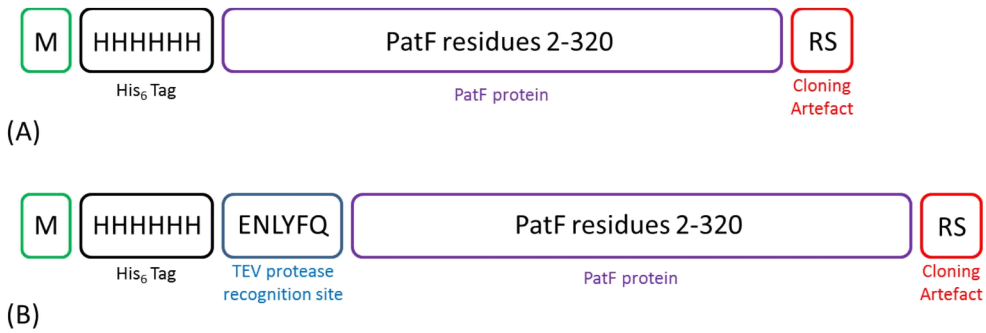


Figure 2.2: Schematic for PatF Constructs. (A) Original PatF construct with N-terminal His_6-tag and C-terminal RS residues. (B) Construct modified from original with the addition of a TEV protease recognition site between the His_6-tag and the protein.

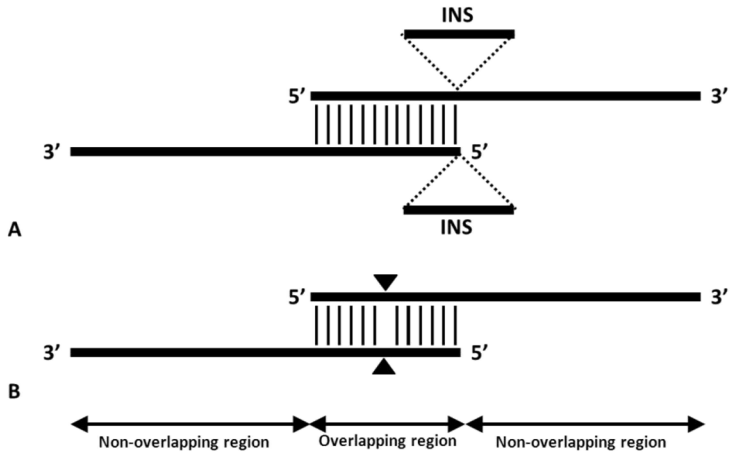


Figure 2.3: Schematic for Primer Design for Site-directed Mutagenesis. Primer design using non-overlapping and overlapping regions for (A) insertion and (B) point mutations. INS indicates the location of insertions, and triangles indicate the locations of mutations in the primer sequences. Figure adapted from Liu & Naismith (2008) [106].

Point mutations of the active site (M136K and H125D/G127R/M136K) were created using the mutagenesis technique using primers shown in Table 2.1. (Note: The triple mutant was made from a single PCR experiment using the single mutant, M136K, as the template.)

5' PatFTEV	gaaaacctgtattttcaggacttgatcgaccgtctgcag
3' PatFTEV	gtcctgaaaatacaggttttcgtggtgatggtgatgatgcat
5' PatF H125DG127R	attggtgtggatctgcgtagcaagttggaggacagcagcgtc
3' PatF H125DG127R	caacttgctacgcagatccacaccaatggtgttattgataat
5' PatF M136K	agcagcgtcaaactgtacattcacatcaaaccggaa
3' PatF M136K	aatgtacagtttgacgctgctgtcctccaacttgct

Table 2.1: PatF Primer Sequences. Mutagenesis oligonucleotide sequences for PatF mutagenesis experiments.

Full length LynF was synthetically cloned into the pJexpress 411 plasmid using optimised codons for *E. coli* (DNA 2.0) [105]. The plasmid consists of full length LynF with an N-terminal His_6-tag and TEV protease site (Figure 2.4 A).

The LynF gene was subsequently PCR amplified using specifically designed primers to yield NcoI and XhoI restriction sites on the 5' and 3' ends respectively (Table 2.2). The amplified DNA was then digested with *NcoI* and *XhoI* restriction enzymes and ligated into the pSUMO plasmid (a gift from C.D. Owen, University of St Andrews) consisting of an N-terminal His_6-tag, SUMO (Small Ubiquitin-like Modifier) protein to aid solubility and TEV protease site (Figure 2.4 B).

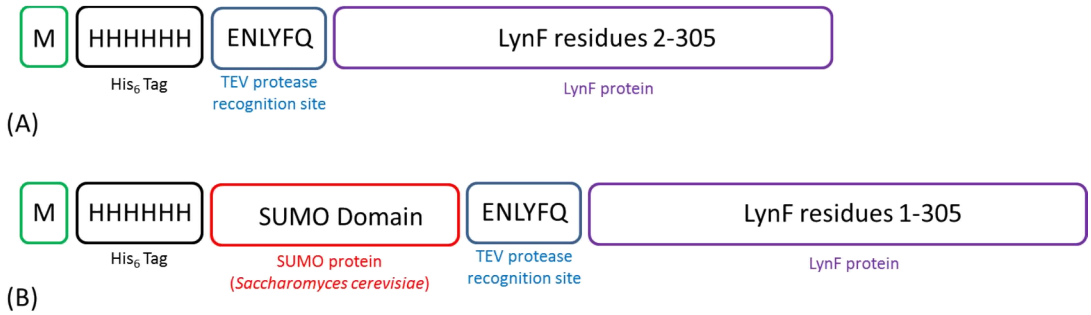


Figure 2.4: Schematic for LynF Constructs. (A) LynF construct with N-terminal His_6-tag and TEV protease recognition site. (B) LynF SUMO construct with N-terminal His_6-tag, SUMO domain and TEV protease recognition site.

5' LynFNcoI	ggcgccatggcgattgcaaaccgtgtaccgtac
3' LynFXhoI	tttctcgagttagccgaagctacgacggta

Table 2.2: LynF Primer Sequences. Oligonucleotide sequences for LynF cloning experiments.

2.2.2 PatF Expression and Purification

PatF was expressed from the pJexpress 411 plasmid using BL21 (DE3) *E. coli* cells grown on auto-induction medium using the Studier method [107] (see Appendix A.1 for media and buffer composition). The cultures were grown at 20 °C, 250 rpm for 48 hours before harvesting by centrifugation.

(L)-Selenomethionine-labelled PatF was expressed in BL21 (DE3) *E. coli* cells, cultures of which were grown in a minimal medium supplemented with glucose-free nutrient mix (Molecular Dimensions) and 5 % glycerol. The medium was inoculated with overnight

culture grown in LB medium which was subsequently washed three times with minimal medium. After 15 min growth at 37 °C, 60 mg L^{-1} (L)-selenomethionine was added to the cultures. An amino acid mix (100 mg L^{-1} lysine, phenylalanine and threonine, 50 mg L^{-1} isoleucine and valine) was added to the cultures at an optical density at 600nm (OD_{600nm}) of 0.6. After 15 min further growth at 37 °C the cultures were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown for 30 hours at 20 °C before harvesting by centrifugation.

Cell pellets of PatF were re-suspended in lysis buffer plus EDTA-free protease inhibitor tablets (Roche) and DNAse at 0.4 mg g^{-1} wet cell pellet. The re-suspension was lysed by passage through a cell disruptor at 30 kPsi (Constant Systems). The lysate was cleared by centrifugation (40,000 x g, 4 °C, 20 min) and then loaded onto a Ni-Sepharose 6 FF column (GE Healthcare) equilibrated in lysis buffer. The column was washed with lysis buffer and then PatF eluted with elution buffer. The protein was then passed over a desalt column (Desalt 16/10, GE Healthcare) in desalting buffer. TEV protease was added at a mass-to-mass ratio of 1:5 TEV protease : protein and incubated for two hours at 20 °C to remove the His_6-tag. The sample was then loaded on to a second Ni-column, in desalt buffer supplemented with 20 mM imidazole. PatF was collected in the flow through. The protein was concentrated to 7.5 ml (Vivaspin concentrators, 10 kDa MWCO) and applied to a Superdex 75 gel filtration column (GE Healthcare) equilibrated in gel filtration buffer. The protein was concentrated to 8 mg ml^{-1} for crystallography. The purity of the protein was confirmed by SDS-PAGE analysis and its identity confirmed by mass spectrometry (MS). All MS experiments in this thesis were carried out by the University of St Andrews BSRC Mass Spectrometry team.

2.2.3 LynF Expression and Purification

LynF was expressed from the pJexpress 411 and pSUMO plasmids using BL21 (DE3) *E. coli* cells, grown on auto-induction medium using the Studier method [107]. The cultures were grown at 20 °C, 250 rpm for 48 hours before harvesting by centrifugation.

Cell pellets of LynF from pJexpress 411 were re-suspended in lysis buffer (Appendix A.2) and sonicated at 10 microns (SoniPrep 150, MSE) and centrifuged at 30,000 x g with the soluble fraction then applied to a His-tag pull down assay (Qiagen BioSprint). The elution and flow-through samples were analysed by SDS-PAGE to determine any soluble expression.

Cell pellets of LynF from pSUMO were processed using the same method as PatF (Section 2.2.2) with the SUMO-tag removed at the same time as the His_6-tag. The only exception to this was that TEV protease was added at a mass to mass ratio of 1:10 and a slight variation of buffers (Appendix A.2). The purity of the protein was confirmed by SDS-PAGE analysis and its identity confirmed by mass spectrometry (MS)

2.2.4 PatF Crystallography

Initial crystallisation trials were set up using the sitting drop vapour diffusion method by means of the Honeybee 963 crystallisation robot (Genomic Solutions). The protein was screened against sparse matrix screens prepared in-house, composed from a range of known crystallisation conditions [108] (Appendix E). 150 nl and 300 nl protein was mixed in sitting drop wells with 150 nl precipitant and equilibrated over a well of 70 µl reservoir solution. Plates were set up and incubated at 20 °C. Optimisation trials were carried out manually using the hanging drop method with drops of 1:1 µl or 1:0.5 µl protein to precipitant ratios equilibrated over a well of 500 µl. Optimisation trials were set up and incubated at 20 °C. Optimal crystals of size 100 x 30 x 30 µm were grown in five days by varying the buffer pH, chemical concentration and altering the chemical composition of the precipitant.

Crystals were cryo-protected in solutions of mother liquor containing up to 30 % glycerol before being flash frozen in liquid nitrogen. Crystals were screened in-house at 100 K using a Rigaku 007HFM rotating anode X-ray generator and a Saturn 944 CCD detector. Datasets were collected in-house or at the I04 beamline at the Diamond Light Source.

2.2.5 PatF Additive Screen

Attempts to improve the quality of optimised crystals were carried out using the Additive Screen [109] (Hampton Research). Crystal trials were set up manually as previously described (Section 2.2.4) with the addition of 0.3 µl additive to the protein-precipitant drop. The precipitant in both drop and reservoir consisted of 0.1 M sodium / potassium tartrate, 26 % PEG 2K MME.

2.2.6 PatF Seeding

For seeded crystallisation trials a single native crystal was transferred initially into 10 µl of fresh mother liquor (0.1 M sodium / potassium tartrate, 26 % PEG 2K MME) and crushed with a microtool (Hampton Research). The entire slurry was then transferred to a tube containing 290 µl mother liquor and briefly vortexed to yield a microcrystal stock. Manual crystallisation trials were set up in hanging drop trays with 2 µl of 4 mg ml^{-1} PatF mixed with 0.5 µl microcrystal stock solution and 1 µl reservoir solution and equilibrated against a 500 µl reservoir of precipitant.

2.2.7 PatF Data Collection and Structure Solution

The crystal of SeMet PatF was cryo-protected in a solution of mother liquor supplemented with 30 % glycerol and flash frozen in liquid nitrogen. A single wavelength anomalous dispersion (SAD) dataset was collected at the Se-K absorption edge at 100 K on beamline IO4 (Diamond Light Source). The data were processed and scaled using the xia2 [110] package incorporating XDS for integration [111] and Scala for scaling [112]. Initial analysis of the processed data was carried out with SHELX C, D and E [113] [114]. The structure was solved using AutoSol as part of the PHENIX suite [115]. Automated model building of the chains was carried out using PHENIX AutoBuild [115]. The model was then refined by iterative cycles of manual rebuilding in Coot [116] and refinement using REFMACv5 [117] as part of the CCP4 suite [118]. PISA [119],

[120] was used to assess the oligomeric state of the protein. TLS restraints were calculated using the TLSMD server [121], [122] and used in refinement [123]. The structure was validated using MolProbity [124]. Structural coordinates were deposited in the Protein Data Bank - PDB code: 4BG2.

2.2.8 Homology Model Building

Homology models of LynF and TruF1 were created using the "one to one threading" module of Phyre2 [125] by inputting the respective protein sequence along with PDB coordinates for the determined PatF structure.

2.2.9 Structural Alignments

Structural alignments of PatF and related proteins were performed in Coot [116] using the Secondary Structure Matching (SSM) Superpose tool. Structural coordinates of related proteins were downloaded from the Protein Data Bank (PDB) [126] or received directly from their authors.

2.2.10 Biochemical Studies of PatF and LynF

Biochemical reactions of both PatF and LynF were set up on the 100 µl reaction scale with 10 µM enzyme, 1 mM amino acid derivative, 1 mM DMAPP, 1 M NaCl, 12 mM $MgCl_2$, 10 mM HEPES pH 7.4 and 3 mM TCEP incubated at 37 °C either overnight or for 72 hours. 50 mM amino acid derivative stocks of Boc-Tyr, Boc-Ser and Boc-Trp were made up in a 50 % DMSO solution due to solubility. DMAPP was purchased from Cayman chemical and solubilised in water to 10 mM. All reactions were analysed by LC-ESI MS (Micromass LCT) using a Phenomenex C18 column run in H_2O / 0.1 % formic acid vs methanol / 0.1 % formic acid gradient.

2.3 Results

2.3.1 PatF Expression, Purification and Crystallisation

PatF was overexpressed with an N-terminal His_6-tag in BL21 (DE3) *E. coli* cells using the Studier auto-induction method [107]. Initial purifications were carried out using a construct lacking a TEV protease site, with the protein applied directly to gel filtration following desalting. The protein yield was 8 mg L^{-1} culture. Initial crystallisation trials of PatF (with tag remaining) against sparse matrix screens were carried out at 8 mg ml^{-1} [108]. Crystal hits were evident in several drops, however the visual morphology was poor and subsequent attempts to optimise these proved difficult. In one experiment a single crystal grown from a precipitant composition of 24 % PEG 5000 MME, 0.15 M sodium / potassium phosphate pH 6.6 was achieved (Figure 2.5). This crystal diffracted to 2.08 Å and a dataset was collected (Table 2.3).

Subsequent attempts to reproduce this result proved difficult with the growth of only fine needles which showed poor diffraction by X-ray.

In order to assess if the His_6-tag was hindering crystallisation, site directed mutagenesis was carried out on the protein to introduce the TEV protease cleavage signal 'ENLYFQ' between the His_6-tag and the protein [127]. The tag was removed during the purification process using TEV protease with approximately 80 % efficiency and the final purified protein placed into crystallisation trials at 8 mg ml $^{-1}$. A final yield of 4 mg L^{-1} culture was achieved (Figure 2.6).

Purified PatF following tag removal gave several hits in crystallisation trials but like the PatF with tag remaining, visual morphology was poor. Single needles (100 x 5 x 5 µm) of poor diffraction quality were evident and optimisation in both nanolitre and microlitre drops yielded no improvement in these crystals.

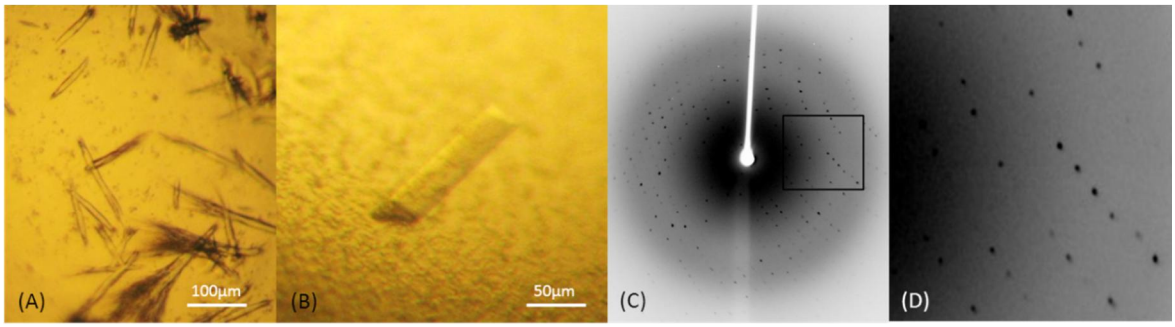


Figure 2.5: Crystallography of PatF. (A) Initial screening of PatF (with tag) in PEG-based screens gave needle-like crystals. These were grown in a condition of 26% PEG 5000 MME, 0.21 M sodium / potassium phosphate pH 7.0. (B) Optimisation of these crystals resulted in the formation of a single 3-dimensional crystal. This crystal was grown in a condition of 24 % PEG 5000 MME, 0.15 M sodium / potassium Phosphate pH 6.6 (C) Single crystal diffraction to 2.08 Å in house. (D) Zoom view of diffraction spots

PatF Native dataset	
Wavelength (Å)	1.54
Space group	P 2_1 2_1 2_1
a, b, c (Å)	46.1, 51.5, 144.1
α, β, γ (°)	90, 90, 90
Resolution (Å)	2.08 (2.19-2.08)
I/σI	16.9 (7.98)
R_{merge} (%)	9.6 (32.5)
Completeness (%)	99.5 (94.4)
Multiplicity	5.9 (4.0)

Table 2.3: Data Collection Statistics for a Native Dataset. The statistics provided are averages with values for the highest resolution shell provided in parentheses. The dataset was collected in house (Rigaku 007-HFM, Saturn 944 CCD).

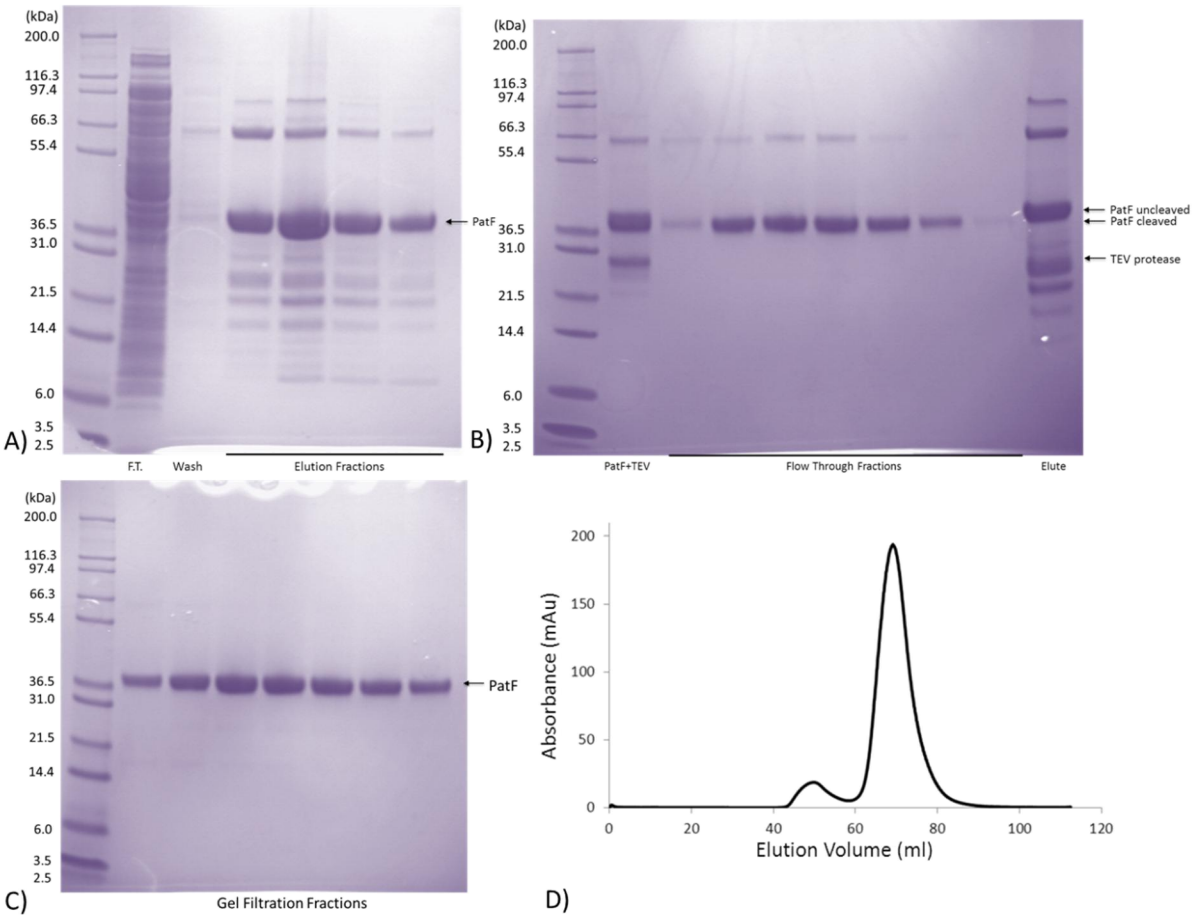


Figure 2.6: Purification of PatF. (A) SDS-PAGE analysis of Ni-sepharose column fractions showing PatF isolated from the soluble lysate. (B) SDS-PAGE analysis of second Ni-sepharose column flow through and elute fraction of PatF treated with TEV protease indicating approximately 80 % cleaved PatF (C) SDS-PAGE analysis of final gel filtration fractions showing the sample to be > 98 % pure. (D) Gel filtration chromatograph of PatF with a single elution peak corresponding to a monomer.

2.3.2 PatF Additive Screening

Additive screens consist of small molecules including detergents, alcohols, salts and metal ions. These additives can aid protein solubility, increase stability between protein molecules and therefore improve crystallisation. PatF was screened against the

Hampton Additive screen (Hampton Research). The additives were added directly to the protein-precipitant drop with precipitant lacking additive in the reservoir. The precipitant was that of the best quality native crystals (0.1 M sodium / potassium tartrate, 26 % PEG 2K MME)

Several conditions improved the visual quality of the crystals with the addition of barium chloride, strontium chloride, taurine or sucrose giving single three-dimensional crystals, however when screened by X-ray these crystals showed no improvement in diffraction quality and in most cases had detrimental effects when compared to native crystals.

2.3.3 PatF Seeding

Multiple iterations of optimisation resulted in no significant improvement in PatF crystals with only thin needles, which diffracted poorly, being produced (Figure 2.8 A). In order to progress to well diffracting three-dimensional crystals, seeding experiments were carried out.

Vapour diffusion crystallisation occurs through the removal of water from the protein-precipitant drop into the precipitant solution in the reservoir. This causes the protein concentration to increase and shift to a supersaturated state, either in the form of precipitation or ideally into the labile zone where nucleation and crystal growth can occur (Figure 2.7). Upon nucleation, the protein concentration in solution will reduce and under the ideal condition will move into the metastable zone resulting in continued growth with no further nucleation [128].

Seeding can be used to bypass the need for a nucleation event and we can place the crystallisation experiment directly into the metastable zone with seeds as the nucleant to achieve sustained crystal growth. Seed crystals can be any crystals of the target protein (no matter how poor) and are normally crushed, vortexed and diluted prior to addition to new crystallisation experiments.

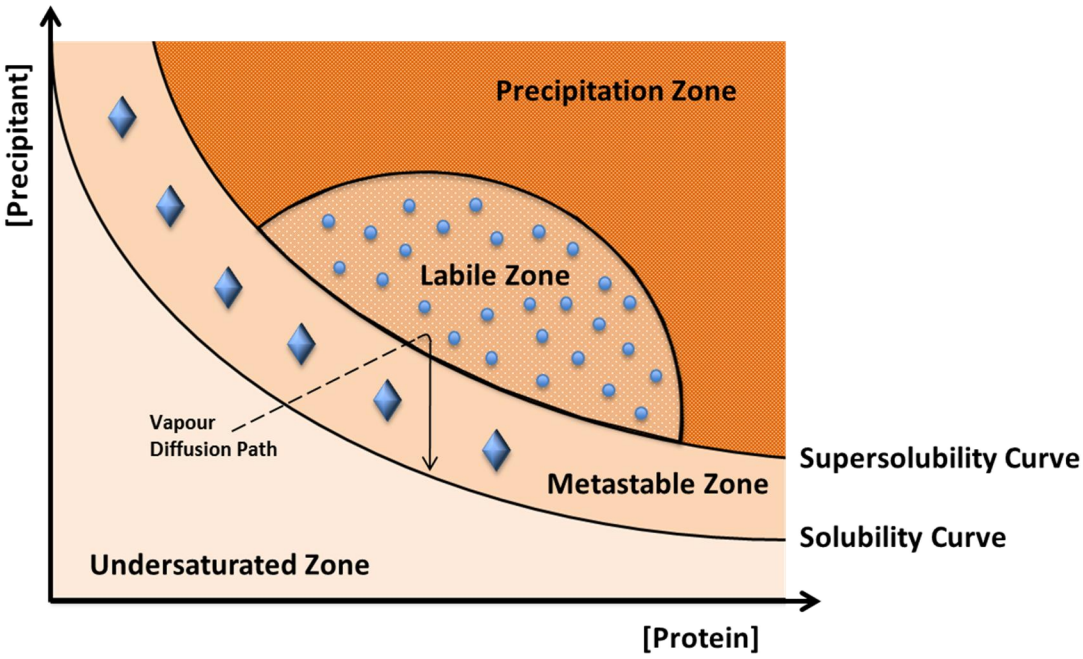


Figure 2.7: Phase Diagram for Crystal Growth. Vapour diffusion crystallisation occurs through the removal of water from protein-precipitant drop. Upon crystal trials the protein will be soluble (undersaturated) however as water is removed it will become supersaturated. If this occurs under the optimum conditions, the protein will move to the labile zone where nucleation can occur. This nucleation will cause a decrease in solution protein concentration and move it to the metastable zone where continued crystal growth but no further nucleation will occur. Too high a precipitant or protein concentration will push the protein to the irreversible precipitation zone. – (Figure adapted from Chayen N. (2004) [128]).

Seeding was successfully utilised for PatF crystallisation. The first round of seeding into new vapour diffusion drops resulted in a decrease in crystal numbers and an increase in crystal volume, however the crystals were still of poor, non-singular quality (Figure 2.8 B). A second round of seeding, using new crystals from the first round as the seed stock, was carried out with the protein concentration lowered to 4 mg ml^{-1}. Single three-dimensional crystals formed which diffracted to 3.0 Å in-house (Figure 2.8 C).

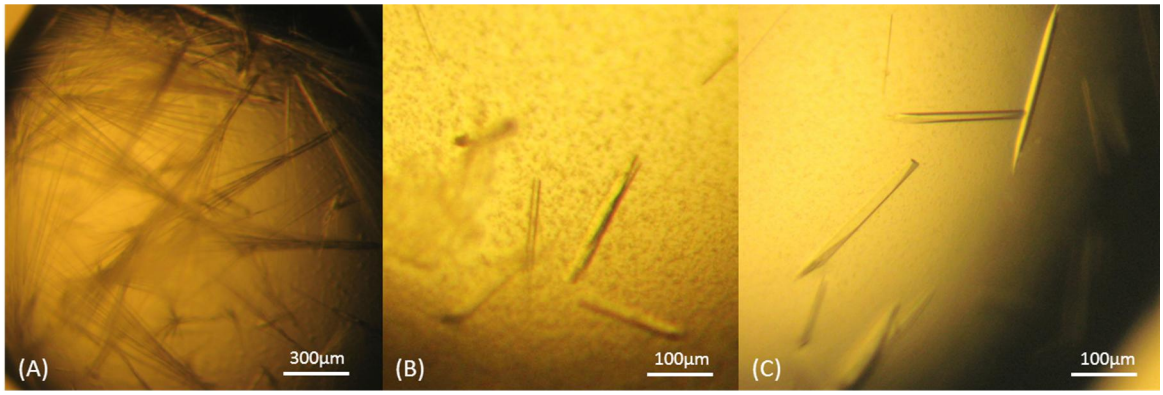


Figure 2.8: Crystals of PatF from Seeding Experiments. (A) Initial needle like crystals of PatF grown from a condition of 0.1 M sodium / potassium tartrate, 26 % PEG 2K MME and used in seeding trials. (B) First round of seeding experiments with crystals of more volume but still not singular. (C) Second round of seeding experiments using crystals from first round and reducing protein concentration to 4 mg ml^{-1}. These single crystals diffract to ~3.0 Å resolution in house.

2.3.4 Selenomethionine-labelled PatF

PatF has no sequence homology greater than 20 % to any other structures currently in the Protein Data Bank (PDB) [126]. The PatF family has been identified as prenyl transferases however attempts to determine the structure by molecular replacement (MR) against known prenyl transferase structures were not successful. It therefore was necessary to determine the phases experimentally [129].

In order to determine phases, heavy atoms are typically incorporated into the protein. Heavy atoms absorb X-rays and then re-emit them at a different phase and wavelength, a process called anomalous dispersion. Friedel's Law states that the intensity of the reflection hkl is equal to that of the reflection $-h$-k-l. (Figure 2.9 A) [129] [130] [131]. Under anomalous dispersion, Friedel's law is broken and the difference in intensities between the Friedel pairs can be used to locate the heavy

atom within the asymmetric unit (Figure 2.9 B). The phases from the heavy atoms can then be used to determine the structure of the protein.

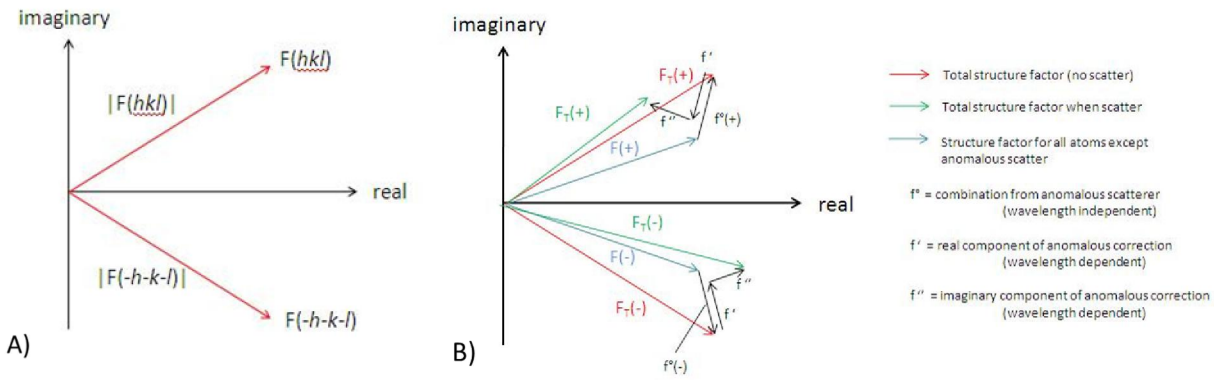


Figure 2.9: Friedel's Law. Schematic diagrams of (A) Structure factors obeying Friedel's law (B) Structure factors when a heavy atom is present in under both scattering and non-scattering wavelengths (Adapted from notes taken directly from Birkbeck University course).

The most efficient and reproducible method of incorporating heavy atoms into the protein is to replace methionine residues with selenomethionines, a derivative containing selenium in place of the sulfur. To achieve this, (L)-methionine is replaced by (L)-selenomethionine in the *E. coli* growth medium resulting in selenomethionine being directly incorporated into the protein during expression [132].

PatF contains four methionine residues, one of which is the first residue that is cleaved off during reaction with TEV protease.

An alternative approach to incorporating heavy atoms is to soak native crystals in solutions of heavy metals [133], however attempts at achieving this with samarium acetate, potassium osmium oxide, and mercury acetate soaks resulted in the deterioration of the crystals.

Purification of SeMet PatF occurs as with the native protein. A final yield of 2 mg L^{-1} cell culture with final purity greater than 98 % was achieved (Figure 2.10 A, B). MS

analysis confirmed that the final purified protein was fully labelled with three selenomethionine residues as expected (Figure 2.10 C).

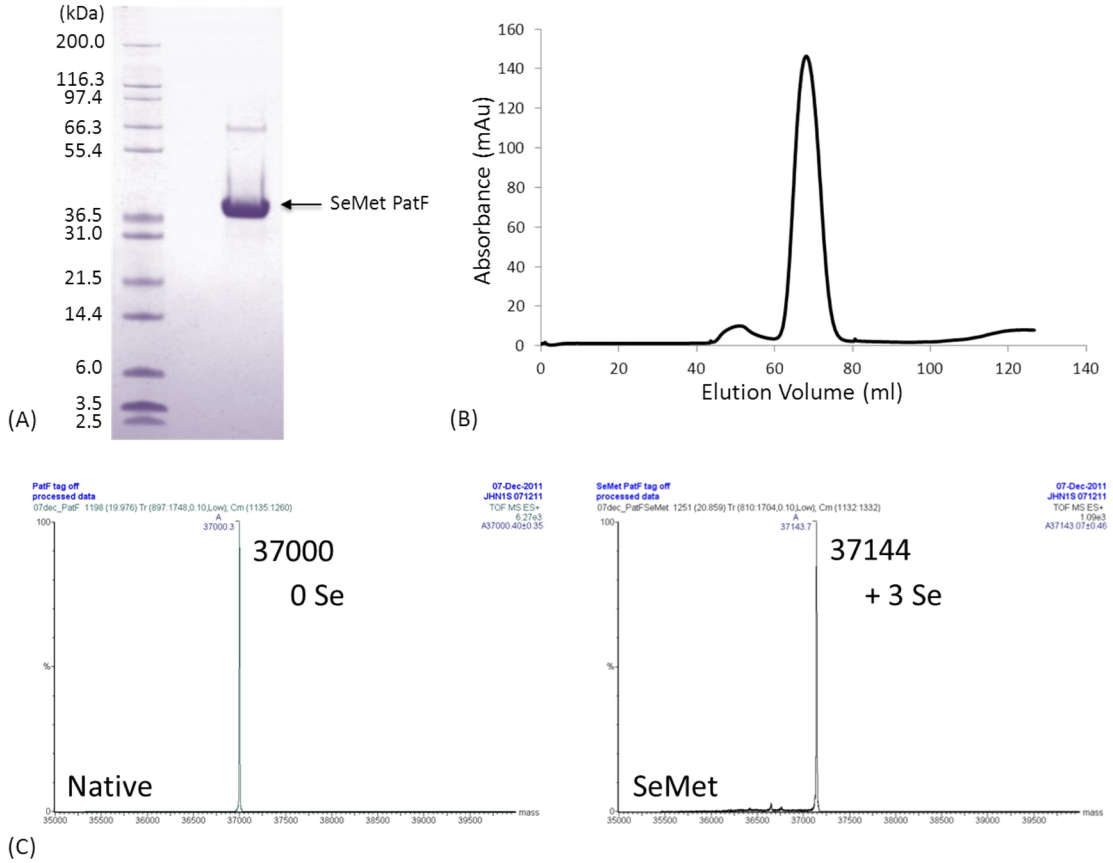


Figure 2.10: Purification of SeMet PatF. (A) SDS-PAGE of final SeMet PatF sample indicating a purity of > 98 % (Minor dimer band due to reducing agent boiling off during sample preparation.) (B) Gel filtration chromatograph for SeMet PatF showing protein eluted a single monomer peak. (C) MS analysis of native and SeMet PatF gives masses of 37000 and 37144 respectively. This corresponds to the incorporation of the expected three selenomethionine residues (100 % Se incorporation).

SeMet PatF was applied to crystallisation trials under the same conditions as for the native protein following tag removal using the micro-seeding method. A single crystal diffracted to 3.2 Å in-house, which was improved to 2.13 Å under synchrotron radiation (Figure 2.11). To achieve anomalous scattering, the crystal was scanned by

fluorescence to assess the optimum wavelength (Figure 2.12) [129]. A dataset was then collected at the selenium absorption edge (wavelength of 0.9790 Å).

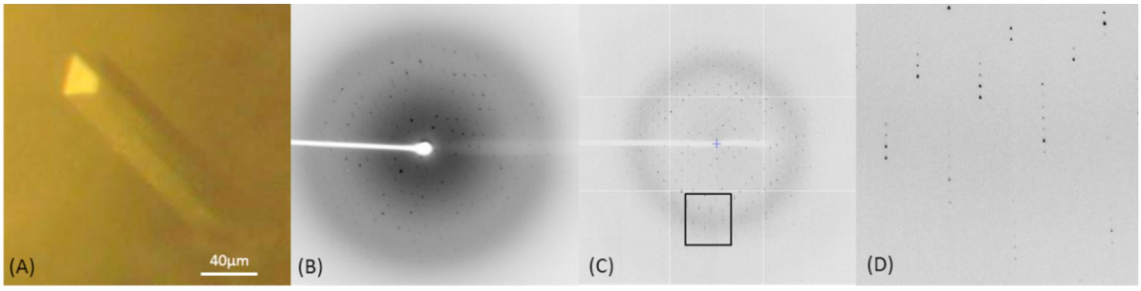


Figure 2.11: Crystallisation of SeMet PatF. (A) The best quality crystals of SeMet PatF were grown in a condition of 0.1 M sodium / potassium tartrate, 26 % PEG 2K MME following seeding. (B) In-house diffraction pattern of SeMet PatF to 3.2 Å. (C) Synchrotron diffraction pattern of SeMet PatF to 2.1 Å (D) Zoom view of synchrotron diffraction

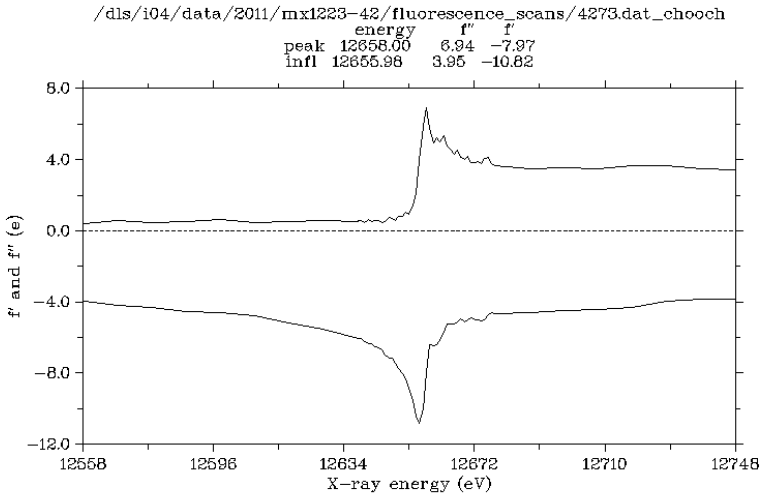


Figure 2.12: SeMet PatF Fluorescence Scan. Fluorescence scan of SeMet PatF crystal collected at beamline I04 (Diamond Light Source) showing the inflection and peak wavelengths.

The data was processed to 2.13 Å in the space group $P2_1$. Matthew's analysis [134] suggested that the asymmetric unit contained two molecules. The processed dataset was analysed first with SHELXC showing a relatively weak anomalous signal overall but with a reasonable signal to approximately 6.5 Å (Figure 2.13 A). The SHELXC output was then processed by SHELXD to a resolution of 6.5 Å which suggested that four

heavy atom sites were present in the asymmetric unit (Figure 2.13 B). The correlation histogram of SHELXD shows a single peak at 39 suggested a solution had been found (Figure 2.13 C). The output of SHELXD was finally carried into SHELXE, however SHELXE failed to find a reasonable solution, i.e. the native and heavy atom do not diverge as would be expected (Figure 2.13 D).

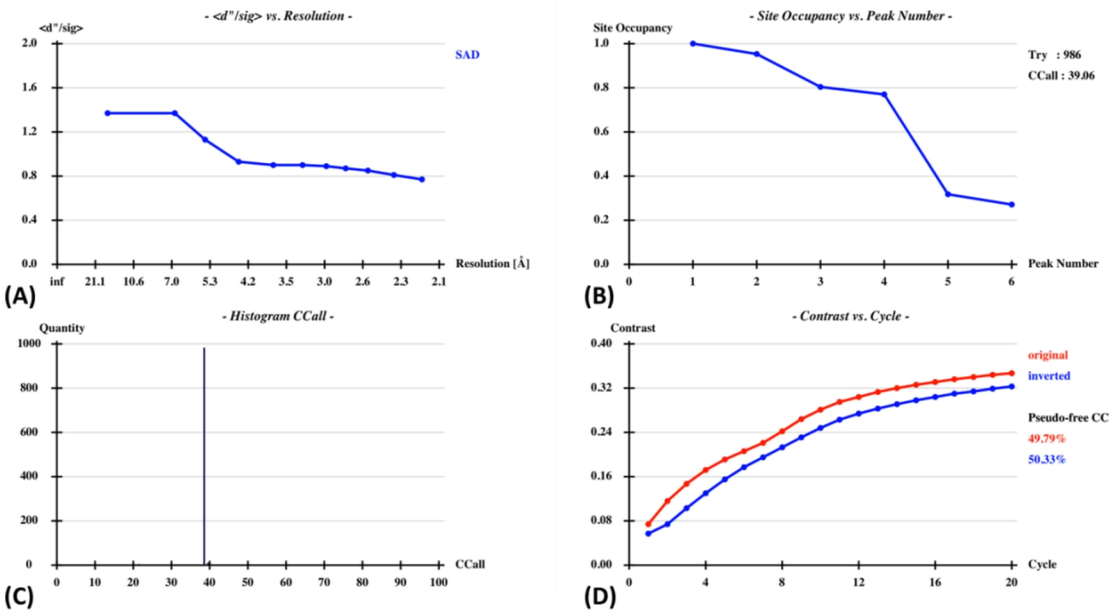


Figure 2.13: SHELX Output Statistics. A) Anomalous signal against resolution for SHELXC with signal strong to ~ 6.5 Å. B) Site occupancy against peak number for SHELXD location of the 4 Se atoms from SAD data C) CC histogram for the selenium-SAD phasing with single peak at 39 and D) Variation of the contrast during SHELXE density modification with a lack of divergence indicating no defined solution (Figures created using HKL2MAP [135]).

As the SHELXC and SHELXD outputs strongly proposed that a solution would be possible, it was decided to attempt to find a solution using PHENIX AutoSol. In agreement with SHELXD, the heavy atom search function of PHENIX AutoSol found four atoms of good occupancy > 0.88. However, unlike SHELXE, a structure solution was found from the heavy atom phasing giving a figure of merit of 0.262 and overall score 24.22 +/- 14.6. As there are three selenomethionines present in the protein, and Matthew's analysis suggests two monomers in the AU, it was predicted that one

selenomethionine in each monomer lies on a disordered part of the structure and could be discarded from the structure solution.

A round of density modification gave an R-work of 0.2827 and visual analysis of the map indicated that the correct solution had been found with secondary structure elements clearly visible in the density (Figure 2.14).

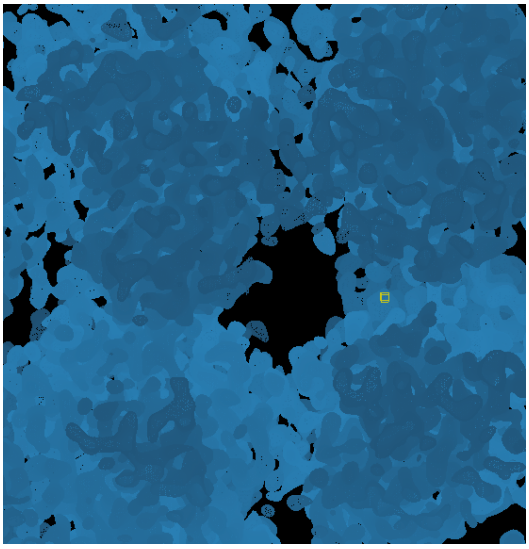


Figure 2.14: Density Modification Map. Density modification output for PatF from PHENIX AutoSol showing clear crystal packing with each macromolecule having elements of secondary structure strongly suggesting a solution had been found for PatF structure. Figure created using Coot [116].

The solution was then processed by PHENIX AutoBuild in order to assign the structure to the electron density. Eight peptide fragments incorporating a total of 531 residues were built into the density across the two chains. A final R-work of 0.2305 and R-free of 0.2813 were observed following automated chain building.

The PHENIX AutoBuild output was manually rebuilt in Coot with significant rebuilding required in the loop regions and the termini of each chain. Each round of rebuilding was refined using Refmac. TLS restraints were calculated using the TLSMD server and used in refinement. All angles were checked by Ramachandran plots in Coot, and poor rotamers analysed on the MolProbity server. These outliers were manually corrected.

The final data collection and refinement statistics are presented in Table 2.4. The final structure was analysed using the MolProbity server [124] with the output statistics presented in Table 2.5.

PatF SAD dataset	
Wavelength (Å)	0.9790
Space group	P 1 2_1 1
a, b, c (Å)	47.1, 135.8, 48.7
α, β, γ (°)	90, 118.5, 90
Resolution (Å)	35.32 - 2.13 (2.19-2.13)
I/σI	14.0 (3.5)
R_{merge} (%)	10.1 (69.3)
Completeness (%)	97.9 (97.5)
Multiplicity	9.5 (9.8)
Anomalous completeness	97.9 (97.6)
Anomalous multiplicity	4.8 (4.9)
Refinement	
R-factor	0.1899
R_{free}	0.2210
R.m.s.d	
Bond Lengths (Å)	0.009
Angles (°)	1.314
B-factors	
All	41.7
Protein	41.6
Water	40.0

Table 2.4: Data Collection and Refinement Statistics for SeMet PatF. An anomalous data set was collected using a single crystal on beamline I04 (Diamond Light Source). Statistics are presented as averages with values for the highest resolution shell included in parentheses.

All-Atom Contacts	Clashscore, all atoms:	3.66		99th percentile* (N=553, 2.13Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	8	1.50%	Goal: <1%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	580	98.81%	Goal: >98%
	MolProbity score^	1.29		100th percentile* (N=11293, 2.13Å ± 0.25Å)
	Cβ deviations >0.25Å	0	0.00%	Goal: 0
	Bad backbone bonds:	0 / 2376	0.00%	Goal: 0%
	Bad backbone angles:	0 / 2963	0.00%	Goal: <0.1%

Table 2.5: MolProbity Statistical Output for PatF Structure. The final coordinate file was processed in the MolProbity server and confirmed as a structure in the 100th percentile in terms of quality when compared to structures of similar resolution [124].

2.3.5 Crystal Structure of PatF

The structure of PatF has been determined by single-wavelength anomalous dispersion (SAD) to 2.13 Å. The structure contains two molecules in the asymmetric unit. PatF is formed of a twelve-stranded beta barrel pore surrounded on the outside by twelve alpha helices in a similar but not identical conformation as the TIM barrel motif [136] (Figure 2.15 A,B). The refined model contains residues 3-196 and 205-307 in chain A, and 3-197 and 205-307 in chain B. The missing residues are the connecting loop between beta strand 8 and alpha helix 8 and the N- and C- termini, all of which are presumed to be disordered. A single disulfide bridge is present at the C-terminus linking residues Cys276 and Cys307. PatF is a globular protein of dimensions 45 x 43 x 53 Å.

The electrostatic potential of PatF (Figure 2.15 C) shows that the pore region, where the putative binding site is located, is strongly electronegative indicating that accepting the proposed DMAPP substrate, which itself is negatively charged due to the phosphate group, will be challenging.

The native dataset of PatF with His_6-tag remaining was determined by molecular replacement (MR) using PHASER [137] with the SeMet PatF structure as the model. The native structure contains one monomer in the asymmetric unit. The resolution difference of 2.13 Å (SeMet) to 2.08 Å (native) is negligible and indeed the native structure has more disordered residues (ordered residues 3-195, 204-305). In addition, there was no evidence of the His_6-tag in the electron density.

Crystals of native PatF with the tag removed diffract to 2.5 Å at best and as a result, the SeMet PatF structure was used in subsequent homology modelling and structural alignments.

2.3.6 Homology Models

Homology models of LynF and TruF1 have been built using Phyre2 [126] by inputting their respective amino acid sequences along with coordinates of the PatF structure. The models appear to show similar overall structure in the form of a beta barrel surrounded by alpha helices (Figure 2.16 A). The significant difference when comparing these models to PatF is in the electrostatic potential. PatF shows large patches of electronegativity through the pore of the structure (Figure 2.15 C), the presumed active site, which makes the likelihood of the negatively charged DMAPP binding particularly low. LynF and TruF1 however do not show large numbers of acidic residues in the pore (Figure 2.16 B, C).

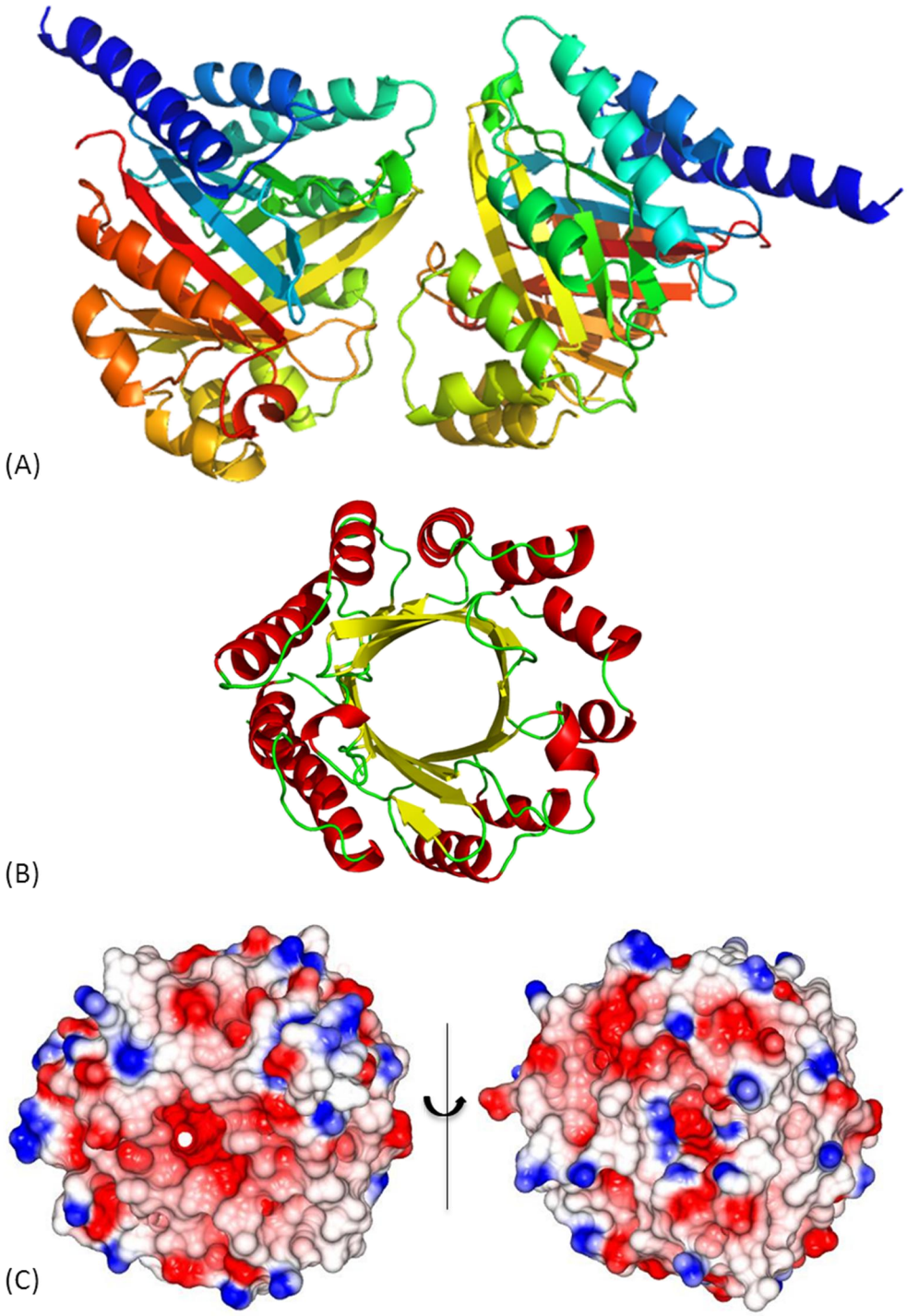


Figure 2.15: The Crystal Structure of PatF. (A) Asymmetric unit of PatF featuring two monomers. (B) Secondary structure analysis of PatF. The monomer consists of twelve β-strands arranged into a β-barrel (yellow) and surrounded on the outside by twelve α-helices (red). (C) Electrostatic surface map representation of PatF rotated around 180°. (Red-negative, blue-positive). Electrostatic surface figures which were produced using CCP4MG [138] [139].

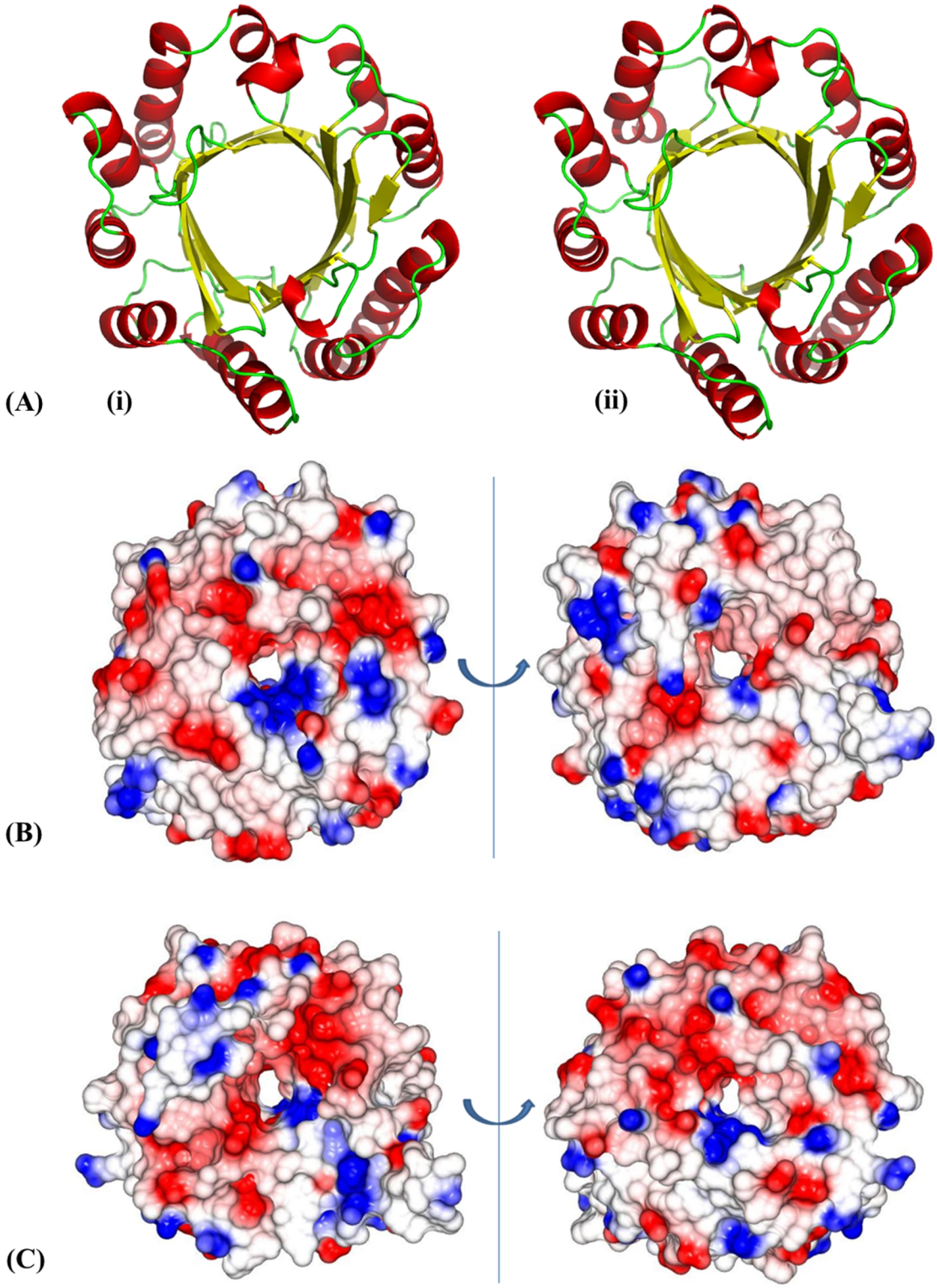


Figure 2.16: Homology Models of LynF and TruF1. (A) Homology models of (i) LynF and (ii) TruF1 both showing a beta barrel (yellow) surrounded by alpha helices (red). (B) Electrostatic potential model of LynF rotated around 180°. (C) Electrostatic potential model of TruF1 rotated around 180°.

2.3.7 Structural Alignments

To further characterise the PatF structure, structural alignments with related proteins were carried out. PatF is presumed to be a prenyl transferase so the structure of dimethylallyl tryptophan synthase (DMATS) from *Aspergillus fumigatus* in complex with an inactive DMAPP mimic and tryptophan [140] was downloaded from the Protein Data Bank (PDB code: 3I4X, Figure 2.17 A). In addition, the structure of PagF (currently unpublished), a member of the PatF family from *Planktothrix agardhii*, which is known to prenylate tyrosine residues was kindly provided by Prof. S. Nair (University of Illinois) (Figure 2.17 B).

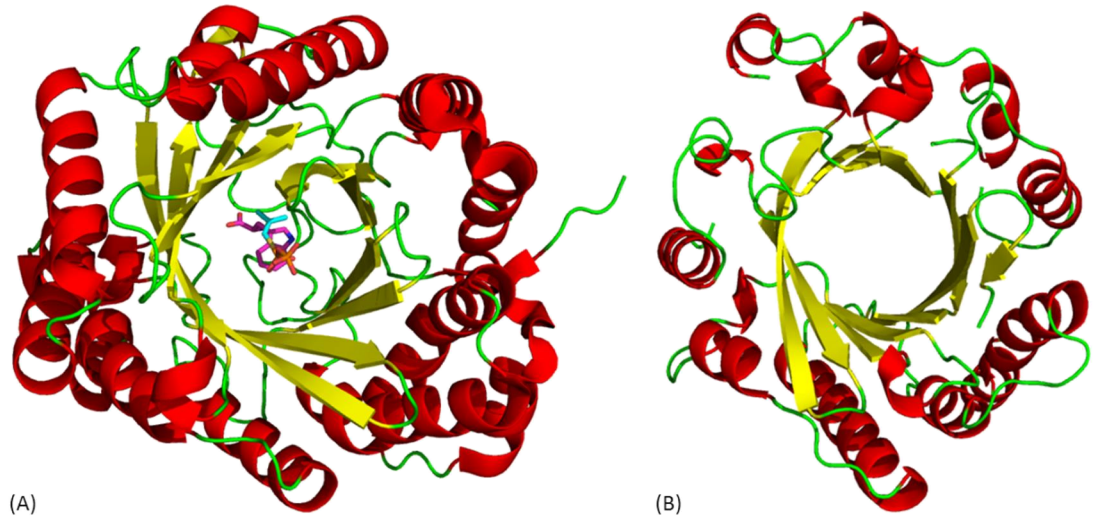


Figure 2.17: Structures of Prenyl Transferases – Secondary structure representations of (A) Dimethylallyl tryptophan synthase from *Aspergillus fumigatus* in complex with tryptophan and a DMAPP derivative (B) PagF, a cyanobactin prenyl transferase from *Planktothrix agardhii* known to prenylate tyrosine residues.

DMATS is considerably different from PatF and PagF in terms of sequence homology (< 5 %) however all three contain the characteristic beta barrel pore with surrounding alpha helices. The presence of the DMAPP mimic in the pore of DMATS allows this structure to be used as a starting model for structural alignments to identify the potential residues involved in PatF/PagF DMAPP binding.

The structures of PatF, PagF and the DMATS complex were aligned in Coot [116] using the SSM superpose tool.

Two key interactions in DMAPP mimic binding which appear conserved in PagF and DMATS are significantly different in PatF. Met136 of PatF is found to be lysine residues in both PagF and DMATS which forms a strong salt bridge with one of the phosphate oxygens of the DMAPP mimic (Figure 2.18). His125 of PatF is an aspartic acid residue in the other structures. The Asp residue forms two salt bridges, one to the lysine mentioned previously and one to an arginine (Arg66 in PatF) stabilising them to form two further salt bridge interactions with the DMAPP mimic (Figure 2.19)

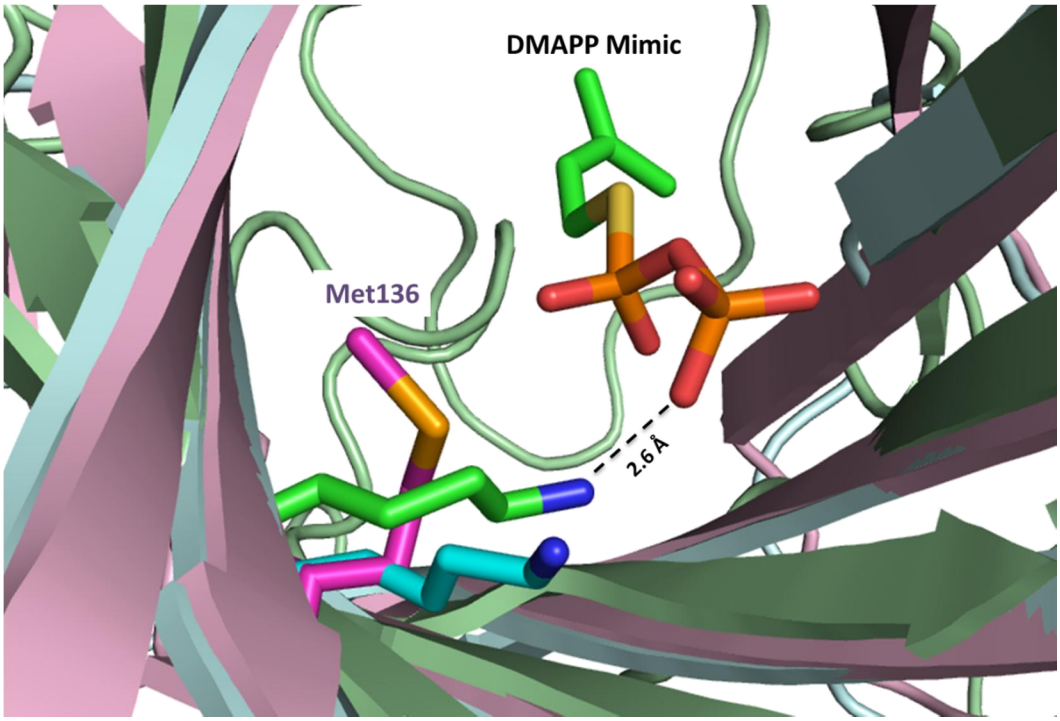


Figure 2.18: Structural Alignment of PatF Met136. The structural alignment of PatF (magenta), PagF (cyan) and DMATS (green) showing the difference of Met136 to Lys in both PagF and DMATS. The salt bridge interaction of the Lys in DMATS to the DMAPP mimic is shown by dashed line.

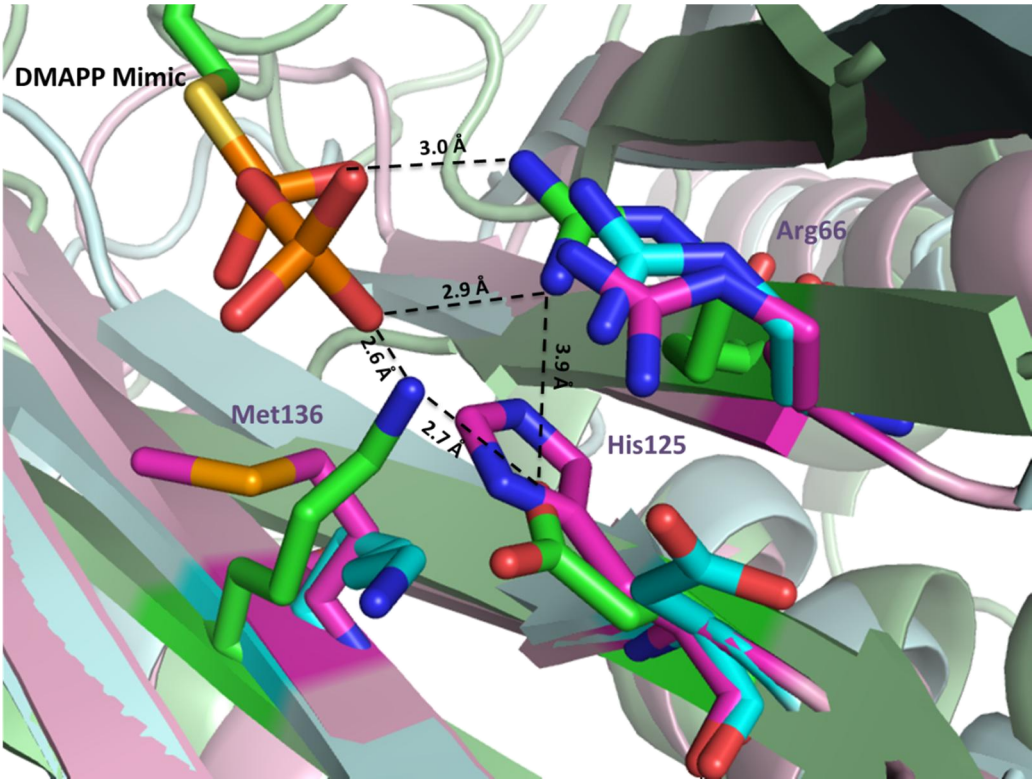


Figure 2.19: Structural Alignment of PatF His125. The structural alignment of PatF (magenta), PagF (cyan) and DMATS (green) highlighting the difference of His125 to Asp in both PagF and DMATS. The stabilising salt bridge interactions in the DMATS structure (dashed lines) are shown to the Arg66 equivalent and the Lys (Met136 in PatF), which in turn form salt bridges with the DMAPP mimic.

2.3.8 LynF Expression and Purification

LynF in pJexpress 411 plasmid was overexpressed with an N-terminal His_6-tag in BL21 (DE3) *E. coli* cells using the Studier auto-induction method [107]. The soluble protein was isolated on a Ni-sepharose column. TEV protease was added to the protein to cleave the His_6-tag and the tag was removed by passage over a second Ni-sepharose column (Figure 2.20). At this stage protein levels were less than 0.2 mg L^{-1} of original cell culture and the protein was precipitating significantly. As a result of these issues, this LynF construct was discontinued.

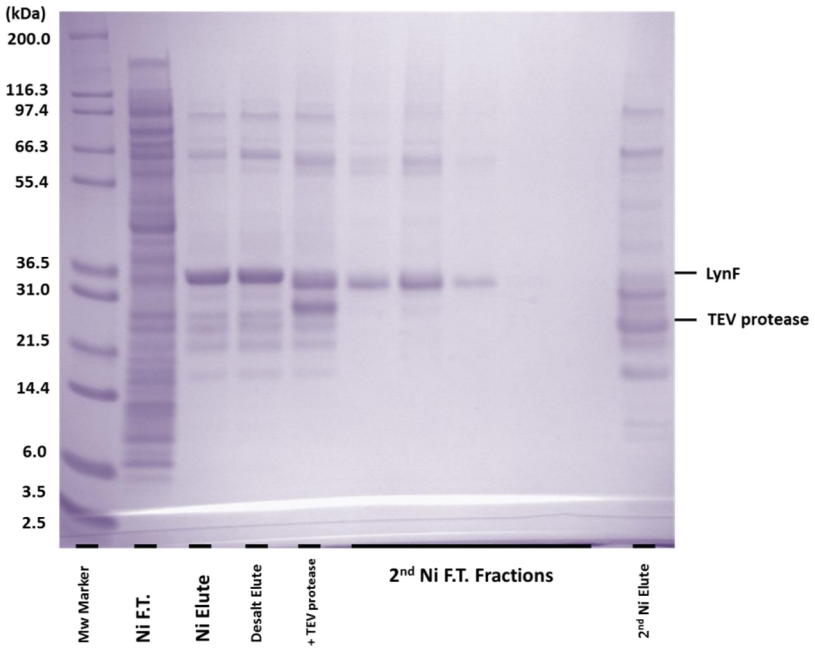


Figure 2.20: LynF Purification (pJexpress 411). (A) SDS-PAGE analysis of LynF purification with Ni-sepharose isolation and TEV protease cleavage steps.

To improve solubility, LynF was re-cloned into the pSUMO plasmid for co-expression with a SUMO domain as a solubility tag. The His_6/SUMO tag was removed with TEV protease with ~95% efficiency. Two significant peaks eluted off the gel filtration column, a highly aggregated peak and a soluble monomer peak. SDS-PAGE analysis confirmed the soluble monomer peak to be pure LynF with the aggregation peak containing a small amount of LynF and significant contaminant proteins (Figure 2.21). The LynF protein was confirmed by MS and a final purified yield of 1.5 mg L^{-1} culture was achieved.

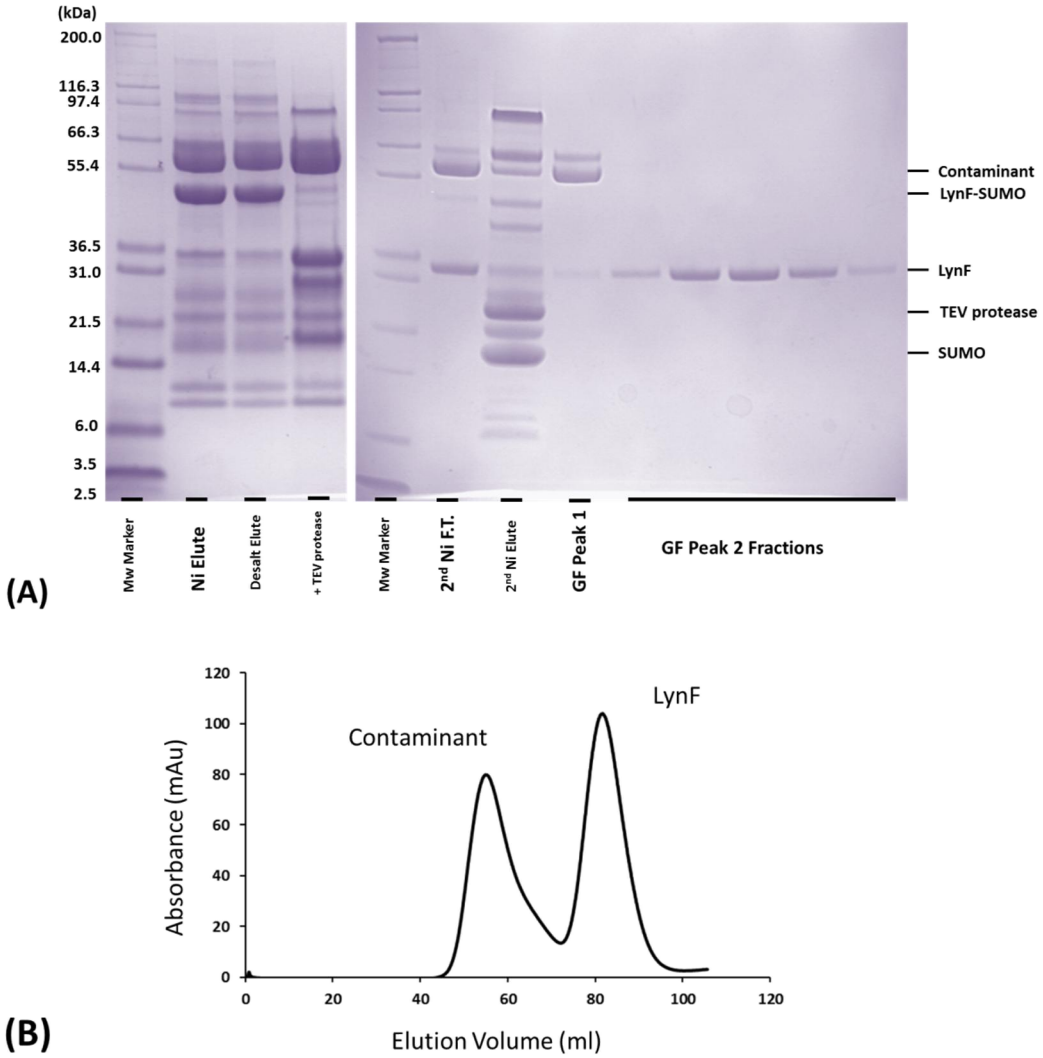


Figure 2.21: LynF Purification (pSUMO). (A) SDS-PAGE analysis of LynF purification from initial Ni-sepharose isolation, TEV protease cleavage to remove His_6/SUMO tag and final gel filtration step. Fractions of peak two contain >98% pure LynF. (B) Gel filtration chromatograph for LynF showing protein eluted as a monomer peak (Peak2) with the contaminants eluting in the void volume (Peak 1).

2.3.9 Biochemical Studies on PatF and LynF

Biochemical reactions of PatF were set up with a range of amino acid derivatives and DMAPP. The reactions were analysed by LC-ESI MS. For the three amino acid derivatives assessed, Boc-Tyr, Boc-Ser and Boc-Trp, no evidence of prenylation was

found when incubated overnight or for 72 hours. All MS spectra showed starting material only.

As a control, a reaction of LynF with Boc-Tyr and DMAPP was set up, repeating the experiment detailed in McIntosh *et al.* (2011) [65]. A peak with m/z = 372.08 corresponding to mass plus sodium of prenylated Boc-Tyr was observed. Examination of the LC peak would suggest that overnight incubation with 10µM LynF prenylated greater than 50 % of the Boc-Tyr. This can be observed by the shift in elution time due to the large hydrophobic group added to the molecule (Figure 2.22)

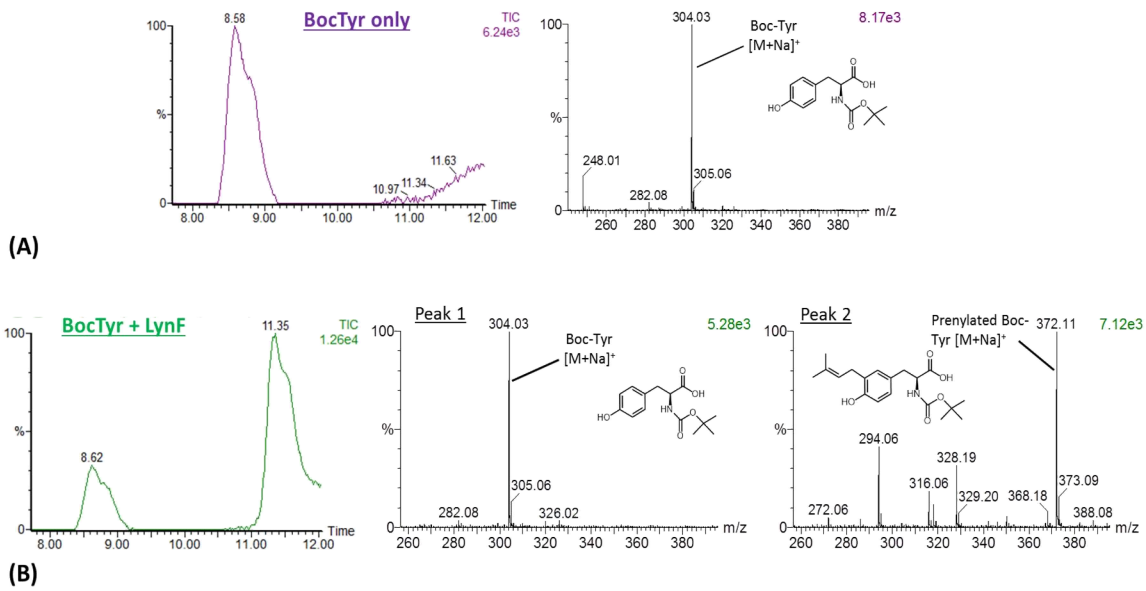


Figure 2.22: LC-ESI MS Spectrum for BocTyr Reactions. (A) Boc-Tyr in reaction conditions without enzyme present. Elution at 8.58 min with $[M+Na]^+$ peak for Boc-Tyr present. (B) Reaction of Boc-Tyr + LynF showing two peaks at 8.62 min and 11.35 min with $[M+Na]^+$ peak for Boc-Tyr and $[M+Na]^+$ peak for prenylated Boc-Tyr respectively.

2.3.10 Active Site Mutants of PatF

PatF mutants M136K and H125D/G127R/M136K were created in order to attempt the introduction of an active site based on the structural and sequence data obtained. Unfortunately, these constructs resulted only in the expression of insoluble protein. This was confirmed by running both the insoluble and soluble fractions of the lysed cells on SDS-PAGE (Figure 2.23).

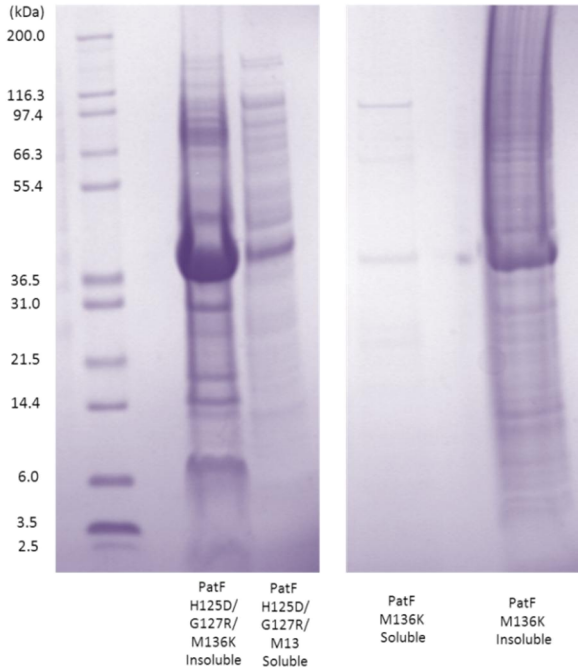


Figure 2.23: SDS-PAGE of PatF Mutants. SDS PAGE analysis of the soluble and insoluble fractions of PatF mutants M136K and H125D/G127R/M136K. The largest bands in the insoluble fraction correspond to the mutated PatF. There are no or minor bands corresponding to soluble protein for both mutants. (Note: The insoluble fraction was suspended in 8M urea to enable SDS-PAGE run).

2.4 Discussion

The PatF crystal structure has been determined to 2.1 Å and confirms it to possess a central beta barrel surrounded by alpha helices in a TIM barrel-like conformation.

Electrostatic potential maps of the PatF structure reveal that the central pore, where the putative binding site is located, is highly electronegative. This suggests that binding of the electronegative DMAPP substrate is unlikely given the potential repulsion interactions which would occur.

Structural alignments of PatF with two other known prenyl transferases, DMATS a tryptophan prenyl transferase in complex with a DMAPP mimic from an unrelated pathway and PagF, a related cyanobactin tyrosine prenyl transferase, has provided us with information on the key interactions involved in DMAPP binding. It appears that two interactions involved in DMATS and PagF are not conserved in PatF and their subsequent equivalents would have a detrimental effect on binding. His125 and Met136 of PatF are conserved as Asp and Lys residues, respectively, in both DMATS and PagF and both residues are involved in DMAPP binding. Interestingly, sequence alignments of the PatF family (Figure 2.24, Appendix B) show that both His125 and Met136 are fully conserved as aspartic acid and lysine/arginine residues respectively in all other PatF family members. Additionally, residue Gly127 is fully conserved as an arginine in all PatF related proteins; however this residue is not located at the binding site but is found at the pore opening. It is possible that this could have an effect on DMAPP entry into the protein although this has not been confirmed.

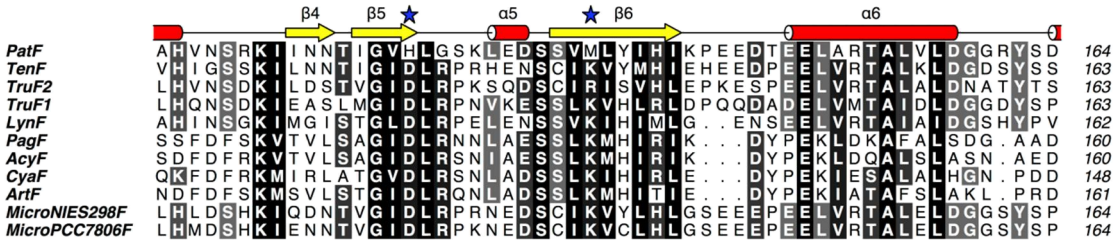


Figure 2.24: Partial Sequence Alignment of PatF. Sequence alignment of PatF with related family members from residues 98 to 146. The key residue changes, H125D and M136K are starred. Secondary structure elements of PatF are also shown above (alpha helices in red, beta sheet in yellow). The full sequence alignment can be found in Appendix B.

Biochemical studies of PatF to date have yielded no prenylated product when substrates of DMAPP and amino acid derivatives Boc-Tyr, Boc-Ser and Boc-Trp have been used. This is most likely due to a lack of binding as discussed previously. The demonstration of activity when using a known active cyanobactin prenyl transferase, LynF, further supports this hypothesis.

Mutation of the PatF active site to align with all other PatF family members resulted in only insoluble protein suggesting that these mutations cause disruptive interactions within the protein.

These studies on PatF have given a strong indication that the protein has become evolutionary inactive with respect to prenyl transferase activity, however this is always difficult to prove conclusively. One final question remains, if PatF is indeed an inactive prenyl transferase, why do PatF gene knockouts result in no patellamide products? Perhaps it has an unknown function which still needs to be investigated.

2.5 Conclusions and Future Work

The crystal structure of PatF has been determined to 2.1 Å resolution. PatF is formed of a twelve strand beta barrel motif surrounded by twelve alpha helices. The function of PatF is still not clear and indeed whether it still retains function despite the protein containing structural characteristics of a prenyl transferase. The structure has shown that the lack of conservation of two and possibly three amino acids with respect to other members of the PatF family could significantly inhibit DMAPP binding. This appears to be confirmed by a lack of activity in biochemical assays. The mutation of PatF to attempt to reintroduce DMAPP binding results in only insoluble protein. PatF has been designated as essential in previous studies, therefore further study will be required to assess what this role may be. LynF has been purified and shown to be active in biochemical assays. Future work will look to examine LynF (and potentially other homologues) in crystallisation trials in the presence of DMAPP (or its mimic) and tyrosine analogues to determine their crystal structures and give an insight into the mechanism of cyanobactin prenyl transferases.

3. Structural and Biochemical Studies of PatGmac

The work in this chapter was a collaboration between Dr Jesko Koehnke (University of St Andrews) and myself. The apo protein structure was determined by Dr Koehnke whilst the protein-peptide complex structure and the mutant activity assays were performed by myself. Conclusions on the enzyme mechanism were a joint contribution from us both.

3.1 Introduction

PatG consists of three domains; an N-terminal oxidase domain, protease / macrocyclase domain and a C-terminal domain of unknown function. The macrocyclase domain (PatGmac) catalyses one of the final steps in patellamide biosynthesis, the C-terminal cleavage and macrocyclisation of the eight amino acid core peptide into the final cyclic product. Structural and biochemical studies of this enzyme should give a greater insight into the mechanism of macrocyclisation and also identify the substrate tolerance of the enzyme.

Bioinformatic analysis of PatGmac identifies it to be a member of the subtilisin-like serine proteases [84], [85] containing the typical active site catalytic triad – Asp, His and Ser. Subtilisin-like proteases accept peptide substrates and the notation P4-P2' is used to donate the substrate residues, with peptide cleavage occurring between residues P1 and P1'.

The native PatGmac structure was determined to a resolution of 2.19 Å by Dr Jesko Koehnke and the structure has given an understanding into how its variation from other subtilisin-like proteases may result in macrocyclisation [141] (PDB code: 4AKS). Sequence alignments of the PatGmac family when compared to other subtilisin-like proteases show a significant insertion of between 35 and 50 amino acids (37 amino acids for PatGmac, Figure 3.1, Appendix C). This insertion in the PatGmac structure has

been defined as the “macrocyclisation insertion” and consists of a helix-turn-helix motif which sits directly over the active site (Figure 3.2 A, B). The removal of this insertion abolishes macrocyclisation activity resulting only in linear proteolysis [141]. In addition, compared to standard subtilisin-like proteases e.g. Bacillus Ak.1 protease [142], the ability of the substrate to form an extended conformation upon binding is hindered by conformational changes in the protein which results in three residues, Met660, Arg686 and Phe684, blocking the traditional P3 and P4 sites. (Figure 3.2 C).

```
PatGmac  LHPHDLDERIPGL.........ADLHNQTLGDPQITIVIIDGD  550
Ak1      WTPNDTY..YQGYQYGPQNTYTDYAWDVTKGSSGQEIAVIDTG   41

PatGmac  PDYTLSCFEGAEVSKVFPYWHEPAEPITPEDYAAFQSIRDQGL  593
Ak1      VDYTHPDLDGKVI.KGY.......................      57

PatGmac  KGKEKEEALEAVIPDTKDRIVLNDHACHVTSTIVG..QEHSPV  634
Ak1      ..........DFVDNDYDPMDLNNHGTHVAGIAAAETNNATGI   90

PatGmac  FGIAPNCRVINMPQDA.VIRGNYDDVMSPLNLARAIDLALELG  676
Ak1      AGMAPNTRILAVRALDRNGSGTLSD......IADAIIYAADSG  127

PatGmac  ANIIHCAFCRPTQTSEGEEILVQAIKKCQDNNVLIVSPTGNNS  719
Ak1      AEVINLSLGCDCHTT....TLENAVNYAWNKGSVVVAAAGNNG  166

PatGmac  NESWCLPAVLPGTLAVGAAKVDGTPCHFSNWGGNNTKEGILAP  762
Ak1      SSTTFEPASYENVIAVGAVDQYDRLASFSNYGT...WVDVVAP  206

PatGmac  GEEILGAQPCTEEPVRLTGTSMAAPVMTGISALLMSLQVQQGK  805
Ak1      GVDIVSTIT.GNRYAYMSGTSMASPHVAGLAALLASQGRNNIE  248

PatGmac  PVDAEAVRTALLKTAIPCDPEVVEEPERCLRGFVNIPGAMKVL  848
Ak1      ......IRQAIEQTA.....DKISGTGTY.FKYGRINSYNAVT  279

PatGmac  FGQ  851
Ak1      Y..  280
```

Figure 3.1: Sequence Alignment of PatGmac and Ak1. Sequence alignment of the macrocyclase domain of PatG (PatGmac) and the subtilisin-like protease Ak1. Corresponding residues are group in black. The large “macrocyclisation insertion” is coloured in cyan.

To further enable the understanding of macrocyclase activity and to identify the key interactions required, a crystal structure of PatGmac in complex with a substrate peptide followed by site specific mutations and subsequent activity assays were targeted.

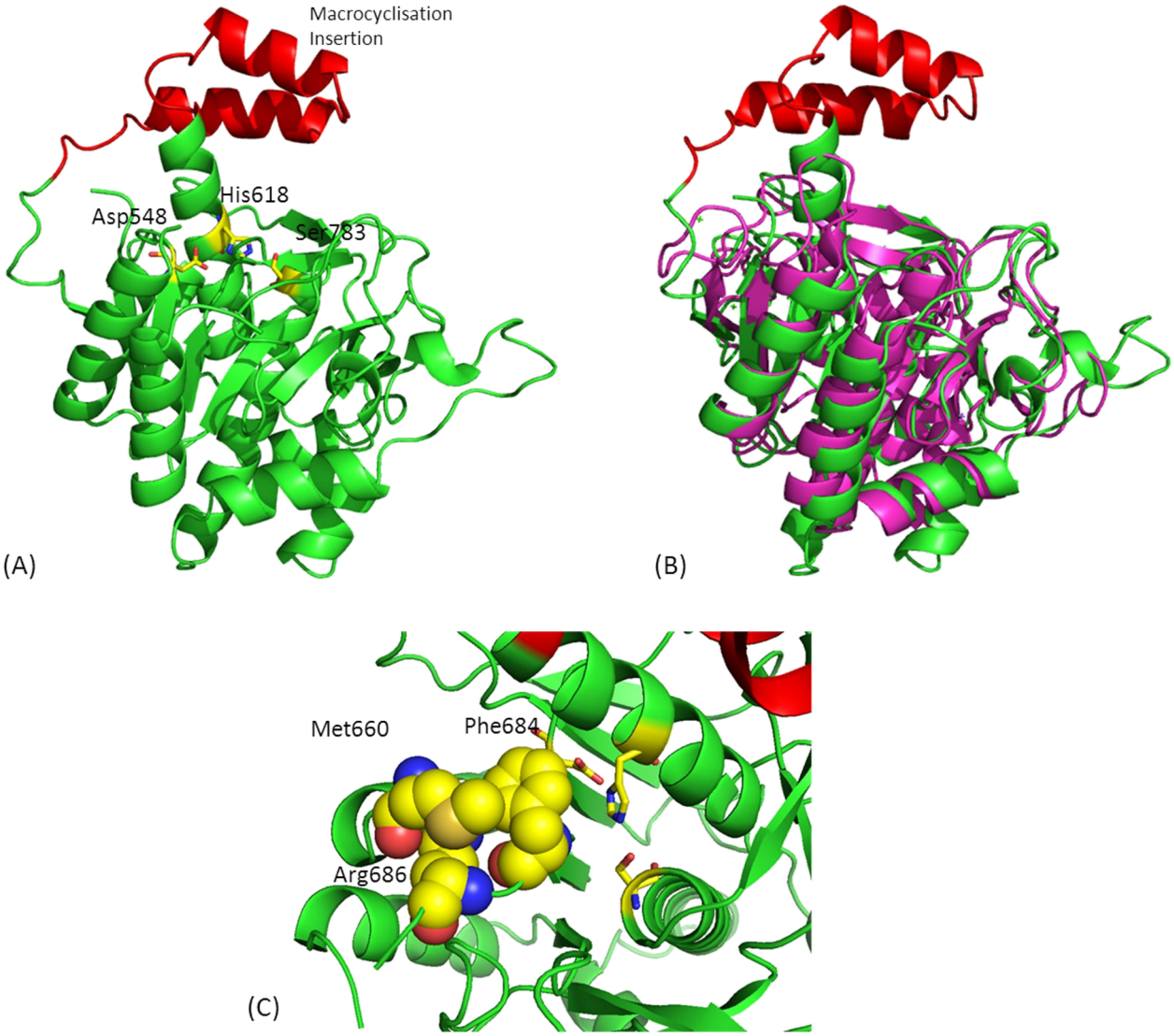


Figure 3.2: The Structure of PatGmac. (A) The crystal structure of PatGmac with the active site residues Asp548, His618 and Ser783 highlighted as sticks and the macrocyclisation insertion coloured in red. (B) The structure of PatGmac (green/red) overlaid with the structure of Ak1, a subtilisin-like protease (magenta) clearly showing the macrocyclisation insertion in PatGmac. (PDB codes: PatGmac: 4AKS Ak1: 1DBI) (C) View of the active site of PatGmac showing the three residues which block the extended conformation, Met660, Arg686 and Phe684 (spheres) and the catalytic triad (sticks).

3.2 Materials and Methods

3.2.1 DNA Cloning

Full length PatG was originally cloned from genomic DNA (*Prochloron sp*.) into the pHISTEV vector by Dr Wael Houssen (University of Aberdeen). The macrocyclase domain 'PatGmac' (PatG residues 492-851) along with its mutants PatGmac H618A, PatGmac K594D, PatGmac K598D, and PatGmac R589D,K594D,K598D were sub-cloned by Dr Houssen and Dr Jesko Koehnke.

3.2.2 Expression and Purification

The inactive mutant PatGmac H618A (PatG residues 492-851) was expressed from the pHISTEV plasmid (Liu *et al.* [143]) with an N-terminal His_6-tag and Tobacco Etch Virus (TEV) protease recognition site using BL21 (DE3) *E. coli* cells grown on auto-induction medium using the Studier method [107] (see Appendix A.3 for media and buffer compositions). The cultures were grown at 20 °C, 250 rpm for 48 hours before harvesting by centrifugation.

Cell pellets of PatGmac H618A were re-suspended in lysis buffer supplemented with DNAse at 0.4 mg g^{-1} wet cell pellet. The re-suspension was lysed by passage through a cell disruptor at 30 kPsi (Constant Systems). The lysate was cleared by centrifugation (40,000 x g, 4 °C, 20 min) and then loaded onto a Ni-Sepharose 6 FF column (GE Healthcare) equilibrated in lysis buffer. The column was washed with lysis buffer and PatGmac H618A eluted with elution buffer. The protein solution was then passed over a desalt column (Desalt 16/10, GE Healthcare) in desalting buffer. TEV protease was added at a mass-to-mass ratio of 1:10 TEV protease : protein and incubated for one hour at 20 °C to remove the His_6-tag. The sample was then loaded on to a second Ni-column, in desalt buffer. PatGmac H618A was collected in the flow through and loaded onto a monoQ column (GE healthcare) equilibrated in Q buffer A. The protein was eluted through a NaCl gradient (Q buffer B) eluting at 350 mM NaCl. The protein was

concentrated to 7.5 ml (Vivaspin concentrators, 10 kDa MWCO) and applied to a Superdex 75 gel filtration column (GE Healthcare) equilibrated in sizing buffer and concentrated to 30 mg ml^{-1} for crystallography. The purity of the protein was assessed by SDS-PAGE analysis and its identity confirmed by mass spectrometry (MS).

Native PatGmac and active site mutants PatGmac K594D, PatGmac K598D and PatGmac R589D K594D K598D were expressed and purified using the same methods as PatGmac H618A.

3.2.3 Crystallography

To obtain a complex structure, the inactive PatGmac H618A was used in crystallisation trials to ensure the substrate peptide wasn't processed. PatGmac H618A was subjected to stochastic crystallisation screening [108] at both 20 °C and 4°C followed by manual optimisation at 4 °C using the same methods as for PatF (see section 2.4.2).
Soaking trials with the substrate mimic were set up by incubating several apo crystals in mother liquor supplemented with 7.5 mM peptide VPAPIPFPAYDG (provided by Dr. Laurent Trembleau, University of Aberdeen) at 4 °C for 72 hours.
For co-crystallisation with the substrate mimic, PatGmac H618A was mixed with a 10 fold excess of peptide VPAPIPFPAYDG at 4 °C overnight prior to being used in crystal trials. The putative complex was then subjected to stochastic crystallisation screening and optimisation as above [108].

3.2.4 Data Collection and Structure Solution

The crystal of PatGmac H618A in complex with VPAPIPFPAYDG peptide was cryo-protected in a solution of mother liquor supplemented with 30 % glycerol and flash frozen in liquid nitrogen. A dataset on a single crystal was collected in-house at 100K on a Rigaku 007HFM rotating anode X-ray generator with a Saturn 944 CCD detector. The data were processed using xia2 [110] and the structure determined by molecular

replacement with PHASER [137] using the structure of PatGmac as a model (PDB: 4AKS). Manual rebuilding was performed with Coot [116] and refinement was carried out using REFMAC5 [117] as part of the CCP4 suite [118]. TLS restraints were calculated using the TLSMD [121], [122] server and used in refinement [123]. The structure was validated using MolProbity [124] and the coordinates deposited in the Protein Data Bank [126] (PDB Deposition ID: 4AKT).

3.2.5 Biochemical Studies

Macrocyclisation reactions of a substrate mimic with both native and active site mutants of PatGmac were used to assess ratios of non-cleaved, linearly cleaved and macrocyclised product.

100 µM peptide VGAGIGFPAYDG (provided by Dr. Laurent Trembleau, University of Aberdeen) , was incubated with 50 µM enzyme in 150 mM NaCl, 10 mM HEPES pH 8.0, 1 mM TCEP for 120 hours at 37 °C. All reactions were carried out at 100 µl scale with samples analysed by LC-ESI MS (Micromass LCT) using a Phenomenex C18 column run in H_2O/0.1% formic acid vs methanol/0.1% formic acid gradient.

3.3 Results

3.3.1 Expression and Purification

PatGmac H618A was overexpressed with an N-terminal His_6-tag and TEV protease recognition site in BL21 (DE3) *E. coli* cells using the Studier auto-induction method [107]. The protein was isolated using the His_6-tag on nickel resin and the tag cleaved at close to 100 % efficiency by the addition of TEV protease. The tag was removed by passage over a second nickel column. The protein was further purified on an ion-exchange mono-Q column with the protein eluting at 350 mM of a NaCl gradient. The sample was subjected to size-exclusion chromatography where PatGmac H618A eluted as primarily a monomer peak with a slight dimer shoulder (Figure 3.3). The addition of the small dimer contaminant has no effect on crystallisation. Final protein yields were 200 mg L^{-1} cell culture.

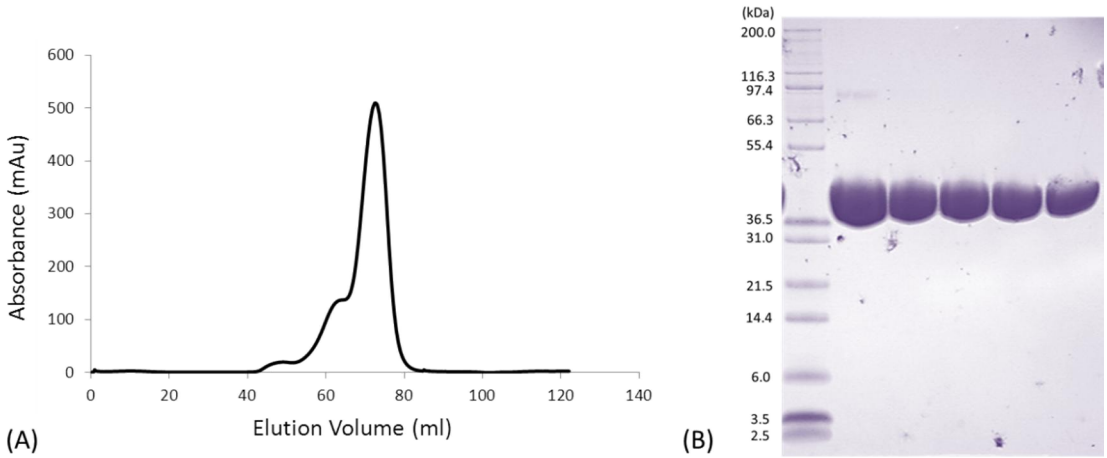


Figure 3.3: Purification of PatGmac H618A. (A) Gel filtration chromatograph of PatGmac H618A showing primarily a monomer peak, with a slight dimer shoulder. (B) SDS-PAGE analysis of PatGmac H618A gel filtration fractions showing > 98% purity.

3.3.2 Crystallography

Stochastic screening [108] of apo PatGmac H618A yielded several crystal hits with the best visually grown at 4 °C in a condition of 22 % PEG 3350, 50 mM calcium acetate. These crystals diffracted to 2.5 Å and belonged to the space group C2, the same as the native PatGmac (PDB: 4AKS). The crystal structure was determined by molecular replacement (MR) using Phaser with the native PatGmac structure as the search model. Apo crystals of PatGmac H618A were then subjected to soaking trials with 7.5 mM peptide VPAPIPFPAYDG at 4 °C both overnight and for three days. Datasets were collected on these crystals and the structure determined by MR, however no density for the peptide was evident.

Stochastic screening of PatGmac H618A in putative complex with 7.5 mM peptide VPAPIPFPAYDG yielded several crystal hits with 1.2 M sodium citrate, 0.1 M sodium cacodylate pH 7.0 giving the best crystals by visual analysis (Figure 3.4 A). Datasets were collected on these crystals again to 2.5 Å and the structure was determined by MR. The structure was analysed in Coot and there appeared to be density in the active site for one of two molecules in the asymmetric unit (AU). The density however was poor and therefore not conclusive suggesting that the peptide may be present but in low occupancy (Figure 3.5 A). In order to improve the occupancy, the co-crystals were subsequently soaked with a further 7.5 mM peptide overnight at 4 °C prior to data collection.

A dataset was collected on the soaked co-crystal to a resolution of 2.6 Å and the structure was again determined by MR (Figure 3.4 B, C). This time there was clear density in the active site of one of the two molecules in the AU which corresponding to residues PIPFPAYDG of the peptide (Figure 3.5 B). The peptide chain was fitted into the structure, and minor conformational changes were manually rebuilt in Coot. The structure was refined using REFMAC [117] and the structure checked by MolProbity

[124]. Final data collection and refinement statistics can be found in Table 3.2. Final MolProbity statistics can be found in Table 3.3

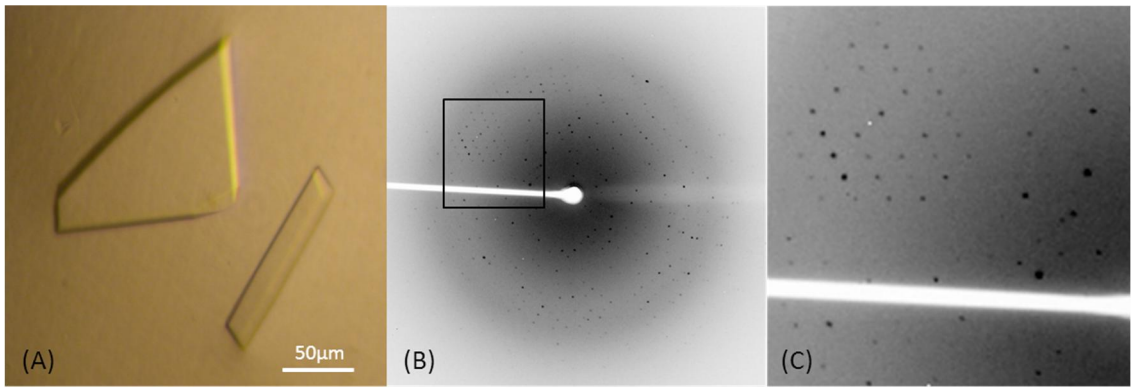


Figure 3.4: Crystallisation of PatGmac H618A in Complex with VPAPIPFPAYDG. (A) Single crystals of PatGmac H618A co-crystallised with VPAPIPFPAYDG peptide grown in a condition of 1.2 M sodium citrate, 0.1 M sodium cacodylate pH 7.0. (B) In-house diffraction pattern from a single crystal co-crystallised with VPAPIPFPAYDG peptide then soaked overnight with additional peptide. (C) Zoom view of diffraction spots

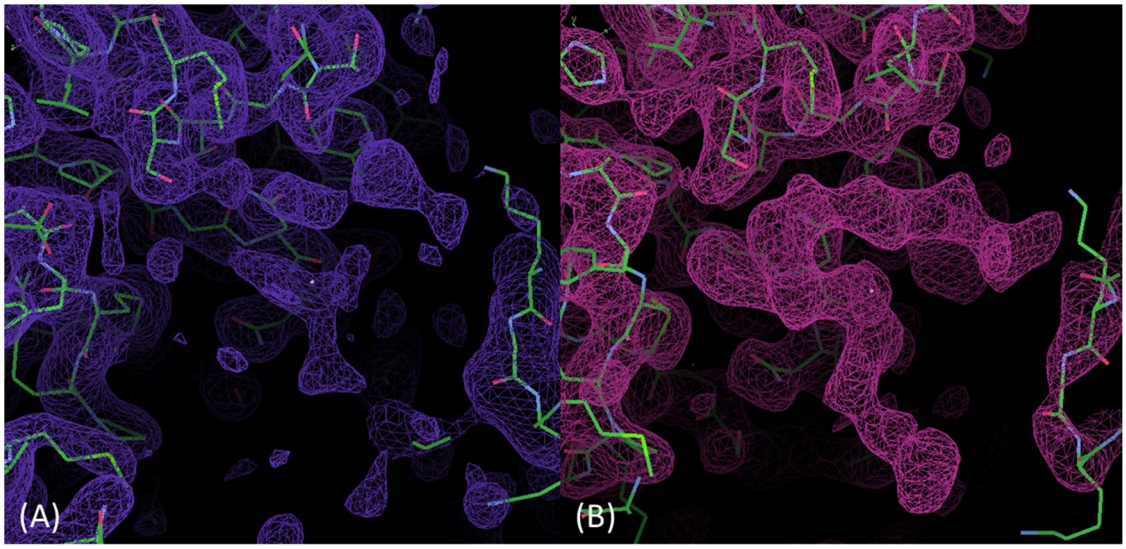


Figure 3.5: PatGmac Electron Density Representations. Electron density of PatGmac crystal structure zoomed in to the active site when (A) co-crystallised with peptide VPAPIPFPAYDG and (B) co-crystallised followed by additional soaking of peptide VPAPIPFPAYDG. Minor density for the peptide can be seen for the co-crystal however not conclusive for chain building. Unambiguous density for the peptide can be seen in the co-crystal plus soaking experiment and could be used for chain building. Both maps are contoured at 1 σ.

PatGmac H618A + Peptide Dataset	
Wavelength (Å)	1.54
Space group	C 1 2 1
a, b, c (Å)	135.6, 67.3, 137.9
α, β, γ (°)	90.0, 116.8, 90.0
Resolution (Å)	2.63 (2.77-2.63)
I/σI	10.1 (2.3)
R_{merge} (%)	10.1 (52.2)
Completeness (%)	99.3 (96.4)
Redundancy	3.7 (3.1)
Refinement	
R-factor	0.191
R_{free}	0.218
R.m.s.d	
Bond Lengths (Å)	0.009
Angles (°)	1.253
B-factors	
All	60.56
Protein	60.70
Peptide	77.98
Water	47.19

Table 3.1: Data Collection and Refinement Statistics of PatGmac H618A Complex with VPAPIPFPAYDG Peptide. Values for the highest resolution shell are provided in parentheses. The dataset was collected in house (Rigaku 007-HFM, Saturn 944 CCD).

All-Atom Contacts	Clashscore, all atoms:	2.63		100th percentile* (N=226, 2.63Å ± 0.25Å)
	Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1000 atoms.			
Protein Geometry	Poor rotamers	4	0.73%	Goal: <1%
	Ramachandran outliers	0	0.00%	Goal: <0.05%
	Ramachandran favored	638	98.15%	Goal: >98%
	MolProbity score^	1.05		100th percentile* (N=6042, 2.63Å ± 0.25Å)
	Cβ deviations >0.25Å	1	0.16%	Goal: 0
	Bad backbone bonds:	0 / 2642	0.00%	Goal: 0%
	Bad backbone angles:	0 / 3292	0.00%	Goal: <0.1%

Table 3.2: MolProbity Statistical Output for PatGmac Structure in Complex with VPAPIPFPAYDG Peptide. The final coordinate file was processed in the MolProbity server and confirmed as a structure in the 100th percentile in terms of quality when compared to structures of similar resolution [124].

3.3.3 Crystal Structure of PatGmac H618A in Complex with VPAPIPFPAYDG Peptide

The structure of PatGmac H618A, the inactive form of PatGmac, in complex with the substrate mimic VPAPIPFPAYDG peptide has been determined by molecular replacement to 2.63 Å (Figure 3.7 A, B). The VPAPIPFPAYDG peptide was chosen as it includes the eight residue core peptide and four residue macrocyclisation signal with proline residues mimicking the heterocycles of a normal substrate. This peptide is however a poor substrate and only achieves approximately 30 % turnover [141].

The difference electron density for the peptide in the active site of one of the molecules in the asymmetric unit was unambiguous for residues PIPFPAYDG (P5 to P4') (Figure 3.7 C). The second molecule in the asymmetric unit contains a hint of density for the peptide but is not strong enough to confirm. The refined model contains residues 514-686, 694-719, 727-747, 754-823, and 833-851 in chain A, and 515-651, 657-688, and 692-851 in chain B.

Residues P5 (Pro) and P4 (Ile) of the peptide make no contact with the protein while P3 (Pro) has some weak van der Waals interactions with the main chain. P2 (Phe) also makes some limited van der Waals contacts and the side chain sits in a shallow pocket.

The Pro of P1 adopts a *cis* peptide conformation and the side-chain makes van der Waals contacts with His618Ala and Val622. The carbonyl of the P1–P1' peptide is oriented for nucleophilic attack by the hydroxyl of Ser783 with the side-chain of Met784 sitting on this face of the carbonyl. The tetrahedral intermediate is stabilised by the side-chain of Asn717 pointing towards the opposite face of the carbonyl. The P1' Ala makes only a few hydrophobic interactions, including contacts with Met784 and the protein backbone. The P2' (Tyr) residue makes extensive contacts with the protein: A π stacking interaction with Phe747, a hydrogen bond to His746 and hydrogen bonds between the main-chain oxygen and the nitrogen of Thr780. The side-chain of P3' (Asp) is oriented towards a large electropositive patch created by Arg589, Lys594, and Lys598. It makes a salt bridge with Lys598 and possibly Lys594, though the side chain of Lys594 is not well ordered. The P4' Gly residue appears to make no direct contact with the protein, although larger residues in this position may be disfavoured by the proximity to Lys594. (Figure 3.6, 3.7 B)

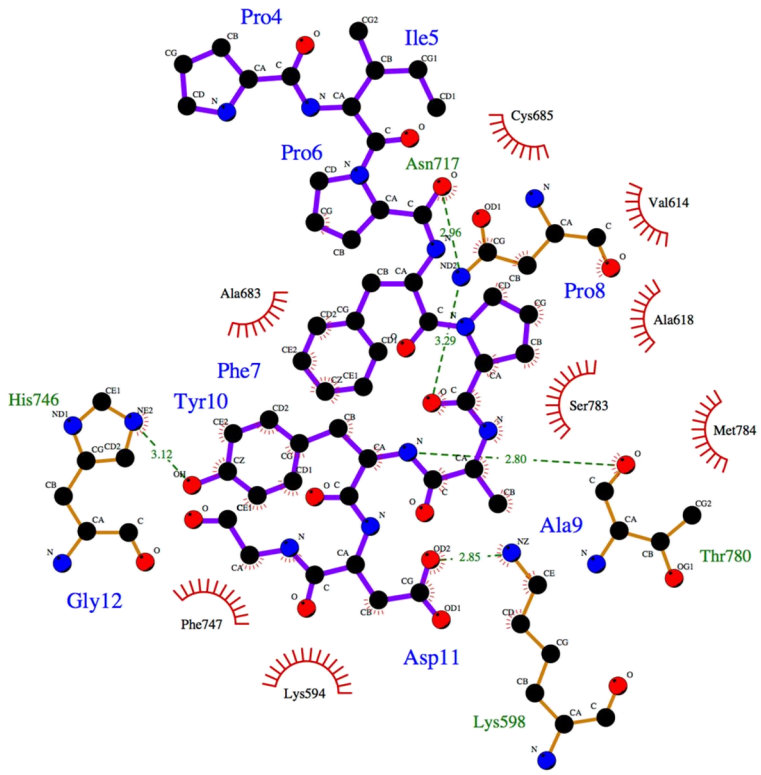


Figure 3.6: PatGmac Binding to VPAPIPFPAYDG Peptide. LigPlot^{+} representation of the residues involved in PatGmac binding of the substrate peptide VPAPIPFPAYDG (only residues 'PIPFAPAYDG' are ordered). Figure was created in LigPlot^{+} [144]

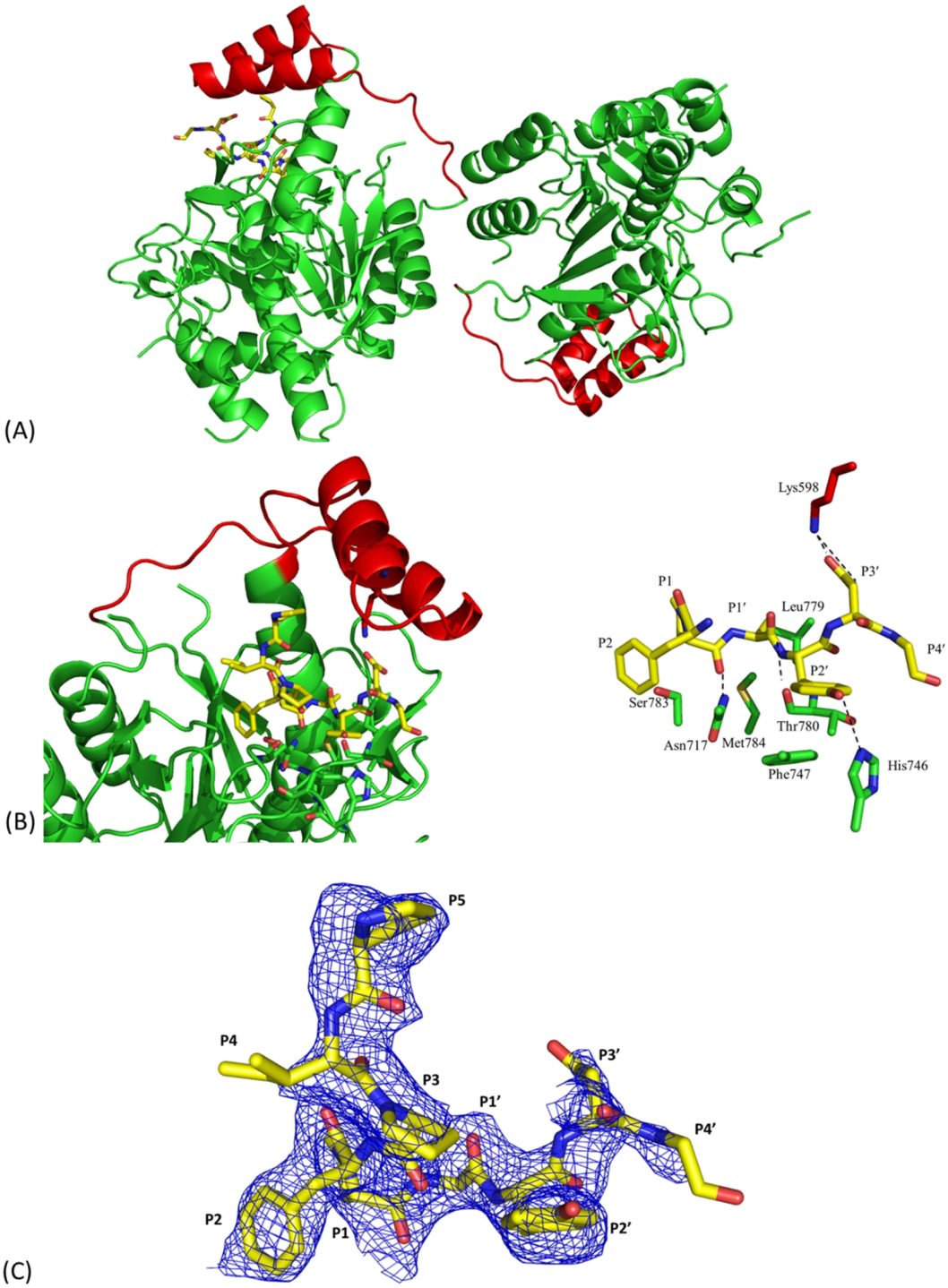


Figure 3.7: The Crystal Structure of PatGmac H618A in Complex with Peptide 'VPAPIPFPAYDG'. (A) Asymmetric unit of PatGmac H618A in complex with peptide 'VPAPIPFPAYDG' featuring two monomers, one with substrate present (yellow sticks), one without. Macrocyclisation insertion residues are shown in red. (B) Two representations of the active site of PatGmac H618A with substrate peptide bound (yellow). (C) Difference electron density (Fo-Fc map) fit for residues 'PIPFPAYDG' in chain A of PatGmac H618A contoured at 2.5σ.

3.3.4 Biochemical Studies on PatGmac

To examine substrate specificity, a range of mutant PatGmac proteins were expressed and purified for use in biochemical reactions with the substrate mimic peptide 'VGAGIGFPAYDG'. The mutants K598D, K594D, R589D/K594D/K598D were made based upon the key interactions in the active site involved in the binding of the AYDG leaving group. The reactions were analysed by LC-MS for unprocessed substrate and linear and cyclic products (Table 3.3). (The data accumulated from this method has some limitations as it is appreciated that certain peptides may ionize better in the MS than others however it should give a good comparison of the products. Additionally, due to the lack of charged N- and C- termini it more likely that the cyclic peptides will not ionize as well as linear peptides.)

	Unprocessed ion count (%) (M + H = 1123)	**Linear ion count (%)** (M + H = 717)	**Cyclic ion count (%)** (M + H = 699)
$PatG_{mac}$	0	0	100
$PatG_{mac}$K598D	0	100	0
$PatG_{mac}$K594D	0	71	29
$PatG_{mac}$ R589D/K594D/K598D	94	6	0

Table 3.3: Relative Ion Counts of Linear Cleaved and Macrocyclised Peptide Substrate.

Native PatGmac accepts the peptide VGAGIGFPAYDG and macrocyclises it with 100 % efficiency when reacted 100 µM peptide : 50 µM PatGmac at 37 °C for 120 hours. PatGmac K598D accepts the peptide however results in only linear product being formed with the complete absence of any macrocyclised peptide. PatGmac K594D gives a 7:3 ratio of linear to macrocyclised product. Finally, PatGmac R589D/K594D/K598D mutant results in mostly unprocessed starting material with a small (6 %) amount of linear product. (Figure 3.8)

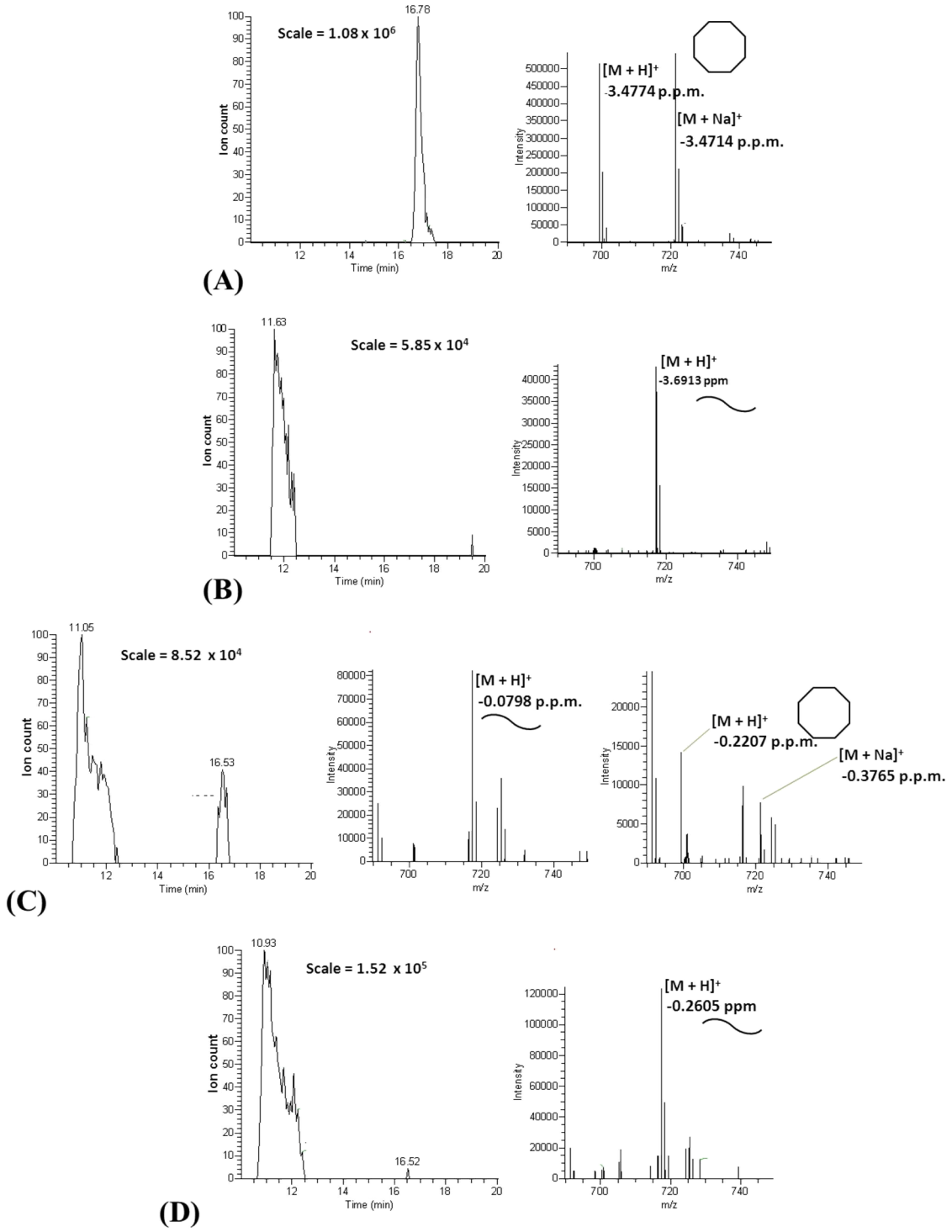


Figure 3.8: Mass Spectrometry Analysis of Peptide VGAGIGFPAYDG after Reaction with Native and Mutant PatGmac. LC-MS analysis showing chromatogram and spectra of VAGAGIGFPAYDG peptide following reaction with (A) native PatGmac resulting in complete macrocyclisation (octagon) (B) PatGmac K598D showing 100 % linear product (waved line). (C) PatGmac K594D gives a mixture of linear and macrocycle products (D) PatGmac R589D/K594D/K598D showing only linear product.

3.4 Discussion

The crystal structure of PatGmac in complex with substrate mimic peptide, VPAPIPFPAYDG, has provided extensive information on the protein-peptide interactions involved in the macrocyclisation process and has also given us some insight into the mechanism of reaction.

The complex structure shows that the P1-P1' bond of the substrate is in the correct position for nucleophilic attack by Ser783 with Asn717 stabilising the tetrahedral intermediate, which would subsequently collapse to the classical acyl-enzyme intermediate.

It is known that there is a requirement for macrocyclisation to have a proline or a heterocycle at the C-terminal end of the modified core peptide (P1) [88]. The complex structure identifies that a proline in this position adopts a *cis* conformation, enabling the peptide to bend back on itself which in turn promotes macrocyclisation. A non-cyclic residue in this position would not be able to adopt a similar conformation without a significant energy penalty [145] [146]

The side chain of P2 (Phe) points into a non-specific shallow pocket while P3 – P5 point away from the protein surface. The non-specific cavity that binds P2 and the lack of any further contacts between PatGmac and the patellamide core peptide is consistent with the known lack of enzyme specificity for the core peptide sequence with the exception of the P1 residue [64] and explains why the core peptide region is hypervariable.

The Ala (P1') and Gly (P4') of the macrocyclisation signature, 'AYDG', appear to make no significant interactions with the protein, while the aromatic ring of Tyr (P2') makes extensive interactions. The Asp (P3') forms salt bridges with one, possibly two, lysine residue(s) (Lys598 and possibly Lys594).

In order to favour the nucleophilic attack of the acyl-enzyme intermediate by the N-terminus of the patellamide modified core peptide over hydrolysis it is essential for PatGmac to protect the acyl-enzyme-intermediate from water which is at 55.5 M. We propose that this is achieved through the requirement of the macrocyclisation signature 'AYDG' remaining bound post-cleavage and working together with the macrocyclisation insertion shield the acyl-enzyme intermediate from water (Figure 3.9).

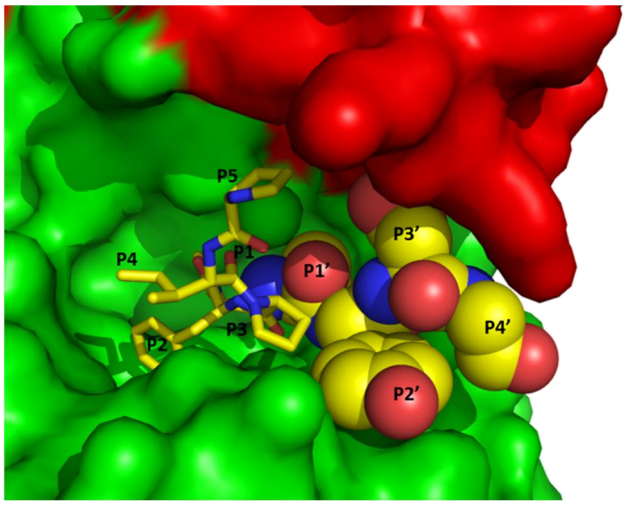


Figure 3.9: PatGmac in Complex with Peptide VPAPIPFPAYDG. Structure of PatGmac with peptide showing how the AYDG leaving group (yellow - spheres) along with the macrocyclisation insertion (red) combine to shield the active site from water, giving the core peptide residues (yellow - sticks) time to orientate for nucleophilic attack of the acyl-enzyme intermediate.

It was shown that removal of the macrocyclisation insertion removes macrocycle-forming activity whilst retaining protease activity [141]. Point-mutations in the macrocyclisation insertion designed to disrupt the interactions it has with the macrocyclisation signature of the peptide have also been found to either reduce (K594D) or abolish (K598D) macrocyclase activity.

The accumulation of both structural and biochemical data has now allowed a mechanism for the macrocyclisation of peptides by PatGmac to be proposed (Figure 3.9)

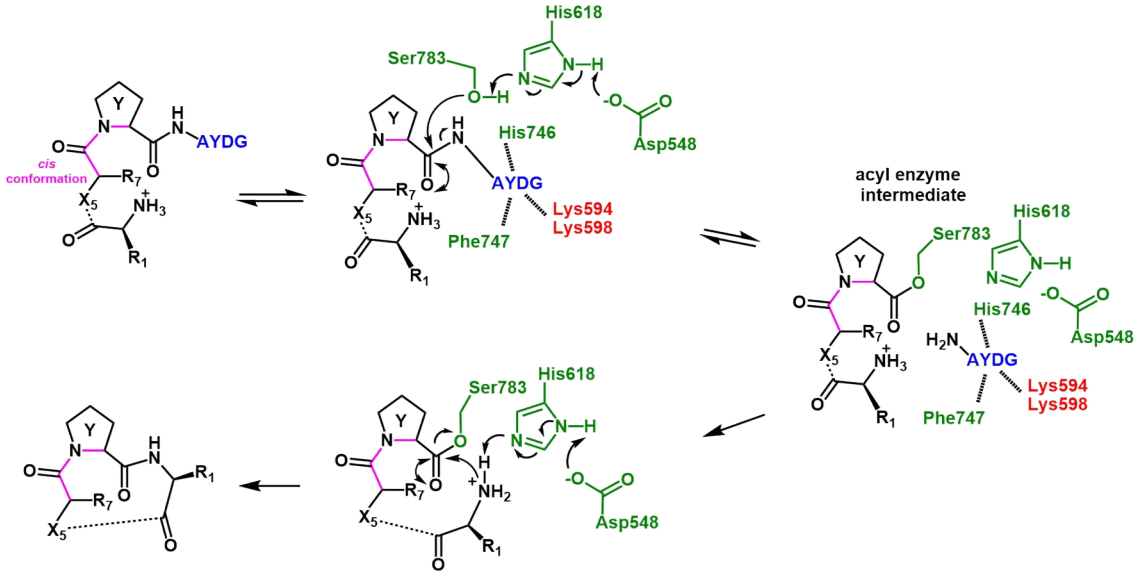


Figure 3.10: Proposed PatGmac Mechanism. The macrocyclase cleaves the P1-P1' bond by means of the catalytic triad (green) forming the classical acyl enzyme intermediate. The AYDG recognition portion of the peptide (blue) is held in place by protein: peptide interactions from the macrocyclisation insert (red) and not released until the N-terminal of the core peptide is in place for the nucleophilic attack of the acyl enzyme intermediate. Following macrocyclisation the enzyme collapses back to its native state releasing the macrocycle. Amino acid Y, which must be either a proline or heterocycle (thiazol(in)e or oxazol(in)e), adopts a *cis* peptide conformation (magenta). Figure adapted from Koehnke *et al.* (2014) [147]

3.5 Conclusions and Future Work

The crystal structure of PatGmac in complex with substrate mimic peptide, VPAPIPFPAYDG, has been determined by molecular replacement to 2.63 Å resolution. This structure has allowed the identification of the key interactions involved in substrate binding and the rationalisation of its macrocyclisation mechanism. The amino acid insertion represents a head group which sits over the active site. The AYDG leaving group of the substrate forms extensive interactions holding it in place within the binding site. Together these two features, shield the active site from water until the N-terminus of the core peptide can orientate itself to attack the acyl-enzyme intermediate to form the cyclic peptide. The presence of a heterocycle or proline in the P1 position is required as these can adopt the *cis* conformation which is needed to allow the peptide to bend back on itself promoting macrocyclisation.

The rate-limiting step involved in macrocyclisation was still to be determined; is it the rate of proteolysis or the time taken for the N-terminus of the peptide to orientate for nucleophilic attack or perhaps given the structural insights it could possibly be the dissociation rate of AYDG? Following our publication of this work [141], Agarwal *et al.* (2012) subsequently published their own structure of PatGmac which independently validated our structure [66]. Their study further explored the rate of proteolysis by mutating the active site serine (Ser783) to a cysteine and a proline (Pro225) to an alanine to form a subtiligase type enzyme as described previously [148] [149] [150]. The change of the oxygen to a sulfur atom should result in a better nucleophile and as a result the attack of the acyl-enzyme intermediate occurs at an increased rate. The changes were implemented and assessed for activity and as expected increased levels of macrocyclisation were observed over the wild type enzyme.

Further work will be carried out to assess if PatGmac can be re-engineered to increase rates, accept alternate peptide motifs and accept longer core peptides and thus produce larger macrocycles. Alternatively, homologues of PatGmac will also be studied

as there are known macrocyclases which can process up to 20 residues in the core peptide [34].

4. *In Vitro* Biosynthesis of Patellamides

4.1 Introduction

Cyclic peptides are common entities in a wide variety of drugs and biotechnology tools [151]. Their synthesis however can be particularly challenging and expensive. Many cyclic peptide natural products have been discovered in a wide range of species and there is much interest in exploiting the biosynthetic pathways for their production to produce novel chemically interesting cyclic peptides.

The patellamides, a member of the cyanobactin superfamily, are cyclic peptides of six to eight amino acids in length and contain D-stereo centres and heterocyclised amino acids in the form of oxazolines (from serine, denoted S^{Oxn} or from threonine, denoted T^{Oxn}) and thiazolines (from cysteine, denoted C^{Thn}), the latter of which can be further oxidised to thiazoles (denoted C^{Thz}) [40].

Patellamide biosynthesis has been examined in some detail, as previously described in Chapter 1, and its products have a range of bioactivities including reversing drug efflux pumps and showing cytotoxicity against leukaemia cells [40], [55]. These compounds have however had very limited development since their initial isolation and this is most likely due to the lack of ability to obtain them in significant quantities. Natural sources are difficult to access and yields from these are low whilst the bacterial strain responsible for production, *Prochloron spp.* can't be cultured, as it requires its symbiont, the sea squirt *Lissoclinum patella*, for sustained growth [152]. Chemical synthesis of the patellamides is possible, but is a complex multi-step process with no fast route to diversity (Chapter 1, [59] , [60]). A recent study has been carried out showing that the patellamide pathway can be utilised *in vivo* using *Escherichia coli* cells to generate patellamides, and patellamide-derived cyclic peptides containing unnatural amino acids [103]. Despite the success in producing these, yields in the

range of µg L^{-1} culture continues to make generating enough material for further studies challenging.

We aimed to explore the *in vitro* biosynthesis of cyclic peptides utilising isolated enzymes from the patellamide pathway (and related pathways) and to re-engineer the pathway where appropriate to improve reaction rates and overall yields.

The work on the nine residue compound 'IITACIMAC' was carried out in collaboration with Rachael Graham (University of St. Andrews, B.Sc project student).

4.2 Material and Methods

4.2.1 DNA Cloning

To explore *in vitro* cyclic peptide production, a PatE precursor peptide (PatE2) was engineered by Dr Jesko Koehnke (University of St Andrews) consisting of a 37-residue N-terminal leader sequence and N- and C- terminal cleavage recognition sites flanking a single core peptide (ITACITFC) corresponding to the natural product Patellamide D. In addition, a C-terminal His_6-tag was added to aid in the purification process (Figure 4.1 A). The construct was cloned into the pBMS23CHis plasmid (gift from Dr Huanting Liu, St Andrews).

Following initial experiments with PatA and discussion with colleagues, it was decided to introduce alternatives to the PatA recognition site for N-terminal cleavage. PatE2 was subsequently mutated by Dr. Wael Houssen (University of Aberdeen) to introduce a lysine between the final residue of the PatApr cleavage site, 'GLEAS', and the first residue of the core peptide to allow for cleavage by trypsin (PatE2K, Figure 4.1 B). A further construct to introduce the TEV protease site 'ENLYFQ' immediately after the 'GLEAS' motif of PatE2 was created using site directed mutagenesis protocols described in section 2.2.1 with primers shown in Table 4.1. (PatE2TEV, Figure 4.1 C).

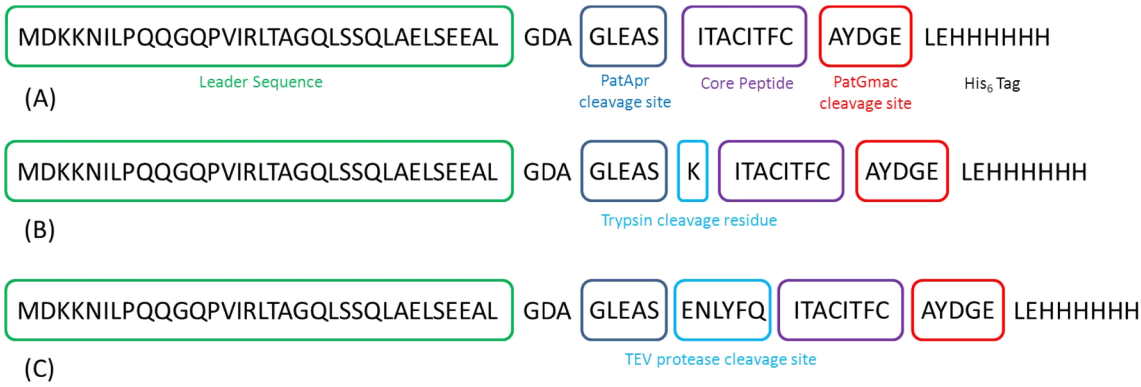


Figure 4.1: **PatE Protein Sequences.** Schematics of (A) PatE2, (B) PatE2K and (C) PatE2TEV precursor peptide sequences identifying leader sequences, cleavage sites, core peptide residues and the C-terminal His_6-tag.

Finally, a range of PatEs based on the PatE2K sequence were created with variations in the core peptide sequence. These were all created by PCR using a standard 5' primer with *Nco1* restriction site, while varying the 3' primer to introduce the new desired core peptide sequence and an *Xho1* restriction site (Table 4.1). These new PatE inserts were separated on an agarose gel and extracted using the Qiagen Gel Extraction Kit. The purified DNA moleule was cleaved using restriction enzymes *Nco1* and *Xho1* and ligated into the pBMS23CHis plasmid (pre-cut with the same enzymes and then alkaline phosphatase treated). The ligated plasmid was used to transform *E. coli* DH5α cells and the single colonies obtained were grown in 10 ml LB culture, mini-prepped (Qiagen MiniPrep Kit) and their sequences confirmed by sequencing (GATC Biotech).

Full length PatD, full length TruD and PatApr (PatA residues 10-289) were synthesised into the pJexpress 411 plasmid using optimised codons for *E. coli* (DNA 2.0) [105]. Each plasmid contained an N-terminal His_6-tag, and TruD and PatApr additionally contained a TEV protease recognition site between the His_6-tag and the protein. PatGmac was cloned as described in section 3.2.1.

4.2.2 PatE Expression and Purification

PatE2 was expressed from the pBMS23CHIS vector in *E. coli* BL21 (DE3) cells grown on auto-induction medium [107] (see Appendix A.4 for media and buffer compositions) for 24 hours at 30 °C, driving the protein to inclusion bodies. Cells were harvested by centrifugation at 4,000 x g for 15 min at 20 °C, re-suspended in urea lysis buffer and lysed by sonication at 15 microns (SoniPrep 150, MSE). The lysate was cleared by centrifugation at 40,000 x g, 20 °C, 20 min followed by passage through a 0.45 µm filter. The cleared lysate was applied to a Ni-Sepharose 6 FF column (GE Healthcare) equilibrated with urea lysis buffer and protein was eluted with urea elution buffer. The protein was then supplemented with 10 mM dithiothreitol (DTT) to induce refolding and subjected to size-exclusion chromatography (Superdex 75, GE Healthcare) in gel

filtration buffer. The protein was visualised by SDS-PAGE and its identity confirmed by MS.

5’ PatE	cttccatggacaaaaaaaacattcta
5’ PatETEV	gctggtttggaagcatctgaaaacctgtattttcagataactgcttgcat cactttttgcgct
3’ PatETEV	ctgaaaatacaggttttcagatgcttccaaaccagcatcgccgagagc
3' core pep ‘VTVCVTVC’	cttctcgagttcaccatcataagcgcacacggtcacgcacacggtcactt tagatgcttccaaaccagc
3' core pep ‘ITACITYC’	cttctcgagttcaccatcataagcgcaataagtgatgcaagcagttattt tagatgcttccaaaccagc
3' core pep ‘IMACIMAC’	cttctcgagttcaccatcataagcgcacgccatgatgcacgccatgattt tagatgcttccaaaccagc
3' core pep ‘IDACIDFC’	cttctcgagttcaccatcataagcgcaaaaatctatgcacgcatctattt tagatgcttccaaaccagc
3' core pep ‘ITACITAC’	cttctcgagttcaccatcataagcgcacgcagtgatgcaagcagttattt tagatgcttccaaaccagc
3' core pep ‘ATACITFC’	cttctcgagttcaccatcataagcgcaaaaagtgatgcaagcagtcgctt tagatgcttccaaaccagc
3' core pep ‘ITACISFC’	cttctcgagttcaccatcataagcgcaaaagctgatgcaagcagttattt tagatgcttccaaaccagc
3' core pep ‘ICACITFC’	cttctcgagttcaccatcataagcgcaaaaagtgatgcaagcgcatattt tagatgcttccaaaccagc
3' core pep ‘IAACITFC’	cttctcgagttcaccatcataagcgcaaaaagtgatgcaagccgctattt tagatgcttccaaaccagc
3' core pep ‘ITAAITFC’	cttctcgagttcaccatcataagcgcaaaaagtgatcgcagcagttattt tagatgcttccaaaccagc
3’ core pep ‘IITACIMAC’	cttctcgagttcaccatcataagcgcacgccatgatgcacgcggtaataa ttttagatgcttccaaaccagc
3' core pep ‘ITVCISVC’	cttctcgagttcaccatcataagcgcacacgctaatgcacacggttattt tagatgcttccaaaccagc
3' core pep ‘ICFPTIC’	cttctcgagttcaccatcataagcgcaaatggtcggaaagcaaattttag atgcttccaaaccagc

Table 4.1: **PatE Primer Sequences.** PCR primer sequences used to generate PatE variants.

Both PatE2K, PatE2TEV and all PatE2K core residue variants were expressed and purified using the same methods as PatE2.

4.2.3 Enzyme Expression and Purification

For *in vitro* processing, each enzyme was individually expressed and purified with the exception of Bovine Trypsin which was sourced from Sigma-Aldrich. PatGmac was expressed and purified as previously described (Section 3.2.2). PatD, TruD and PatApr were expressed from the pJexpress 411 vector. All three were expressed using BL21 (DE3) *E. coli* grown on auto-induction medium using the Studier method [107] (see Appendix A.5 – A.7 for media and buffer compositions). The cultures were grown at 20 °C, 250 rpm for 48 hours before harvesting by centrifugation.

PatApr and TruD were both purified by re-suspending the cell pellets in lysis buffer plus DNAse (0.4 mg g^{-1} wet cell pellet). TruD was additionally supplemented with EDTA-free protease inhibitor tablets (Roche). The re-suspended cells were passed through a cell disruptor (Constant Systems) at 30 kpsi. The lysate was cleared by centrifugation (40,000 x g, 4 °C, 20 min) and then loaded onto a Ni-Sepharose 6 FF column (GE Healthcare) equilibrated in lysis buffer. The column was washed with lysis buffer and the protein eluted with elution buffer. The protein was then passed through a desalting column (Desalt 16/10, GE Healthcare) in desalting buffer. Tobacco Etch Virus (TEV) protease was added at a ratio of 1 mg TEV protease per 10 mg protein and incubated for 1 hour at 20 °C to remove the His_6-tag. The sample was then loaded on to a second Ni-column, in desalt buffer. The protein was collected in the flow through, concentrated to 7.5 ml (Vivaspin concentrators) and subjected to size exclusion chromatography. PatApr was applied to a Superdex 75 gel filtration column (GE Healthcare), while TruD was applied to a Superdex 200 gel filtration column (GE Healthcare), both equilibrated in sizing buffer.

PatD was purified by re-suspending the cell pellets in lysis buffer plus DNAse (0.4 mg g^{-1} wet cell pellet) and EDTA-free protease inhibitor tablets (Roche). The re-suspended

cells were passed through a cell disruptor (Constant Systems) at 30 kpsi. The lysate was cleared by centrifugation (40,000 x g, 4 °C, 20 min) and then loaded onto a Ni-Sepharose 6 FF column (GE Healthcare) equilibrated in lysis buffer. The column was washed with lysis buffer and the protein eluted with elution buffer. PatD was concentrated to 7.5 ml and applied to a Superdex 200 gel filtration column (GE Healthcare) equilibrated in gel filtration buffer.

The purity of all proteins was confirmed by SDS-PAGE analysis and their identity confirmed by mass spectrometry (MS).

4.2.4 Heterocyclisation

Purified PatE2 and PatE2K precursor peptides were subjected to *in vitro* processing. The incubation of 100 µM PatE2/PatE2K with 5 µM PatD or TruD in 150 mM NaCl, 10 mM HEPES pH 7.4, 5 mM ATP, 5 mM $MgCl_2$ at 37 °C for two hours was carried out for heterocyclisation. Alternatively, to reduce heterocyclase enzyme usage, 100 µM PatE2/PatE2K was incubated with 1 µM PatD or 0.5 µM TruD in 150 mM NaCl, 10 mM HEPES pH 7.4, 5 mM ATP, 5 mM $MgCl_2$ at 37 °C for 16 hours to drive heterocyclisation to completion. Heterocyclised peptides were purified from the PatD / TruD on an S75 gel filtration column (GE Healthcare) and the heterocyclisation state of the peptide was confirmed by MS.

4.2.5 N-terminal Core Peptide Cleavage

Heterocyclised (by either TruD or PatD) PatE2 pro-peptide was incubated with 20 µM PatApr at 37 °C for 200 hours to drive N-terminal cleavage to completion. The cleaved peptide was separated from the leader sequence and PatApr by passage through a Superdex S30 column (GE Healthcare) pre-equilibrated in 150 mM NaCl, 20 mM Bicine pH 8.1. PatE2K (after reaction with PatD or TruD) was N-terminally cleaved by

incubation with 1:250 bovine trypsin (Sigma) at 37 °C for two hours and the cleaved peptide purified as for PatE2.

4.2.6 C-terminal Core Peptide and Macrocyclisation

100 µM heterocyclised and N-terminally cleaved PatE2/PatE2K was incubated with 20 µM PatGmac for 72 hours at 37 °C in 20 mM Bicine pH 8.1, 500 mM NaCl, 5 % DMSO to drive the completion of macrocyclisation as determined by MALDI TOF MS.

4.2.7 Cyclic Peptide Purification

The final macrocycles derived from PatE2/PatE2K were purified by reverse-phase HPLC by passage over a Jupiter C_4 column (Phenomenex) then subsequent passage of macrocycle peaks over an Eclipse C_{18} column (Agilent). The sample was applied to both columns and eluted with a 5 - 95 % MeOH/water gradient. Peak fractions were identified by MALDI TOF MS.

4.2.8 Mass Fragmentation

Mass fragmentation of both HPLC purified macrocycles was carried out by Dr. Matt Fuszard and Dr. Sally Shirran. The samples were applied to LC/MS/MS on an AB SCIEX Triple TOF 5600 system. Peaks were assigned based on the expected chemical structures.

4.2.9 NMR Spectroscopy

^{1}H Nuclear Magnetic Resonance (NMR) spectroscopy was carried out on the macrocycle resulting from PatE2K treatment with TruD, trypsin and PatGmac (cyclo[ITAC^{Thn}ITFC^{Thn}]) following HPLC purification. The sample was analysed on a 500 MHz NMR spectrometer (Bruker) in 20 mM sodium potassium phosphate pH 6.8. ^{1}H NMR spectroscopy was also carried out on cyclo[IMAC^{Thn}IMAC^{Thn}] with the sample analysed on a 500 MHz NMR spectrometer (Bruker) in 100 % deuterated methanol. NMR experiments were carried out and the data processed by Dr. Uli Schwarz-Linek (University of St Andrews).

4.3 Results

4.3.1 DNA Cloning

PatE2 and PatE2K constructs were supplied by collaborators. PatE2TEV and all core peptide variants were cloned by PCR using the PatE2K DNA as a template. The newly synthesised DNA inserts were analysed by argarose gel electrophoresis to confirm they were of correct size (Figure 4.2) and then DNA sequenced (GATC Biotech) to ensure correct sequence.

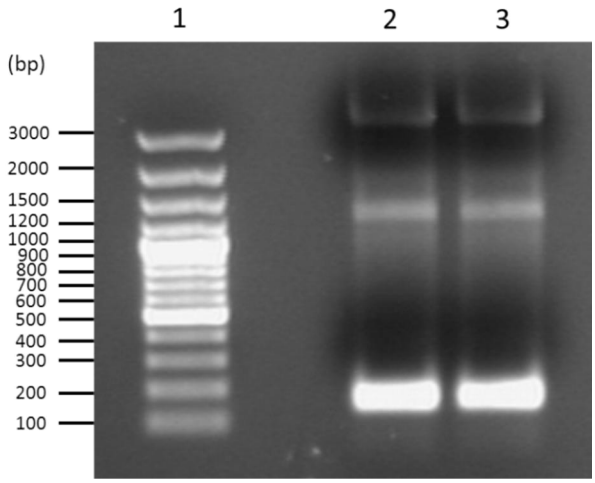


Figure 4.2: **PatE DNA Cloning.** Representative examples of PatE inserts (200 bp) run on agarose gel electrophoresis. Lane 1) 100bp plus DNA markers (Life Technologies), Lane 2) PatE2K-'VTVCVTVC' and Lane 3) PatE2K-'ITACITYC'.

4.3.2 PatE Expression and Purification

PatE2 and PatE2K were overexpressed with a C-terminal His_6-tag in BL21 (DE3) *E. coli* cells using the Studier auto-induction method [107]. The peptide was solubilised from inclusion bodies in urea lysis buffer, isolated by Ni-NTA chromatograpy and refolded using DTT. The peptide elutes off a gel filtration column as a single peak at approximately 75 ml (Figure 4.3). The final protein yield was between 200 and 250 mg L^{-1} culture.

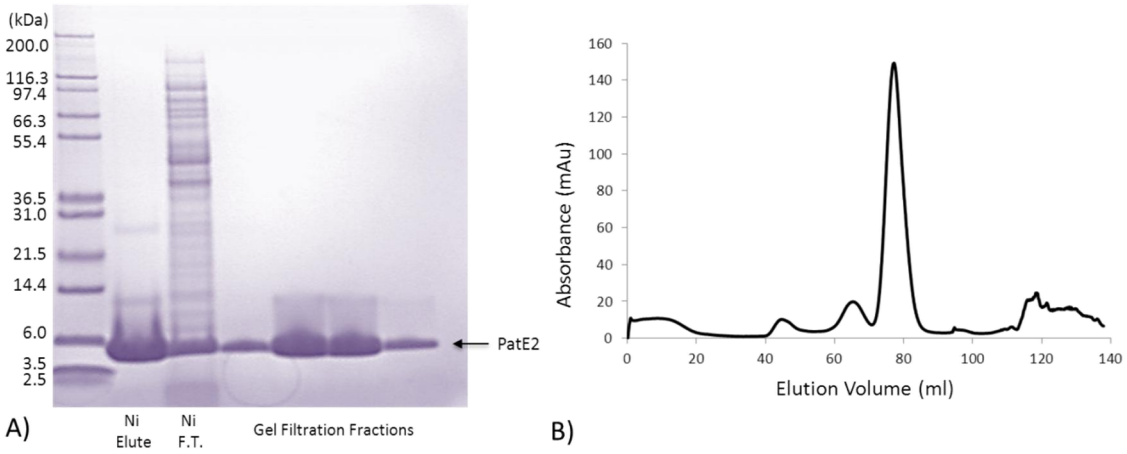


Figure 4.3: Purification of PatE2 Precursor Peptide. A) SDS-PAGE analysis of PatE2 showing > 98% purity. B) Gel filtration UV chromatograph of PatE2 showing protein elutes as a single peak. Due to the unfolded nature of PatE2 it runs aberrantly on SDS-PAGE ~5 KDa and elutes at 75 ml on gel filtration.

4.3.3 Enzyme Expression and Purification

PatGmac, PatApr and TruD were all overexpressed with an N-terminal His_6-tag and TEV protease recognition site in BL21 (DE3) *E. coli* using the Studier auto-induction method. The proteins were isolated using the His_6-tag on nickel resin and the tag removed by addition of TEV protease with cleavage efficiency close to 100 %. PatGmac was further purified on an ion-exchange mono-Q column with the protein eluting at 350 mM of a NaCl gradient. PatGmac and PatApr were finally subjected to size exclusion on an S75 gel filtration column (GE Healthcare) while TruD was subjected to size exclusion on an S200 gel filtration column (GE Healthcare).

PatD was overexpressed with an N-terminal His_6-tag in BL21 (DE3) *E. coli* cells using the Studier auto-induction method. The protein was isolated using the His_6-tag on nickel resin and then subjected to size exclusion on an S200 gel filtration column (GE Healthcare).

PatApr, PatGmac and TruD all eluted from gel filtration as monomers, with PatGmac and TruD showing slight dimer contaminants which have no detrimental effects on subsequent biochemical reactions. PatD eluted from gel filtration as a dimer (Figure 4.4 A). All proteins were analysed by SDS-PAGE (Figure 4.4 B) and their identities were confirmed by MS. Final protein yields are summarised in Table 4.2

Protein	Yield (mg protein L^{-1} cell culture)
PatE2 / PatE2K	200-250
PatApr	200-250
PatGmac	200-250
TruD	40-60
PatD	20-30

Table 4.2: Protein Purification Yields. Final purified protein yields per litre of cell culture for all proteins involved in cyclic peptide *in vitro* biosynthesis.

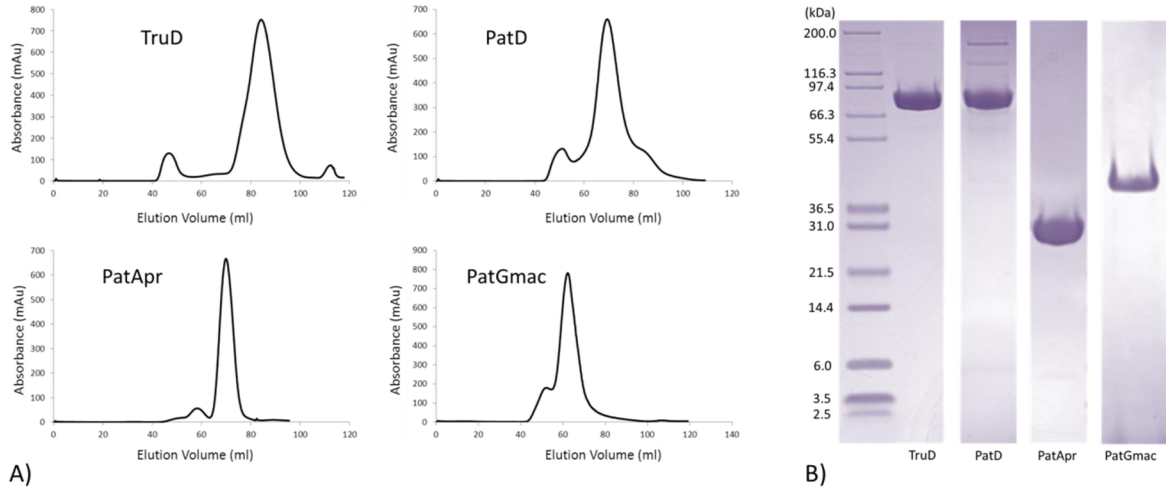


Figure 4.4: Enzyme Purification. (A) Gel filtration chromatographs of each enzyme showing primarily monomer elutes for TruD, PatApr and PatGmac and dimer elute for PatD (B) SDS-PAGE analysis of the final purified protein for enzymes TruD, PatD, PatApr and PatGmac showing > 98 % purity for all but PatD which shows > 95 % purity.

4.3.4 Heterocyclisation

Incubation of PatE2/PatE2K with TruD results in a loss of 36 Da by MS corresponding to the expected two water losses resulting from cysteine heterocyclisation (two heterocycles). The incubation of PatE2/PatE2K with PatD results in the loss of 72 Da corresponding to both threonine and cysteine heterocyclisation (total of four heterocycles) (Figure 4.5).

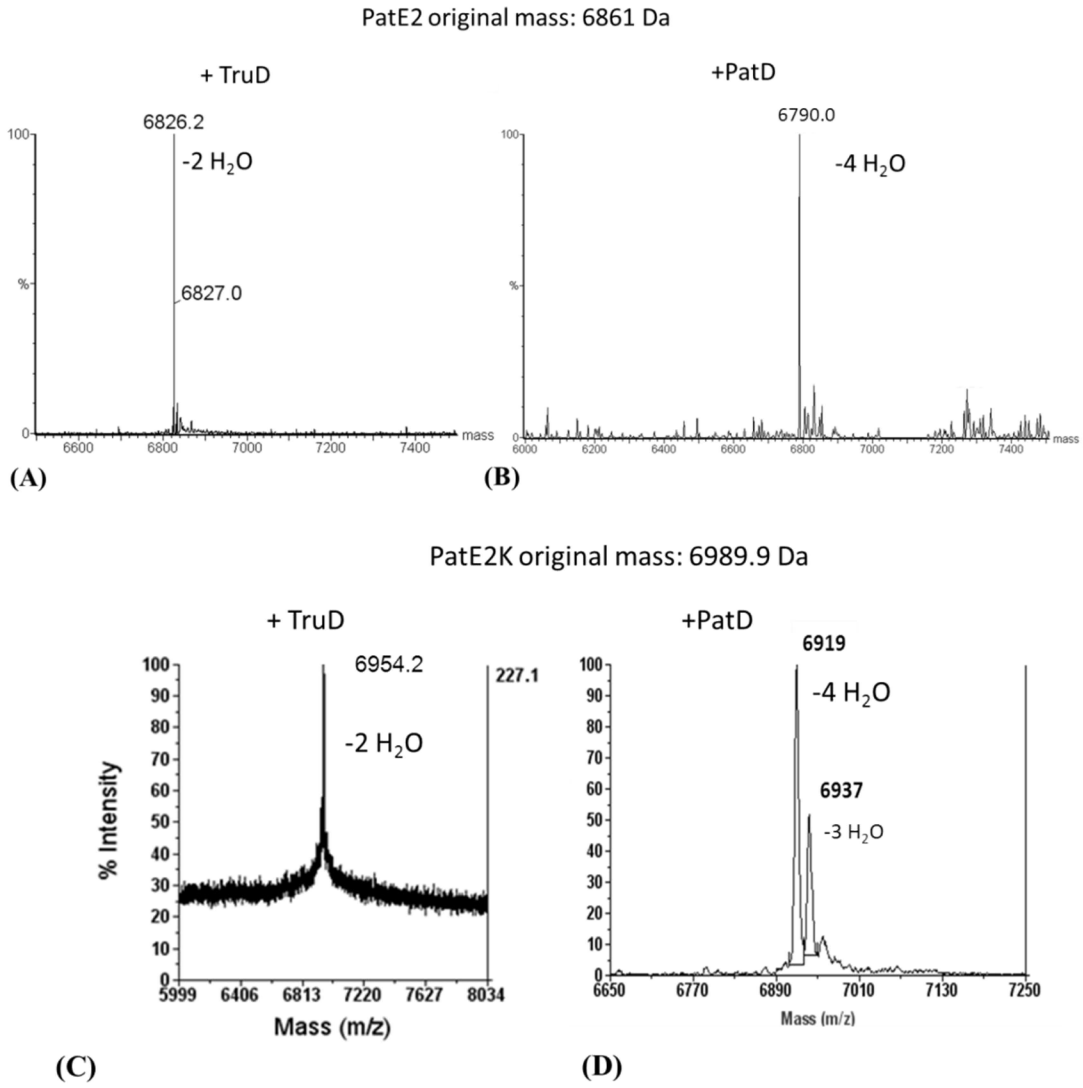


Figure 4.5: MS Analysis of Heterocyclisation. MALDI TOF MS spectra of peptide products from the reactions of A) PatE2 + TruD - $[M+H]^+$ of 6826, B) PatE2 + PatD - $[M+H]^+$ of 6790 C) PatE2K + TruD - $[M+H]^+$ of 6954 and D) PatE2K + PatD - $[M+H]^+$ of 6919 (Note: E2K + PatD shows a secondary $[M+H]^+$ peak of 6937 corresponding to the formation of only 3 heterocycles).

The heterocyclised peptide was purified from the heterocyclases by gel filtration (Figure 4.6). The heterocyclases can be recycled at a yield of up to 80 % recovery and remain active. The enzyme still achieves the same level of heterocyclisation as when freshly purified and has no adverse effect on the reaction rates.

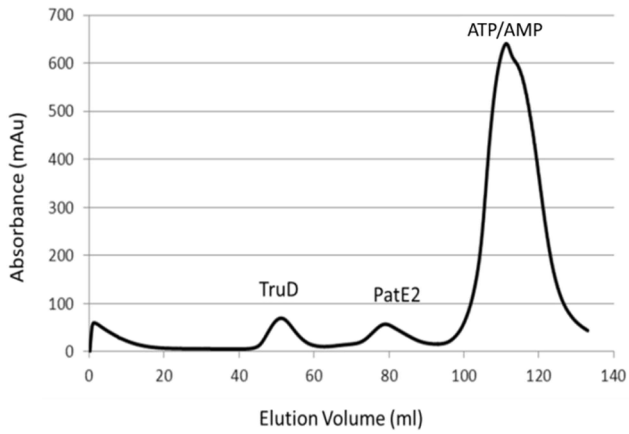


Figure 4.6: Purification of PatE2 from TruD. S75 gel filtration UV chromatogram of the reaction between PatE2 and TruD showing three distinct elution peaks corresponding to TruD, heterocyclised PatE2 and ATP/AMP.

4.3.5 N-terminal Core Peptide Cleavage

PatApr mediated cleavage of heterocyclised PatE2 is a slow process and takes at least 200 hours at 37 °C to achieve completion. The protease cleaves between the 'GLEAS' recognition site and the first residue of the core peptide. The cleaved peptide was purified from the leader sequence and PatApr by separation on a Superdex S30 column (GE Healthcare). MS analysis was used to confirm N-terminal cleavage of both TruD and PatD treated peptides (Figure 4.7).

The slow processing by PatApr led us to re-engineer PatE2 to contain a lysine residue after the 'GLEAS' PatApr recognition sequence (PatE2K). PatE2K was treated with both heterocyclases and the additional residue was found to have no detrimental effect on

heterocyclisation. Incubation of the peptide with 1:250 trypsin resulted in complete N-terminal cleavage after just two hours at 37 °C.

Additionally, as an alternative to trypsin which would still allow for the use of lysine and arginine residues in the core peptide, we introduced a TEV protease site 'ENLYFQ' immediately after the GLEAS sequence of PatE2 (PatE2TEV). Tests of the peptide with heterocyclase TruD revealed no detrimental effects on heterocyclisation with two heterocycles formed. The N-terminal leader was removed by addition of 1 mg TEV protease per 10 mg PatE and purified as for PatE2/PatE2K.

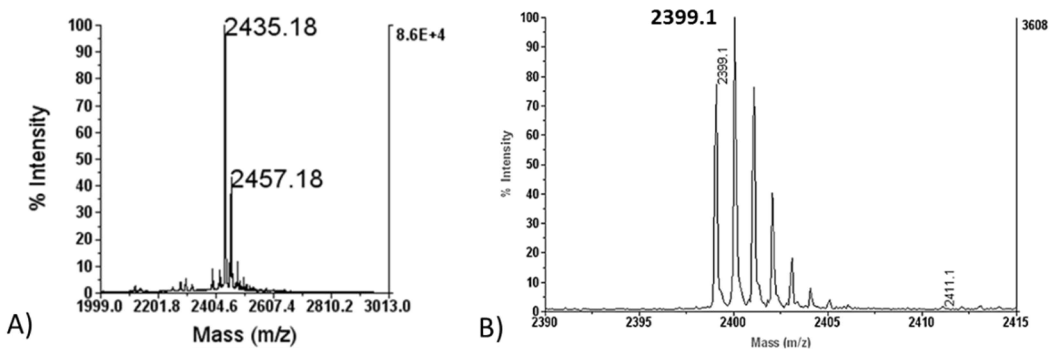


Figure 4.7: MS Analysis of N-terminal Cleavage. MALDI TOF MS spectra of the purified PatE2K peptide following leader sequence removal with trypsin after initial treatment with either A) TruD - $[M+H]^+$ of 2435 and $[M+Na]^+$ of 2457 or B) PatD - $[M+H]^+$ of 2399.

4.3.6 C-terminal Core Peptide Cleavage and Macrocyclisation

Heterocyclised (with either TruD or PatD) and N-terminally cleaved PatE2/E2K were macrocyclised by incubation with PatGmac at 37 °C. Initial reactions were carried out in 150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP and found to be slow (> 200 hours for completion). Switching the reaction to published buffer conditions of 500 mM NaCl, 20 mM bicine pH 8.1 and 5 % DMSO resulted in complete macrocyclisation in under 48 hours [141]. The formation of the macrocycles cyclo[ITAC^{Thn}ITFC^{Thn}] and

cyclo[IT^{Oxn}AC^{Thn}IT^{Oxn}FC^{Thn}], derived from TruD and PatD treatment respectively, were confirmed by MALDI TOF MS (Figure 4.8).

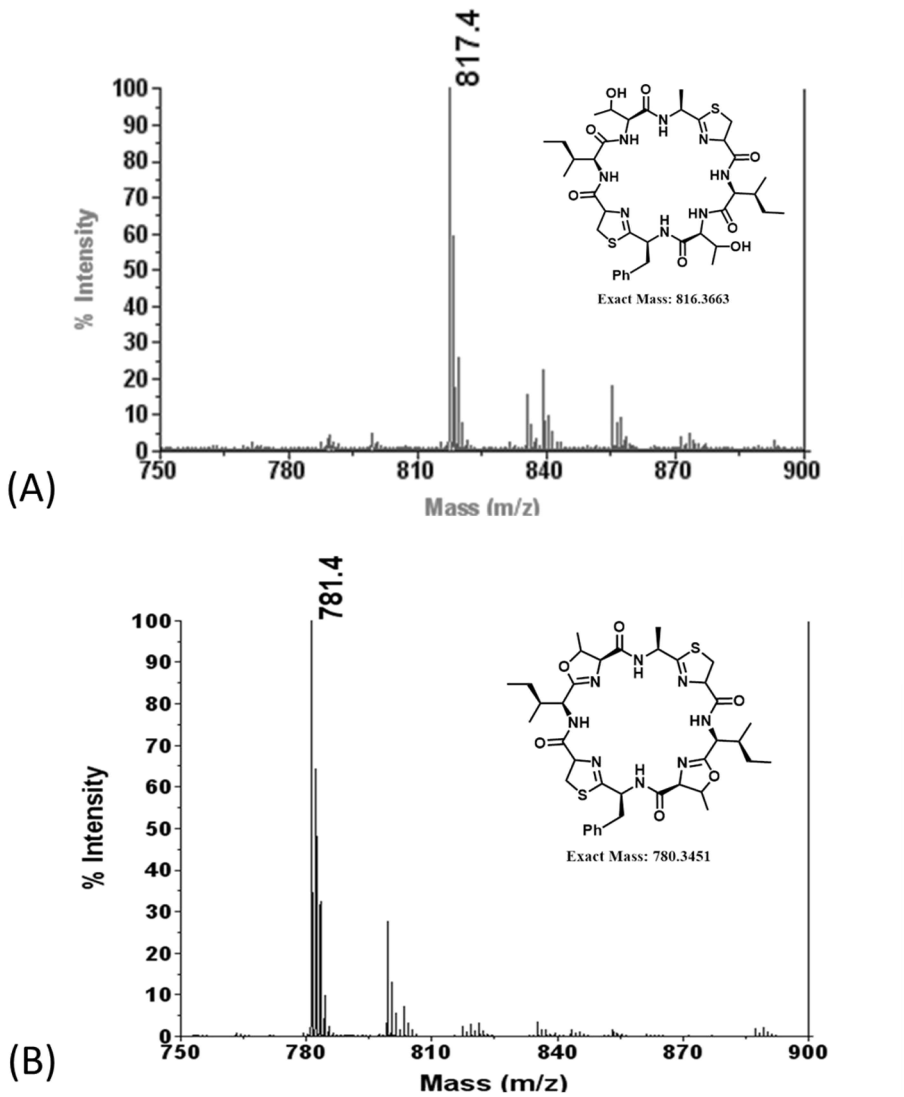


Figure 4.8: MS Analysis of Macrocyclisation. MALDI TOF MS spectrum and associated chemical structures of the macrocyclised peptides A) cyclo[ITAC^{Thn}ITFC^{Thn}] ([M+H]$^{+}$ = 817 Da) and B) cyclo[IT^{Oxn}AC^{Thn}IT^{Oxn}FC^{Thn}] ([M+H]$^{+}$ = 781).

4.3.7 Cyclic Peptide Purification

The macrocycles were purified by HPLC by passage over a C_4 column then subsequent passage of macrocycle peaks over a C_{18} column. The sample eluted off both columns at between 70 and 90 % methanol (Figure 4.9). Peak fractions were identified by MALDI-MS.

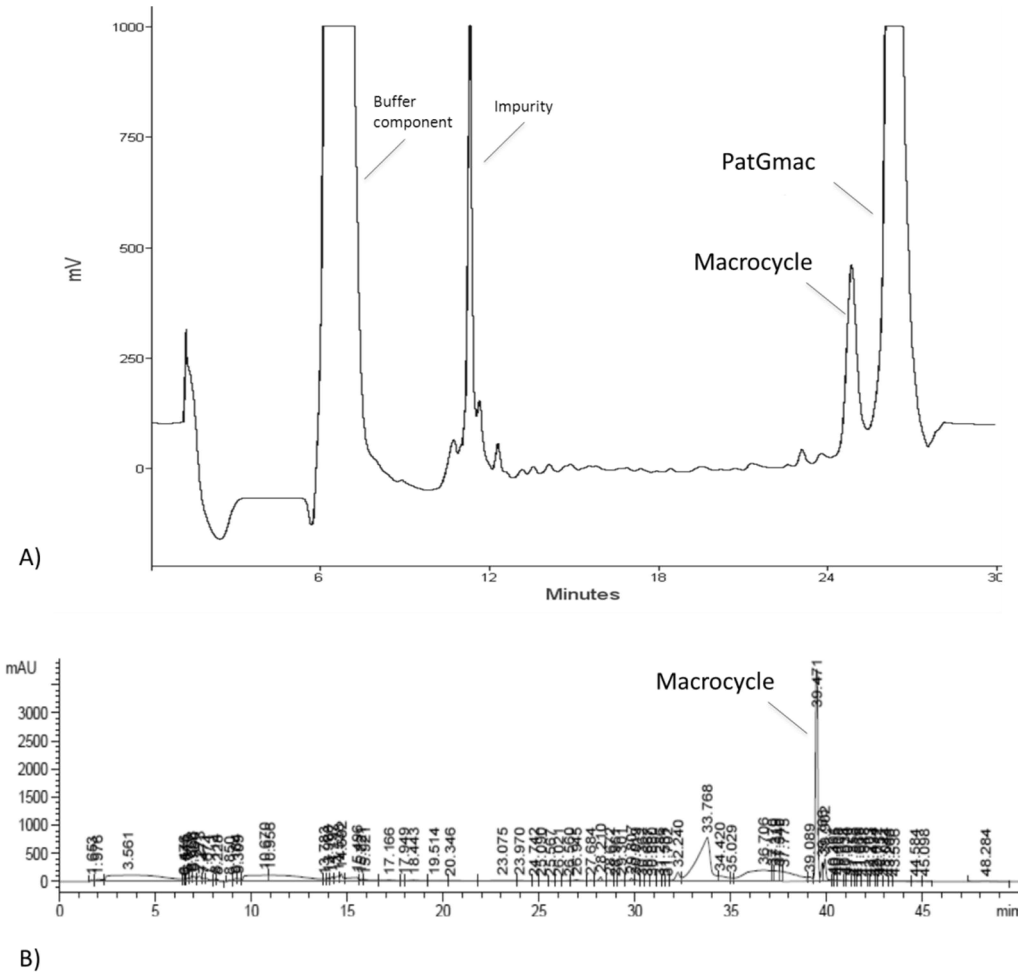


Figure 4.9: HPLC Purification of Macrocycle. HPLC chromatogram of the macrocycle cyclo[ITAC^{Thn}ITFC^{Thn}] when run over (A) an initial C_4 column followed by (B) a subsequent C_{18} column.

4.3.8 Mass Fragmentation

Mass fragmentation of both the TruD and PatD treated PatE2K derived macrocycles (cyclo[ITAC^{Thn}ITFC^{Thn}] and cyclo[IT^{Oxn}AC^{Thn}IT^{Oxn}FC^{Thn}] respectively) was carried out by Dr Matt Fuszard (Figure 4.10 and 4.11 respectively). The fragments were assigned based on the hypothesised structures and were confirmed to correspond to the expected macrocycles.

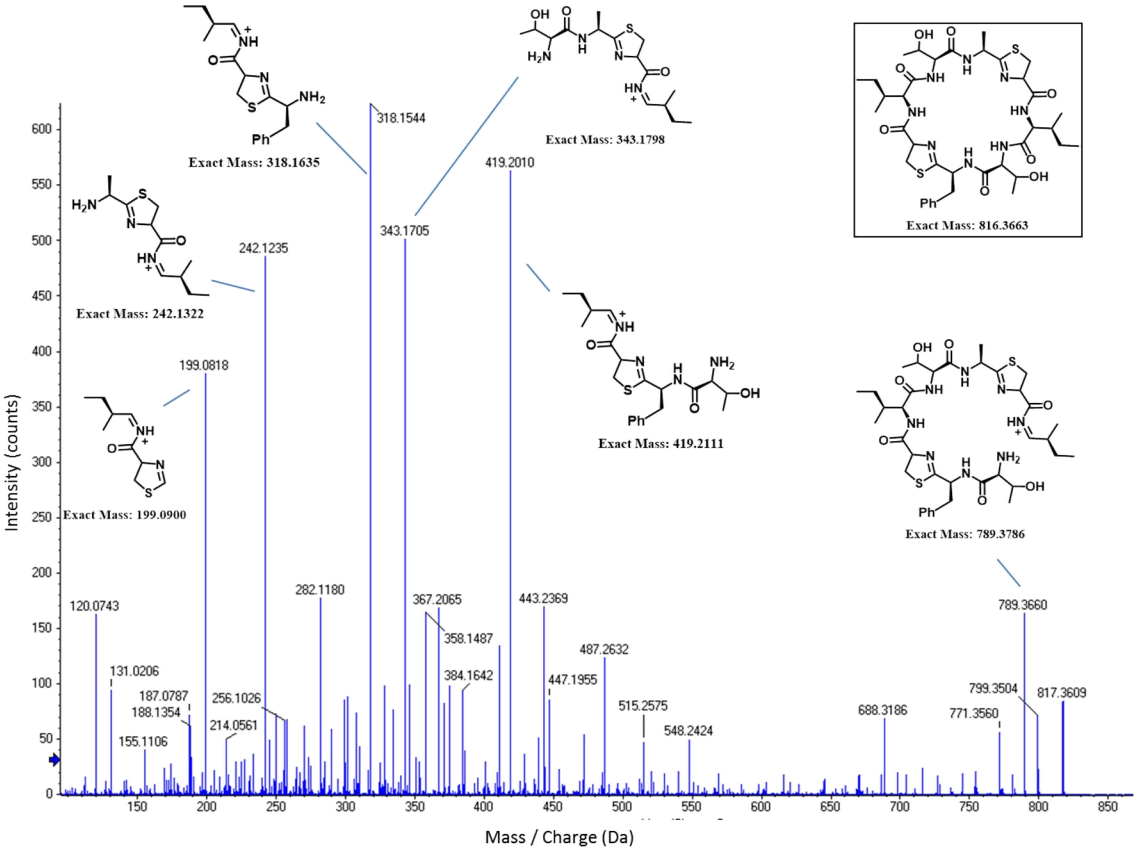


Figure 4.10: Mass Fragmentation of cyclo[ITAC^{Thn}ITFC^{Thn}]. LC/MS/MS fragmentation of cyclo[ITAC^{Thn}ITFC^{Thn}], a macrocycle derived from TruD-treated PatE2K, confirming the presence of two heterocycles, one in Pos1 and the other in Pos5.

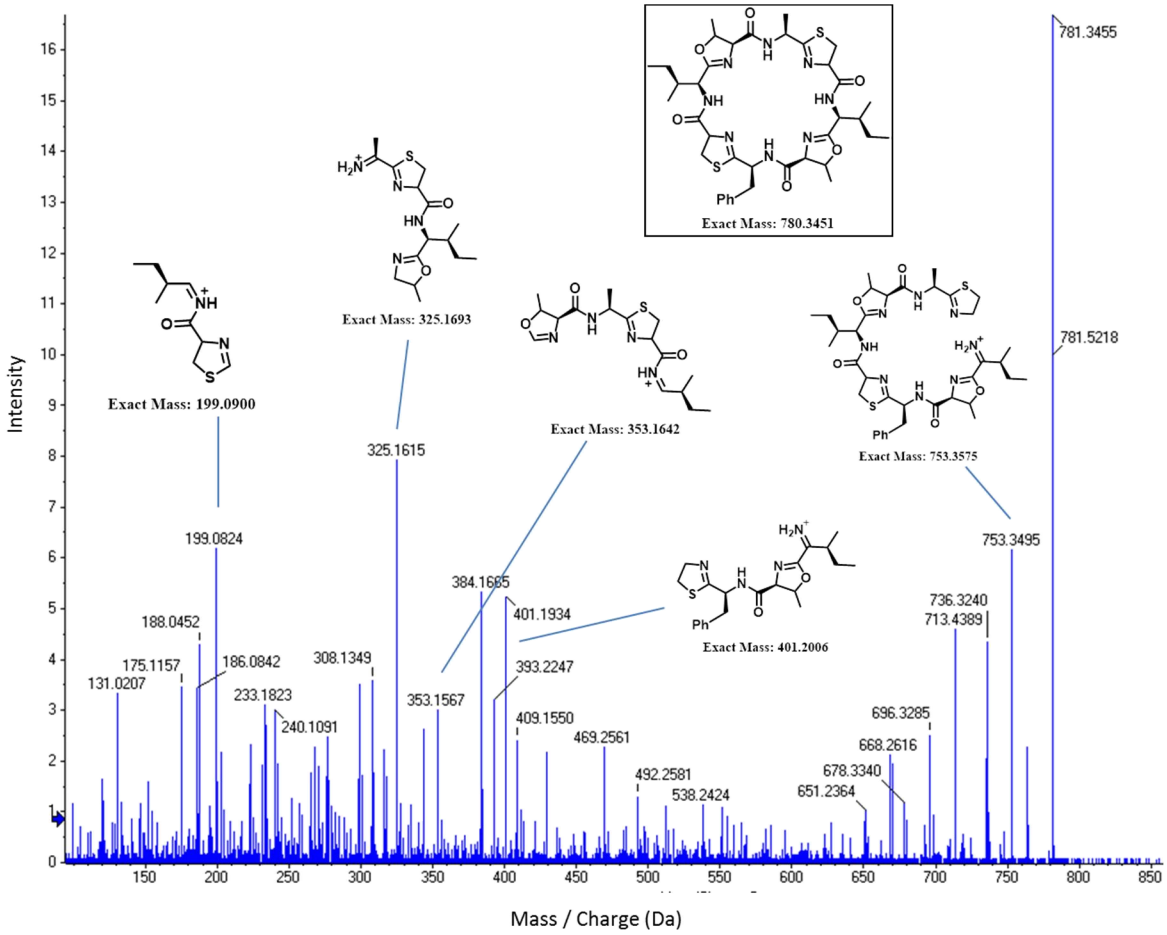


Figure 4.11: Mass Fragmentation of cyclo[IT^{Oxn}AC^{Thn}IT^{Oxn}FC^{Thn}]. LC/MS/MS fragmentation of cyclo[IT^{Oxn}AC^{Thn}IT^{Oxn}FC^{Thn}], a macrocycle derived from PatD-treated PatE2K, confirming the presence of four heterocycles in positions 1, 3, 5 and 7.

4.3.9 NMR Spectroscopy

Preliminary ^{1}H NMR spectroscopic analysis was carried out on the TruD-treated PatE2K macrocycle cyclo[ITAC^{Thn}ITFC^{Thn}] by Dr. Uli Schwarz-Linek. The sample was run in 20 mM sodium phosphate pH 6.8. Analysis of the spectra and assigning of peaks appears to confirm the structure of the expected macrocycle (Figure 4.12). Water peak suppression was applied to the spectra (4.5-5ppm).

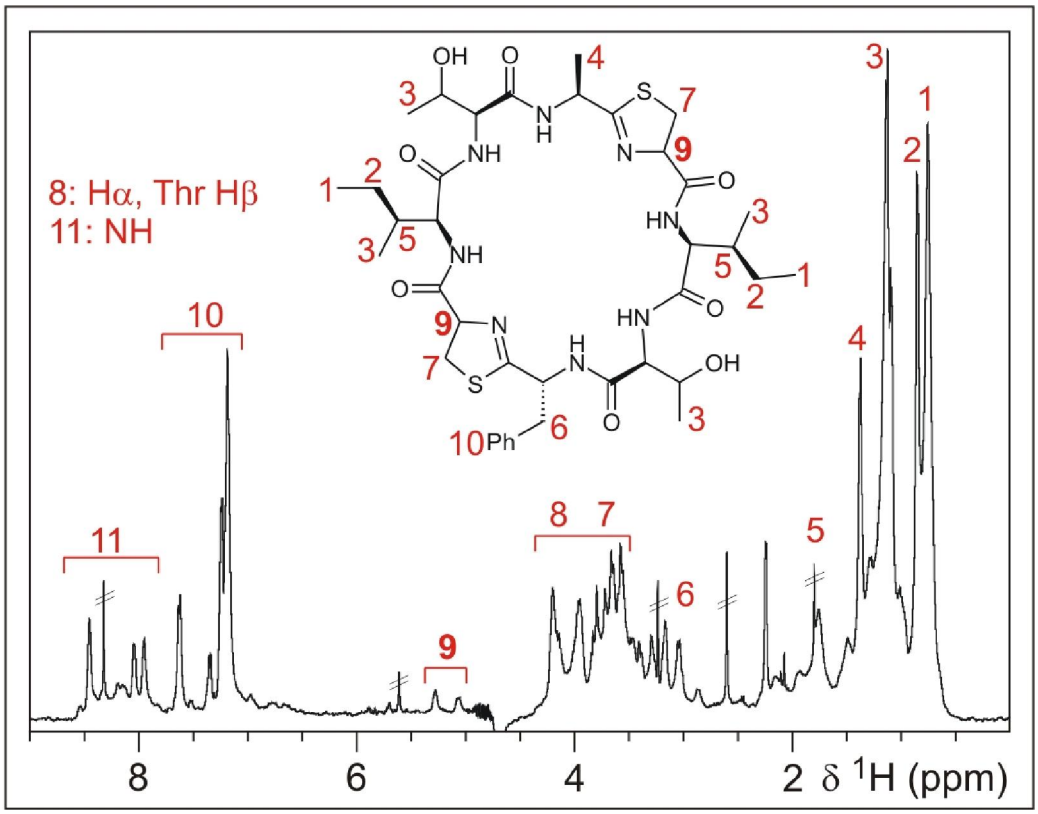


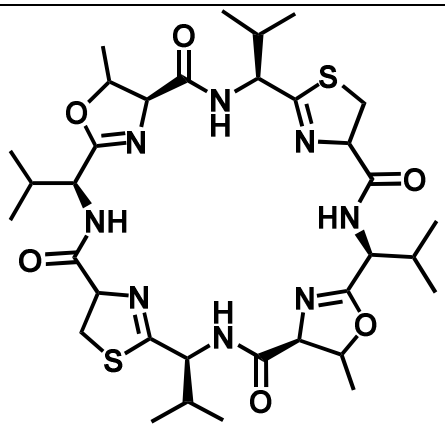
Figure 4.12: ^{1}H NMR Spectrum of cyclo[ITAC^{Thn}ITFC^{Thn}]. ^{1}H NMR spectrum of cyclo[ITAC^{Thn}ITFC^{Thn}], a macrocycle derived from TruD-treated PatE2K. Peaks have been assigned which correspond to the hydrogens as displayed on chemical structure. Small molecule peaks have been discarded by double strikethrough. Water suppression has been implemented between 4.5 and 5.0 ppm.

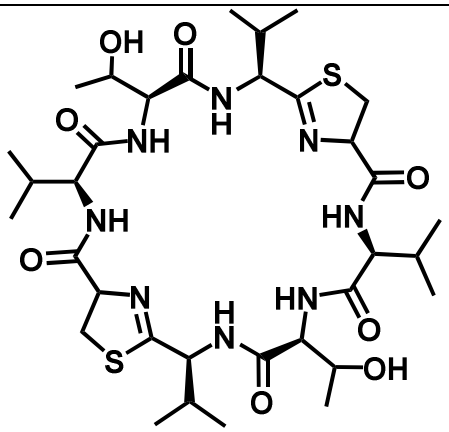
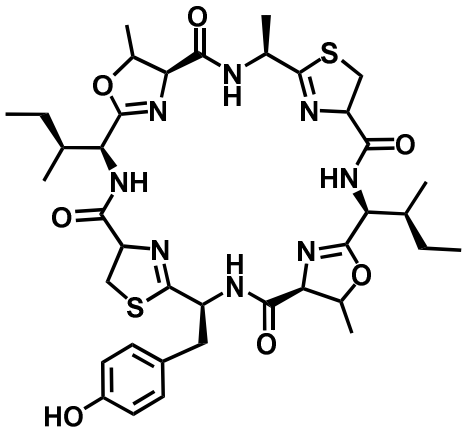
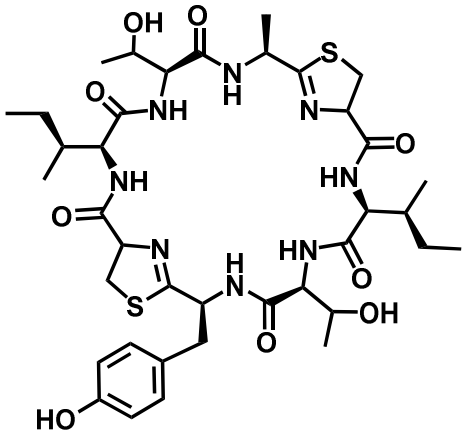
4.3.10 Further Compounds

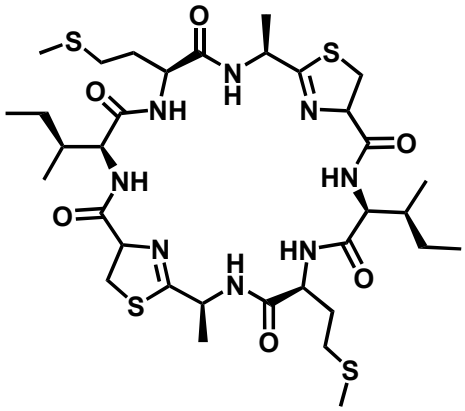
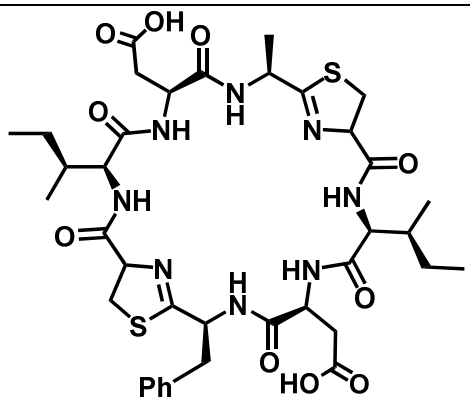
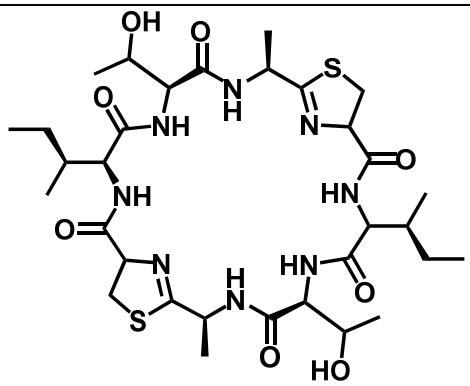
In order to further explore the diversity and the flexibility of the *in vitro* process, several variants of PatE2K were cloned, expressed, purified and taken through the process. These samples varied only in the core peptide sequence which, along with their subsequent macrocyclic compounds, are summarised in Table 4.3.

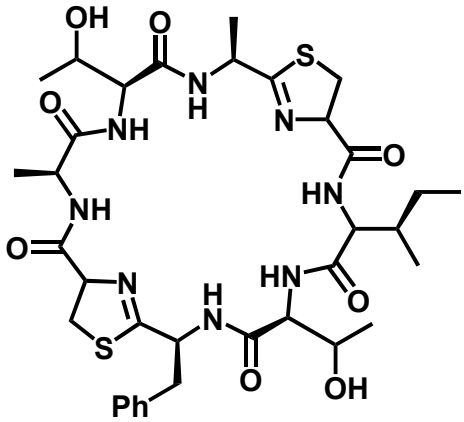
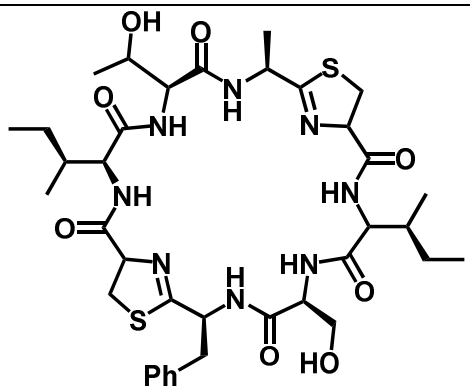
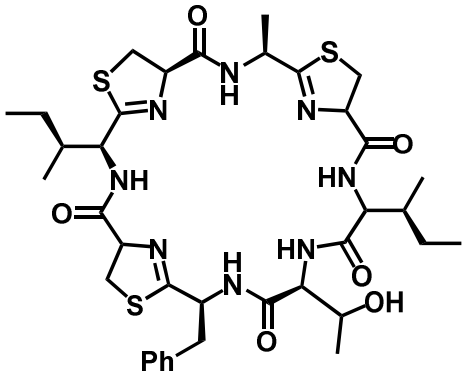
The PatE2K variants were expressed and purified in the same way as PatE2K. Each was taken through the *in vitro* process with either PatD or TruD used for heterocyclisation, the N-terminal cleavage carried out by trypsin and finally macrocyclisation with PatGmac. The final compounds were HPLC-purified.

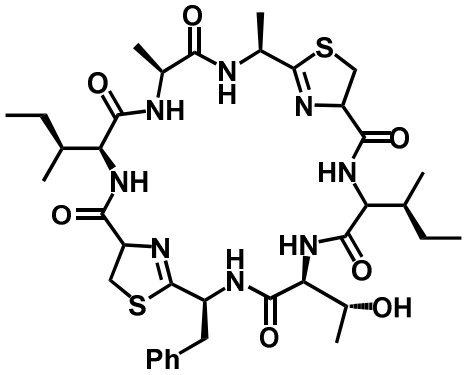
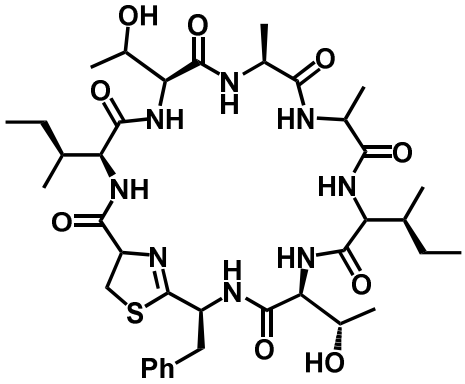
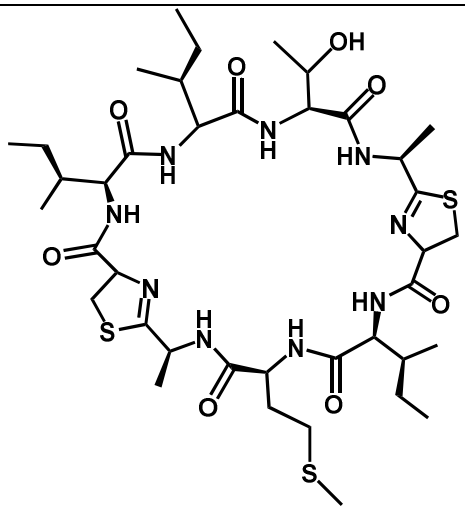
MALDI-TOF-MS was used to confirm the correct mass of each compound and mass fragmentation was carried out on a representative selection of the compounds to confirm the *in vitro* pathway was accurate. All MS and fragmentation data can be found in Appendix D.

Core Peptide Sequence	Heterocyclisation Treatment	Expected Chemical Structure	Appendix Data
VTVCVTVC	+ PatD	 cyclo[VT^{Oxn}VC^{Thn}VT^{Oxn}VC^{Thn}]	MS, MSMS

VTVCVTVC	+ TruD	 cyclo[VTVC^{Thn}VTVC^{Thn}]	MS, MSMS
ITACITYC	+ PatD	 cyclo[IT^{Oxn}AC^{Thn}IT^{Oxn}YC^{Thn}]	MS, MSMS
ITACITYC	+ TruD	 cyclo[ITAC^{Thn}ITYC^{Thn}]	MSMS

IMACIMAC	+ TruD	 cyclo[IMAC^{Thn}IMAC^{Thn}]	MS, MSMS
IDACIDFC	+ TruD	 cyclo[IDAC^{Thn}IDFC^{Thn}]	MSMS
ITACITAC	+ TruD	 cyclo[ITAC^{Thn}ITAC^{Thn}]	MS, MSMS

ATACITFC	+ TruD	 cyclo[ATAC^{Thn}ITFC^{Thn}]	MS, MSMS
ITACISFC	+ TruD	 cyclo[ITAC^{Thn}ISFC^{Thn}]	MS, MSMS
ICACITFC	+ TruD (3het)	 cyclo[IC^{Thn}AC^{Thn}ITFC^{Thn}]	MS

IAACITFC	+ TruD	 cyclo[IAAC^{Thn}ITFC^{Thn}]	MS
ITAAITFC	+ TruD	 cyclo[ITAAITFC^{Thn}]	MS
IITACIMAC	+ TruD	 cyclo[IITAC^{Thn}IMAC^{Thn}]	MS, MSMS

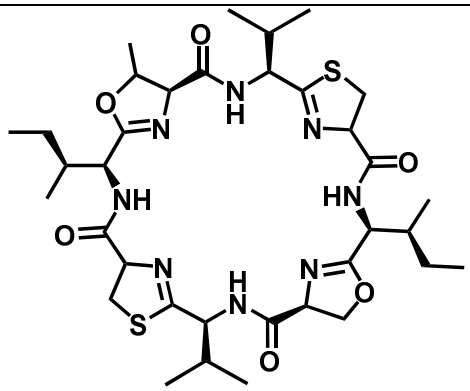
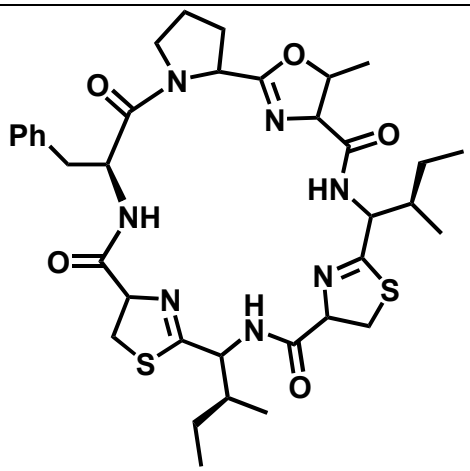
ITVCISVC	+ PatD	 cyclo[IT^{Oxn}VC^{Thn}IS^{Oxn}VC^{Thn}]	MS
ICFPTIC	+ PatD	 cyclo[IC^{Thn}FPT^{Oxn}IC^{Thn}]	MS

Table 4.3: *In vitro* Biosynthesis Derived Cyclic Peptides. Core peptide sequence, heterocyclisation treatment and final chemical structure of fifteen cyclic peptides derived from the reconstituted patellamide pathway. Additional data listed can be found in Appendix D.

^{1}H Nuclear Magnetic Resonance (NMR) spectroscopic analysis was carried out on the cyclo[IMAC^{Thn}IMAC^{Thn}] macrocycle in 100 % deuterated methanol by Dr. Uli Schwarz-Linek (Figure 4.13). Proton peaks were assigned and are annotated in the chemical structure. The use of deuterated methanol produces a sharper NMR spectrum than that of the PatE2K macrocyclic product run in 20 mM sodium phosphate pH 6.0 (Figure 4.12).

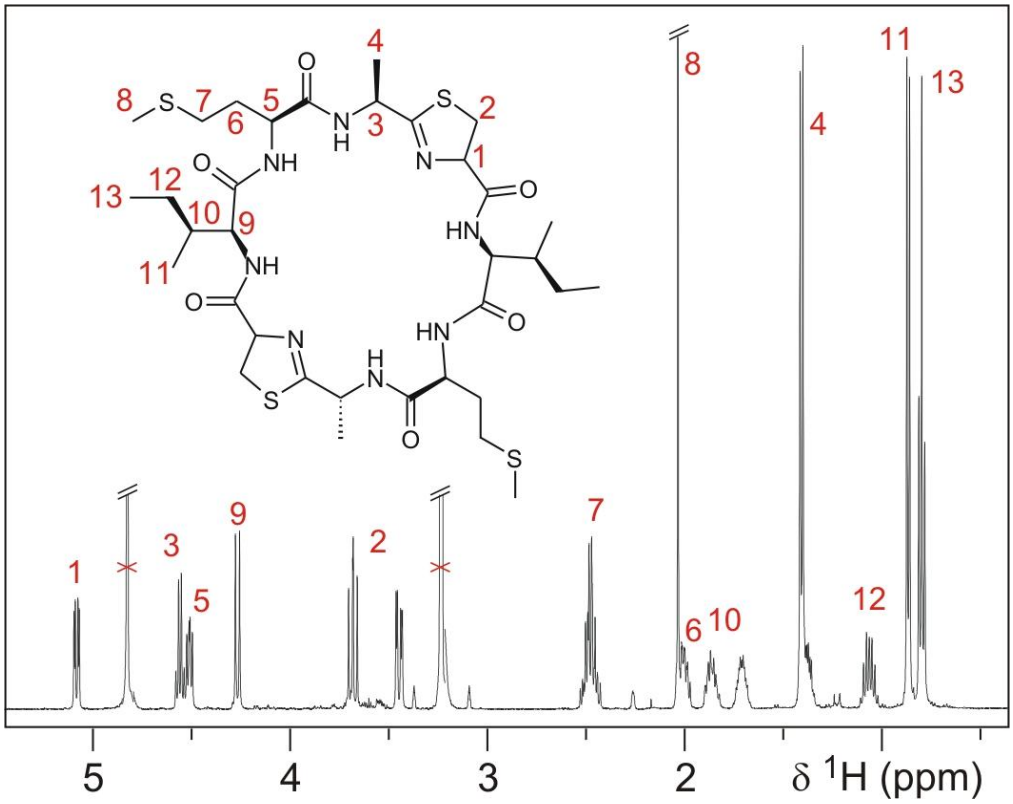


Figure 4.13: ^{1}H NMR Spectrum of cyclo[IMAC^{Thn}IMAC^{Thn}]. ^{1}H NMR spectrum of the cyclic peptide cyclo[IMAC^{Thn}IMAC^{Thn}] derived from TruD treated PatE(IMACIMAC). Peaks have been assigned which correspond to the hydrogens as displayed on chemical structure. Small molecule peaks have been discarded by double strikethrough.

4.4 Discussion

Using isolated enzymes from the patellamide and related pathways, cyclic peptides have been produced by an *in vitro* biosynthesis on a milligram scale.

The heterocyclisation step is a fast process and can be completed in less than two hours with the ability to recycle the heterocyclase enzyme for future reactions. Recycling of the heterocyclases is an important feature as they show the lowest expression levels (< 40 mg L^{-1} culture) but also that enzyme recycling is a cost efficient method in any bio-industrial process.

The N-terminal cleavage of the precursor peptide by PatApr is particularly slow with a minimum of 200 hour incubation required for full cleavage. This is most likely due to the requirement of the leader to remain bound in order for the heterocyclase to work. *In vivo* both of these enzymes will be present at the same time, but *in vitro* this can be controlled by the timing of enzyme addition. In order to overcome the slow processing, the PatE2 precursor peptide was re-engineered to introduce either a lysine residue or the TEV protease recognition signal 'ENLYFQ' between the PatA recognition sequence 'GLEAS' and the core peptide. Treatment of PatE2K and PatE2TEV with both TruD and PatD confirmed that the additional residue(s) had no effect on heterocyclase activity. N-terminal cleavage of the core peptide can now be efficiently carried out using trypsin or TEV protease in approximately two hours. This represents a 100 fold increase in reaction rates.

The final macrocyclisation by PatGmac under our initial conditions was slow (> 200 hours), however optimisation of the conditions with an increase in both salt concentration and buffer pH plus the addition of 5 % DMSO increased the reaction rate allowing completion in under 48 hours.

Overall, the steps from precursor peptide to macrocycle have been reduced from approximately 400 hours down to 50 hours. It is possible that with further optimisation we can reduce this further.

The diversity of the pathway has been demonstrated by processing *in vitro* a range of PatE precursor peptides containing broad diversity in the core peptide sequence and thus generating multiple cyclic peptides.

Using this process, 1 to 5 mg macrocycle from 100 mg precursor peptide has been achieved, which equates to around 5 to 10 mg macrocycle L^{-1} total cell culture. These yields are based on dried solid following HPLC purification which we appreciate has limitations (salts, etc). A more accurate quantification method is currently being investigated by Greg Mann, University of St Andrews and initial experiments suggest our dried solid contains 40 - 60 % macrocycle. This confirms that our process is still on the mg scale. In contrast, *in vivo* studies by Tianero *et al.* were only able to generate 5 to 174 µg macrocycle L^{-1} cell culture [103]. This therefore indicates that the *in vitro* process is able to generate larger quantities of macrocycle for use in further studies, e.g. mode of action studies and toxicology screening.

The benefits of this *in vitro* biosynthesis over *in vivo* include that all enzymes can be produced in large quantities and stored at -80 °C with no detrimental effects, allowing multiple reactions to be carried out per enzyme batch, as well as enzyme recycling (as demonstrated for the heterocyclases). Also, the ability to take one precursor peptide and process it in different ways by enzyme switching (TruD/PatD, oxidised/unoxidised (see Chapter 6)) could lead to multiple final products from a solitary starting material. Finally, cyanobactins are known to be toxic to certain types of cells and therefore *in vivo* there is the possibility that they will cause host cell death; the *in vitro* biosynthesis eliminates this risk as only the non-toxic precursor protein is produced *in vivo*.

On the final macrocyclic compounds the oxidation of thiazolines to thiazoles, which appear in the majority of patellamide natural products, was not carried out. It is widely believed that this oxidation is catalysed by the N-terminal domain of PatG which bioinformatic analysis has identified as an oxidase domain and this is studied further in Chapter 6. It has however been discovered that in a macrocycle sample derived from PatE2K (PatD treated) stored at room temperature for four weeks and re-analysed by MALDI-MS, that there is the presence of a minor peak of 777 Da equivalent to a loss of 4 Da. This may correspond to the oxidation of the two thiazolines and may suggest that this process occurs slowly (Figure 4.14).

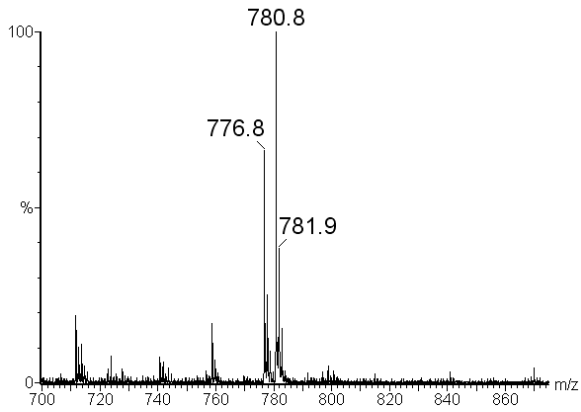


Figure 4.14: Potential Oxidation of Thiazolines. MALDI TOF MS spectrum of cyclo[IT^{Oxn}AC^{Thn}IT^{Oxn}AC^{Thn}] stored at room temperature for 4 weeks showing two peaks; 781 corresponding to the macrocycle $[M+H]^+$ and 777 which may represent an oxidised derivative $[M+H]^+$.

Finally, natural patellamides contain two D-stereoisomers, however at present we have been unable to ascertain the epimeric state of our macrocycles. The process to which epimerisation occurs in natural patellamides is not yet known and may be either an enzymatic reaction or may be spontaneous.

4.5 Conclusions and Future Work

A library of cyclic peptides containing thiazolines and oxazolines have been produced *in vitro* on the mg scale utilising isolated enzymes from the Patellamide (and related) biosynthetic pathway(s). These compounds have been confirmed by mass spectrometry and NMR and their production represents a ten to one hundred fold yield increase over previous *in vivo* studies of the same pathway. The modification of our PatE to accommodate cleavage by the more efficient trypsin protease has reduced processing times by almost 200 hours over the natural protease PatApr. This represents the first stage in our pathway engineering to improve both the speed and yield of cyclic peptide production and we will explore further improvements.

At the time of this work a cyanobactin oxidase had not yet been isolated and confirmed active for use in the pathway. However progress has since being made on this and is described in detail in Chapter 6. Once established fully, the oxidation step can be implemented into the *in vitro* biosynthesis pathway.

Prenylation has been determined to not occur in the patellamide biosynthetic pathway (Chapter 2), yet it should be possible to introduce a level of diversity through the use of related prenyl transferases. The prenylation of tyrosine residues by LynF has already been demonstrated (Section 2.3.9) and as such we aim to implement a prenylation step into our *in vitro* biosynthesis. Obtaining other prenyl transferases which act on different amino acids could also increase diversity.

Epimerisation has still not been confirmed to be either enzymatic or spontaneous but is likely a key component of the patellamide which results in inhibition of target proteins. Studies on epimerisation are being carried out by Greg Mann (University of St Andrews) and if reaction conditions can be identified it could be included in our compound production.

Finally, once a significant library of cyclic peptides has been established, with mg quantities of each, biological assays allowing the generation of structure-activity relationships (SAR) could be undertaken. Logically, the first target would be p-glycoprotein (p-GP) as previous studies have identified that the natural product Patellamide D inhibits p-GP [55].

5. Selenazoline Incorporation into Macrocyclic Peptides

5.1 Introduction

Selenazolines, an alternative to thiazolines where the sulfur is replaced by selenium, pose an intriguing motif for medicinal chemistry, where they can act as thiazoline substitutes whilst retaining similar activities. Kim *et al.* [153] have shown that the introduction of selenazolines in place of thiazolines in *O*-GlcNAcase inhibitors does not significantly alter activity in human cells. Remarkably, this is despite a 70 fold decrease in activity in protein assays, where the thiazoline form is more potent. This is believed to be a result of selenazoline containing compounds been internalised by cells more efficiently than their thiazoline containing counterparts [153].

Compounds containing selenazolines and their oxidised derivatives 1,3-selenazoles are chemically difficult to synthesise and also require the use of toxic precursors such as selenocyanide [154] or seleno-urea [155].

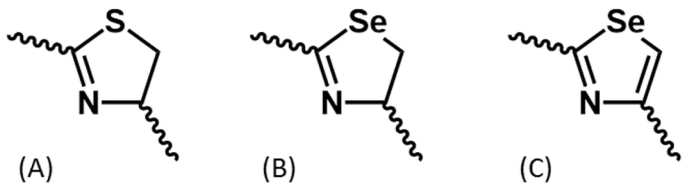


Figure 5.1: Selenium Containing Heterocycles. Chemical structures of A) thiazoline B) selenazoline and C) selenazole.

Tao *et al.* (2011) have chemically synthesised patellamide-like macrocycles containing selenazoles [156] which target the p-glycoprotein as has been demonstrated for Patellamide D [55] (Figure 5.2). The process of making these compounds does however lead to significant risks in addition to the toxic precursors, with H_2Se being produced, which can decompose to H_2 gas [156]. The ability to produce these macrocycles using a biological process with significantly reduced risks would be of high interest.

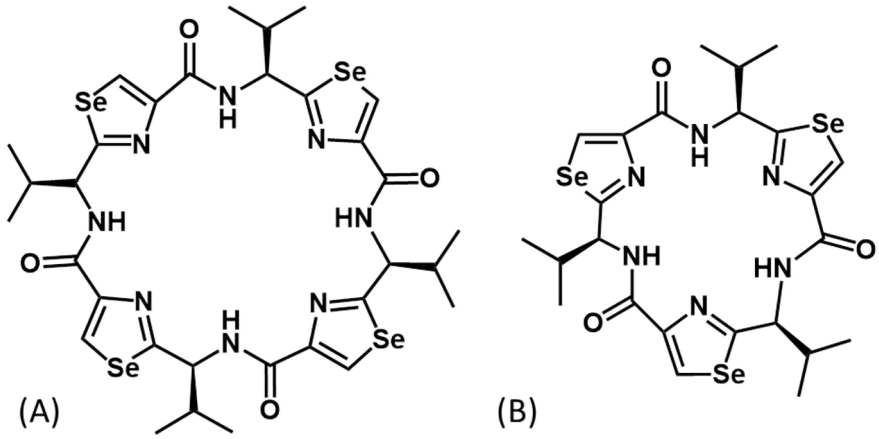


Figure 5.2: Chemical Structures of Selenazole Macrocycles. Structures of macrocyclic compounds containing selenazoles chemically synthesised by Tao *et al*. A) QZ60Se-SSSS B) QZ59Se-SSS [156].

Cyanobactin heterocyclases PatD and TruD process cysteine residues into thiazolines and it is possible that they may accept selenocysteine as a substrate to generate selenazolines. If successful, the resultant peptides could be used to make macrocycles using additional cyanobactin biosynthetic enzymes.

Salgado *et al.* (2011) [157] have highlighted a method of incorporating the selenocysteine amino acid into proteins expressed in *E. coli* cells by growing the cells on medium containing the selenocysteine precursor seleno-L-cystine.

We proposed to utilise the Salgado method to generate a PatE precursor peptide containing selenocysteine in the core peptide sequence. We would then use this SeCys PatE to assess the substrate tolerance of the heterocyclase enzyme TruD, i.e. can it heterocyclise SeCys to form selenazolines, and if successful, subject the SeCys PatE to the full *in vitro* process described in Chapter 4 to generate selenazoline containing cyclic peptides.

The work in this chapter was carried out in collaboration with Falk Morawitz (University of St. Andrews, M. Chem project student)

5.2 Materials and Methods

5.2.1 Expression and Purification

Selenocysteine-labelled PatE2K was expressed in BL21 (DE3) *E. coli* cells, cultures of which were grown in a minimal medium supplemented with glucose-free nutrient mix (Molecular Dimensions) and 5 % glycerol (see Appendix A.8 for media and buffer compositions). The medium was inoculated with overnight culture grown in Luria-Bertani (LB) medium which was subsequently washed three times with minimal medium. The cultures were grown to an optical density (OD) at 600 nm wavelength of 0.6 then an amino acid mix was added. Fifteen minutes after amino acid mix addition, 0.1 g L^{-1} seleno-L-cystine and 1 mM isopropylthio-β-D-galactoside (IPTG) were added. The cultures were then grown at 37 °C overnight driving the protein to inclusion bodies before harvesting by centrifugation. SeCys PatE2K was purified using the same protocols as outlined for native PatE2K in section 4.2.1. Enzymes used in the reactions were purified as stated in section 4.2.2

5.2.2 Heterocyclisation

100 µM PatE2K-SeCys was incubated with 0.5 µM TruD in 150 mM NaCl, 10 mM HEPES pH 7.4, 5 mM ATP, 5 mM $MgCl_2$ at 37 °C for 16 hours to drive heterocyclisation to completion. The heterocyclised PatE2K-SeCys was purified from TruD on an S75 gel filtration column (GE Healthcare). The heterocyclisation state of the peptide was analysed by MS.

5.2.3 N-terminal Cleavage

PatE2K-SeCys, post treatment with TruD, was N-terminally cleaved by incubation with 1:500 mass to mass ratio of PatE to bovine trypsin (Sigma) at 37 °C for four hours. The cleaved peptide was separated from the leader sequence and trypsin by passage over

a Superdex S30 column (GE Healthcare) pre-equilibrated in 150 mM NaCl, 20 mM Bicine pH 8.1

5.2.4 Macrocyclisation, Purification and Characterisation

100 µM heterocyclised and N-terminally cleaved PatE2K-SeCys was incubated with 20 µM PatGmac for 48 hours at 37 °C in 20 mM bicine pH 8.1, 500 mM NaCl, 5% DMSO to drive the completion of macrocyclisation. The final SeCys macrocycle was purified from PatGmac by reverse-phase HPLC using the same methods as for native peptides (Section 4.2.6). The SeCys macrocycle was analysed by mass fragmentation as for the native peptides (Section 4.2.7).

5.3 Results

5.3.1 Expression and Purification

PatE2K-SeCys precursor peptide was overexpressed with a C-terminal His_6-tag in BL21 (DE3) *E. coli* using a minimal medium containing seleno-L-cystine. An indicator of selenium incorporation was the distinct presence of a red colour in the culture. The peptide was solubilised from inclusion bodies in urea buffer, isolated by His_6-tag and refolded using DTT. The peptide elutes off gel filtration as a single peak. The final protein yield was 22 mg L^{-1} culture (Figure 5.3). Mass spectrometry analysis of the purified protein confirmed its identity and indicated that the major species contains a single selenium atom while minor species contain no or two seleniums.

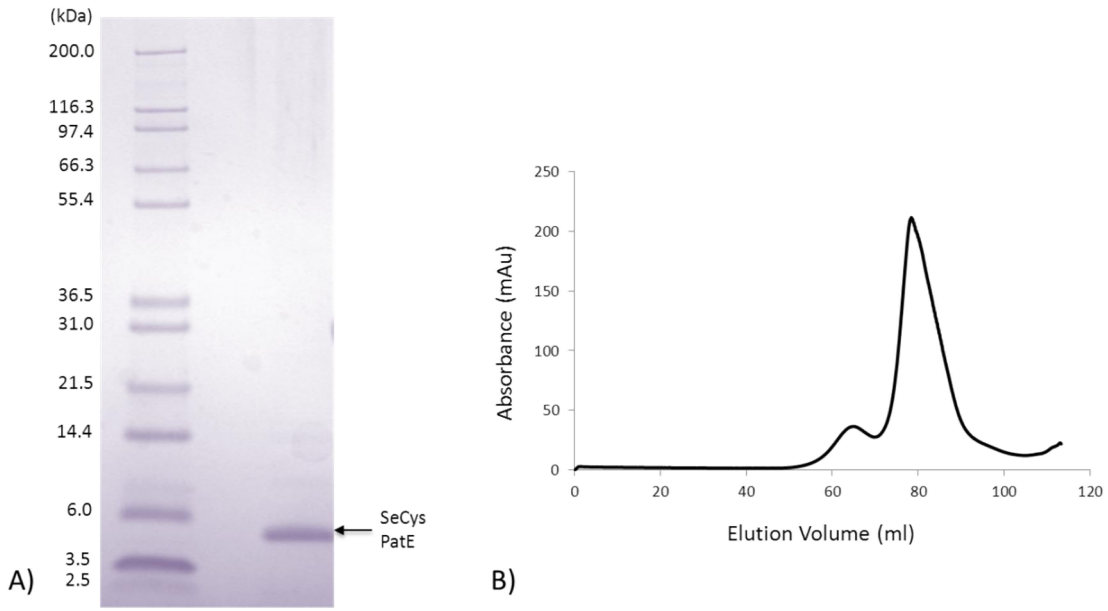


Figure 5.3: Purification of PatE2K-SeCys. (A) SDS-PAGE analysis of final PatE2K-SeCys purified sample showing >98% purity. (B) Gel filtration chromatograph for PatE2K-SeCys (Note as with PatE2K, due to the unfolded nature of the peptide, the protein travels further on SDS-PAGE and elutes earlier on gel filtration).

5.3.2 Heterocyclisation

Incubation of PatE2K-SeCys with TruD results in a loss of 36 Da by mass spec (Figure 5.4). This corresponds to the loss of two water molecules as a consequence of two heterocycles forming. As the majority of the starting substrate contains one selenocysteine, it can be confirmed that with the formation of two heterocycles, one will be a thiazoline while the other will be a selenazoline.

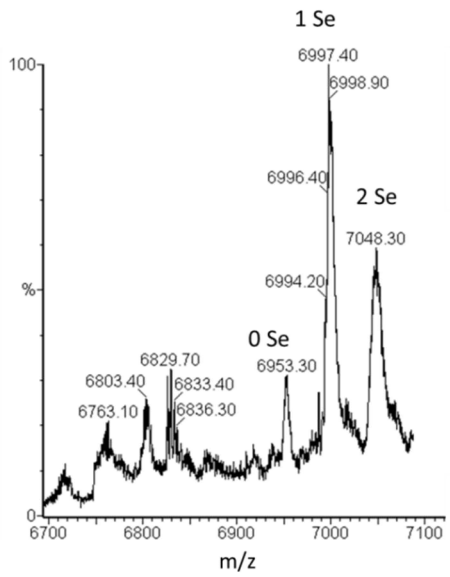


Fig 5.4: Heterocyclisation of PatE2K-SeCys. MALDI-TOF-MS analysis of the peptide product from reaction of PatE2K-SeCys with TruD showing a primary mass of 6997 $[M+H]^+$ corresponding to two water losses for the majority one selenium species. The minor peaks correspond to two water losses for both no (6953 $[M+H]^+$) and two (7048 $[M+H]^+$) selenium species.

5.3.3 N-terminal Cleavage

Initially, trials of SeCys derived PatE2 (with no Lys) were hampered by the fact that the slow rate for PatApr to N-terminally cleave (200 hours for native PatE2) resulted in precipitation of the peptide. PatE2K-SeCys however could be cleaved efficiently in just

over four hours with only a small amount (< 5 %) of precipitation. The cleaved peptide was purified and confirmed by MS (Figure 5.5).

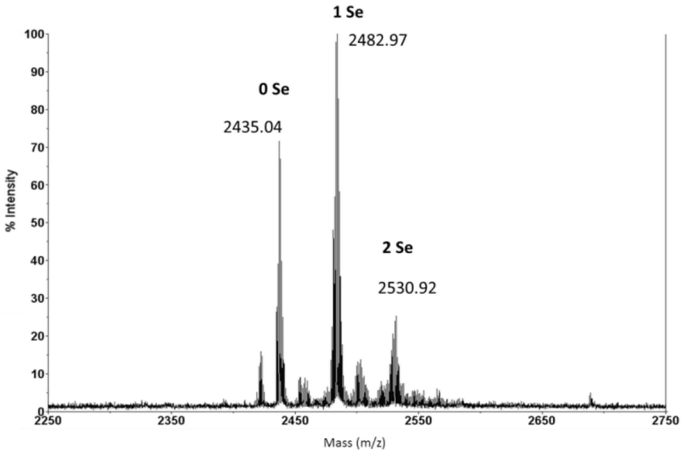


Figure 5.5: N-terminal Cleavage of PatE2K-SeCys. MALDI-MS analysis of the peptide product from reaction of TruD treated PatE2K-SeCys with trypsin showing a primary mass of 2483 $[M+H]^+$ corresponding to N-terminally cleaved peptide containing one selenium. The minor peaks correspond to cleaved peptide containing no ($[M+H]^+$ of 2435) and two ($[M+H]^+$ of 2531) seleniums.

5.3.4 Macrocyclisation

The TruD and trypsin treated PatE2K-SeCys sample was finally macrocyclised by incubation with PatGmac. MS analysis confirms the removal of 18 Da corresponding to the water loss which occurs upon macrocyclisation (Figure 5.6). The majority of the sample contains a single selenium atom and it is likely that two different macrocycles have formed, with Se at either position 1 or 5 (Figure 5.7). There are also low levels of macrocycles containing 0 and 2 selenium atoms.

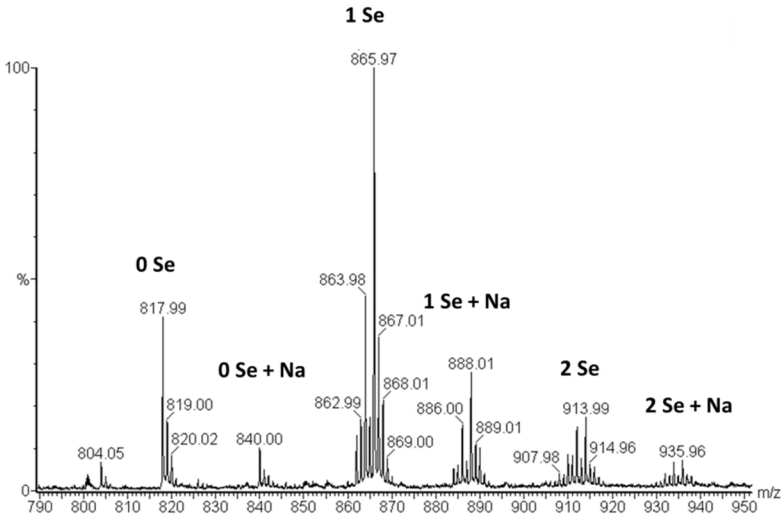


Figure 5.6: Macrocyclisation of PatE2K-SeCys. MALDI-MS analysis of PatE2K-SeCys following treatment with TruD, Trypsin and PatGmac. The primary peak represents a macrocycle containing a single selenium ($[M+H]^+$ of 866 and $[M+Na]^+$ of 888). Minor peaks also exist for no selenium ($[M+H]^+$ of 818, $[M+Na]^+$ of 840) and two selenium ($[M+H]^+$ of 914, $[M+Na]^+$ of 936) containing macrocycles

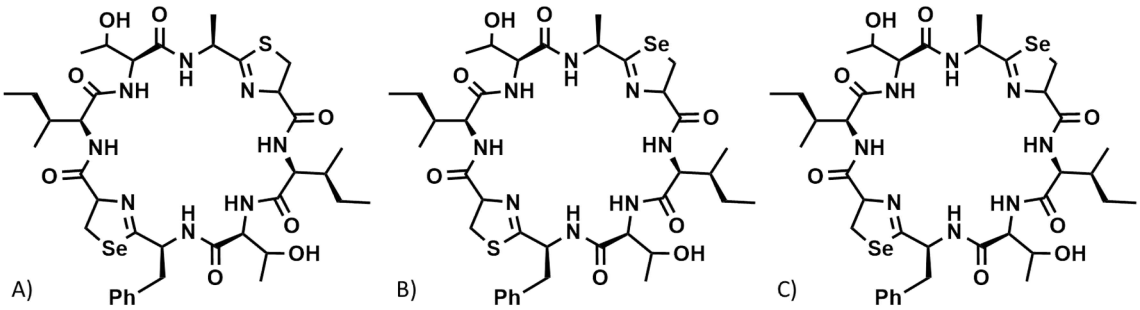


Figure 5.7: Chemical Structures of Selenazoline Containing Cyclic Peptides. Structures hypothesised based on MS data for the PatE2K modified core peptide when grown in SeCys media. Macrocycles (A) cyclo[ITAC^{Thn}ITF(SeCys)^{Sen}] (B cyclo[ITA(SeCys)^{Sen}ITFC^{Thn}]) and (C) cyclo[ITA(SeCys)^{Sen} ITF(SeCys)^{Sen}].

5.3.5 Cyclic Peptide Purification

Selenazoline containing macrocycles were purified by reverse-phase HPLC in 100 % methanol (Figure 5.8). Macrocycles containing no, one and two seleniums eluted off the C_{18} column as separate peaks and their identity confirmed by MS.

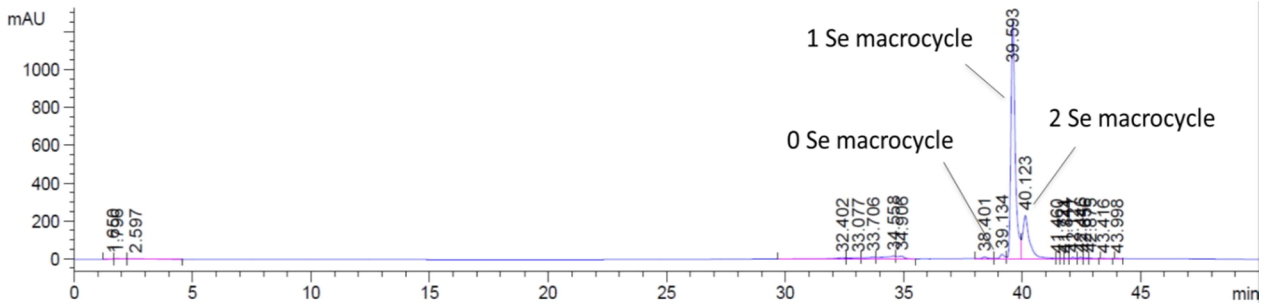


Figure 5.8: SeCys Macrocycle Purification. Reverse-phase HPLC chromatogram of the purification of SeCys macrocycles on a C_{18} column.

5.3.6 Mass Fragmentation

LC/MS/MS spectra were recorded for both one and two selenium containing macrocycles by Dr. Matt Fuszard (University of St Andrews). The single selenium containing macrocycles fragment in a manner which confirms the presence of two species; one containing a selenazoline in the P1 position with a thiazoline in P5, cyclo[ITAC^{Thn}ITF(SeCys)^{Sen}], while the second species has the selenazoline and thiazoline switched around, cyclo[ITA(SeCys)^{Sen}ITFC^{Thn}] (Figure 5.9). The macrocycles containing two selenium atoms fragments as would be expected for two selenazolines per macrocycle, cyclo[ITA(SeCys)^{Sen}ITF(SeCys)^{Sen}] (Figure 5.10).

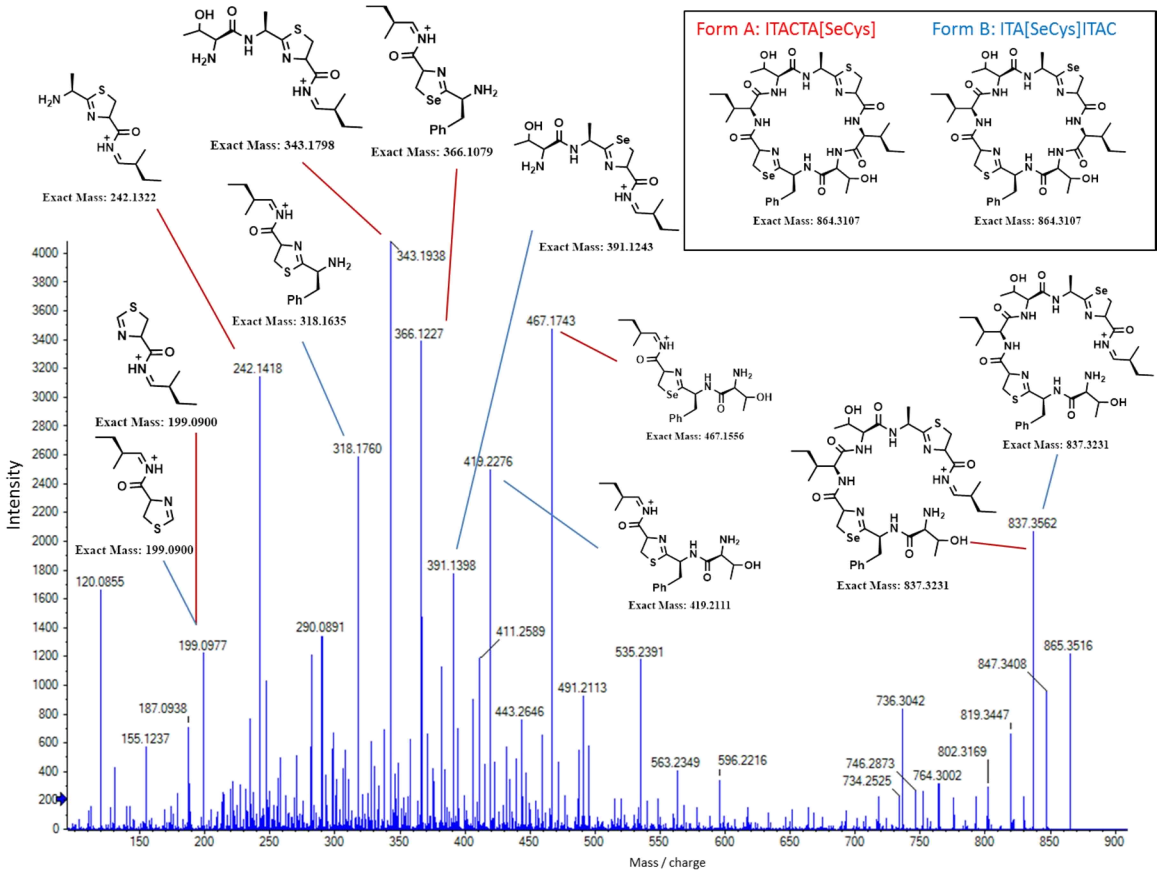


Figure 5.9: Mass Fragmentation of Single Selenazoline Containing Macrocycle. LC/MS/MS spectrum of mixed species of single selenazoline containing macrocycles showing fragments which confirm the identity of two species; Form A: cyclo[ITAC^{Thn}ITF(SeCys)^{Sen}] and Form B: cyclo[ITA(SeCys)^{Sen}ITFC^{Thn}].

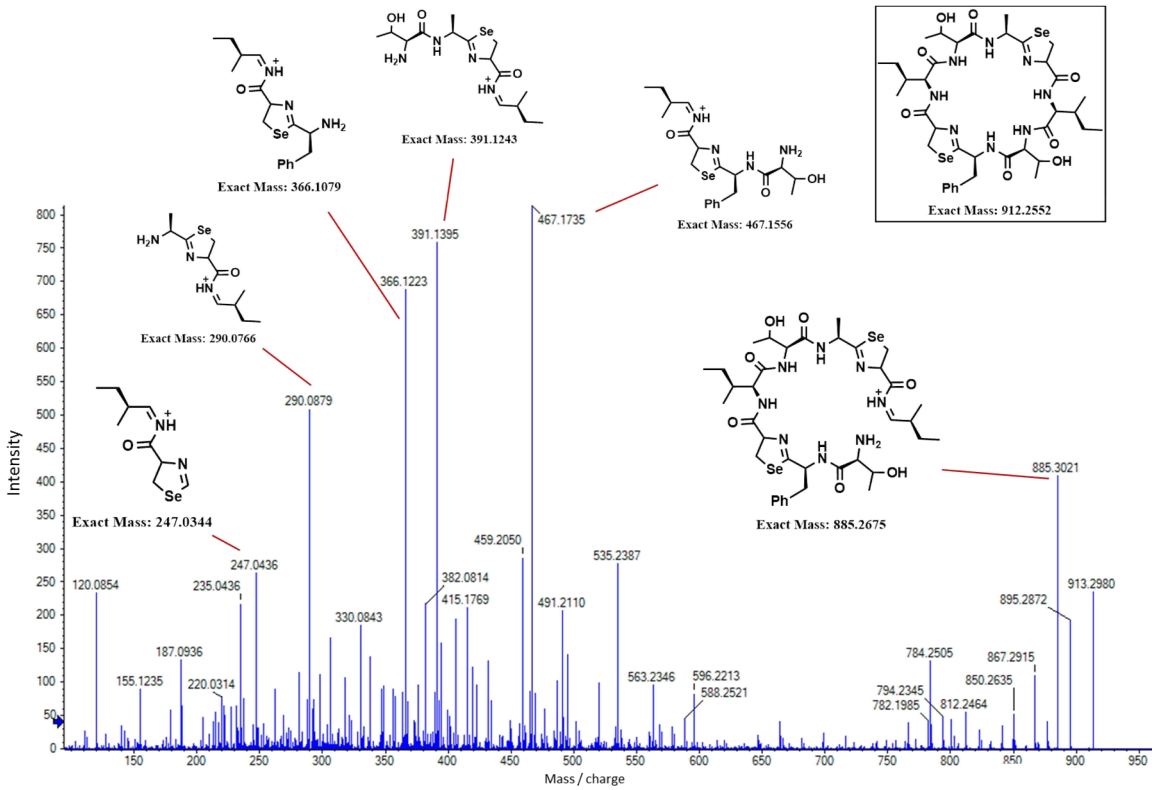


Figure 5.10: Mass Fragmentation of cyclo[ITA(SeCys)^{Sen}ITF(SeCys)^{Sen}]. LC/MS/MS spectrum of mixed species of single selenazoline containing macrocycles showing fragments confirming the presence of cyclo[ITA(SeCys)^{Sen}ITF(SeCys)^{Sen}].

5.4 Discussion

Incorporation of selenium into the PatE2K precursor peptide has been achieved by expressing it in *E. coli* BL21 (DE3) cells grown on a minimal medium supplemented with selenocysteine. Mass spectrometry analysis shows the incorporation of primarily a single selenium per peptide, with small levels of peptide containing no incorporation or the incorporation of two seleniums. The single selenium species was presumed to result in two different precursor peptides forming, one with core peptide sequence ITACITF(SeCys) and the other ITA(SeCys)ITFC. In future experiments it will be possible to re-engineer the core peptide region of the PatE precursor peptide to include only a single cysteine residue which we would expect to result in a greater selenium incorporation. The PatE2K-SeCys was taken through the *in vitro* process optimised for native peptides in Section 4. The loss of two water molecules during treatment with TruD confirms selenazoline formation. The heterocyclised peptides were then cleaved and macrocyclised to yield selenazoline containing cyclic peptides.

Reverse-phase HPLC was used to separate the 0, 1 and 2 selenium containing macrocycles and LC/MS/MS analysis was carried out to confirm their structures. The single selenium containing macrocycle fragmentation pattern clearly showed the presence of two species corresponding to the expected sequences cyclo[ITAC^{Thn}ITF(SeCys)^{Sen}] and cyclo[ITA(SeCys)^{Sen}ITFC^{Thn}]. LC/MS/MS of the two selenium containing macrocycles confirmed the presence of two selenazolines; cyclo[ITA(SeCys)^{Sen}ITF(SeCys)^{Sen}].

This study represents an enzymatic route to selenazoline containing natural products and allows scope for a library of diverse selenazoline containing macrocycles to be produced. To date, no natural products containing selenazoline or selenazoles have been isolated directly from living organisms and the only known compounds have been created by chemical synthesis.

5.5 Conclusions and Future Work

We have shown that we can introduce unnatural amino acids into a macrocycle through the direct incorporation of seleno-cysteine into the precursor peptide during expression. LC/MS/MS analysis confirms the presence of selenium and strongly support the macrocyclic structure. Seleno-cysteine is considerably less toxic than other selenium containing molecules and so offers a safer and greener method of selenazoline production.

Future work will be to oxidise selenazoline containing macrocycles to contain selenazoles and also further explore *in vitro* synthesis to incorporate other unnatural entities into cyclic peptides.

6. Structural and Biochemical Studies on Cyanobactin Oxidases

6.1 Introduction

The oxidation of heterocyclised residues is an important step in the majority of cyanobactin pathways with thiazolines and in some pathways also oxazolines being oxidised to form thiazoles and oxazoles respectively (Figure 6.1) [22].

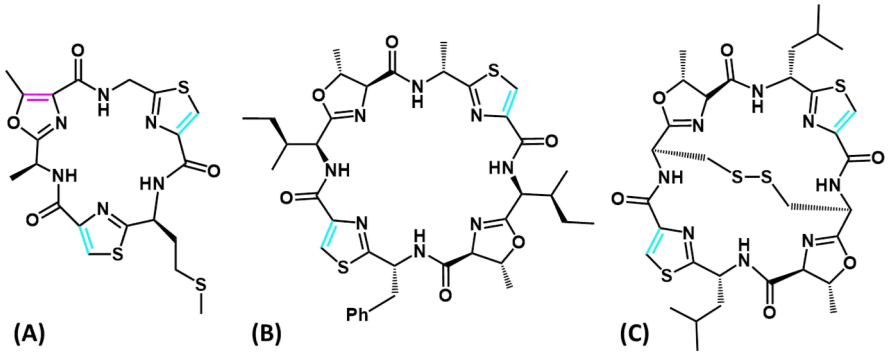


Figure 6.1: Cyanobactin Oxidation: Chemical structures of a range of cyanobactins with oxidised heterocycles. The oxidation of thiazoline rings forming a thiazole is highlighted in cyan whilst the oxidation of an oxazoline ring forming an oxazole is highlighted in magenta. A) Tenuecyclamide C [36], B) Patellamide D [42] and C) Ulithiacyclamide A [39].

In the patellamide biosynthetic pathway only thiazolines are oxidised to form thiazoles while the oxazolines are left in their reduced state [40]. To date it has not been confirmed how the oxidation process is carried out, although bioinformatic analysis of the patellamide gene cluster highlights a putative oxidoreductase domain in the N-terminal portion of PatG [47]. In most other cyanobactin pathways an oxidase domain is also present; either contained in the PatG equivalent or as an isolated protein [34], and there is a high level of homology between them (Figure 6.2). This would strongly suggest that these domains are indeed responsible for the oxidation of the thiazolines and/or oxazolines.

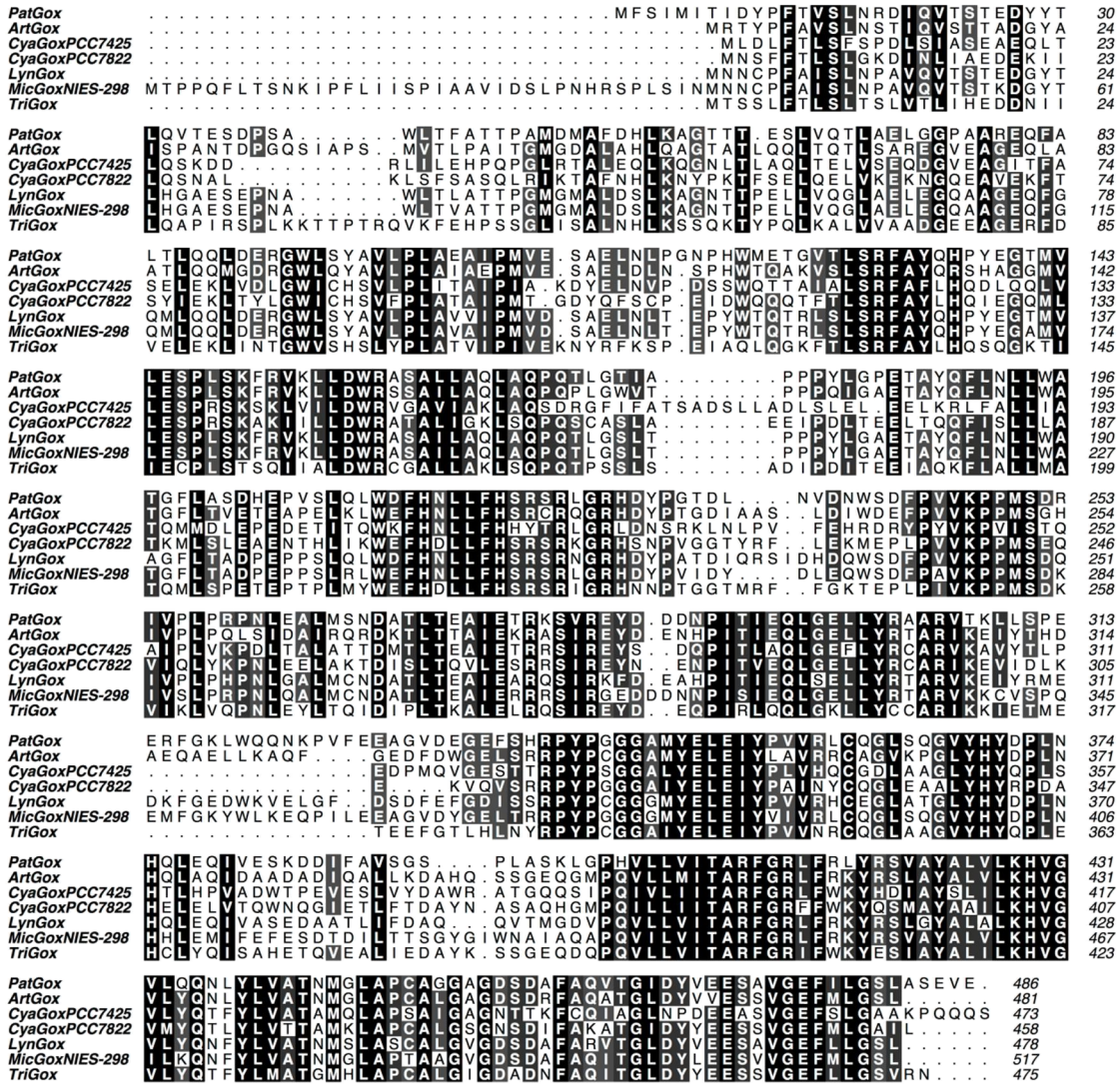


Figure 6.2: PatGox Homologue Sequence Alignment: Sequence alignment of the N-terminal domains (including oxidase) of PatG and its homologous domains from ArtG (Arthrospira platensis), CyaG (Cyanothece sp. PCC 7425 and PC7822), LynG (Lyngbya sp.), MicG (Microcystis aeruginosa NIES-298) and TriG (Trichodesmum erythraeum IMS101)

The activities of cyanobactins can vary greatly dependent on the oxidation states of their heterocycles. Studies on the lissoclinamides have shown that the oxidation of one thiazoline heterocycle (lissoclinamide 5 over lissoclinamide 4) can result in a 100-fold decrease in cytotoxicity in T24 bladder carcinoma cells (Figure 6.3) [42] [54].

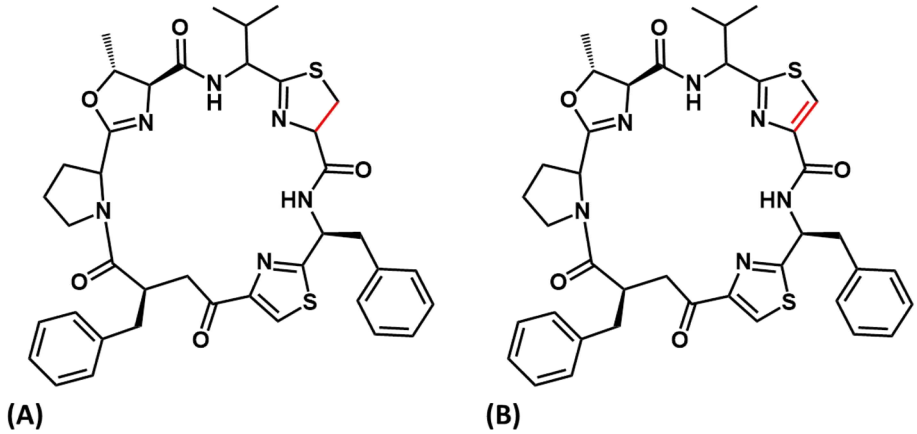


Figure 6.3: Lissoclinamide Oxidation: Chemical structures of A) lissoclinamide 4 and B) lissoclinamide highlighting the single oxidation state difference between the two which results in a 100 fold difference in cytotoxicity against T24 bladder carcinoma cells [42] [54].

As discussed, to date little is known of the oxidation of cyanobactins and of the oxidase domains found in the gene cluster. BcerB, an oxidoreductase enzyme from *Bacillus cereus* has been studied previously by Melby *et al.* (2012) through its use as a substitute for the native enzyme in the TOMM's pathway described previously (Chapter 1). The study showed that BcerB can oxidise both thiazolines and oxazolines to form thiazoles and oxazoles respectively [96] (Figure 6.4). Unlike the patellamide pathway, the TOMMS pathway generates linear peptides so BcerB must function on a linear substrate as well. In order to generate substrate processing, the enzyme had to be purified with the co-factor flavin mononucleotide (FMN) bound confirming its requirement in the oxidation process.

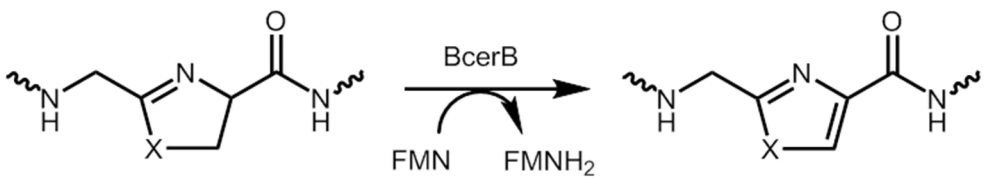


Figure 6.4 –BcerB Utilisation in TOMMs pathway – The oxidation of heterocyclic residues by BcerB protein with the FMN co-factor reduced to $FMNH_2$. X = O or S.

PatGox and BcerB catalyse similar reactions, the oxidation of heterocyclic amino acids, although BcerB has a wider substrate tolerance (Figure 6.5). The two proteins have homology of approximately 40 % suggesting that they will have a similar structure. It is also likely that PatGox will require FMN for activity as in addition to BcerB it is also required in the epothilone and bleomycin biosynthesis pathways as means to oxidise thiazolines [97]. One interesting observation between the protein sequences of PatGox and BcerB is that the former has a 197 amino acid addition at the N-terminus (Figure 6.6). This small N-terminal domain of unknown function is also conserved within cyanobactin homologues (Figure 6.2)

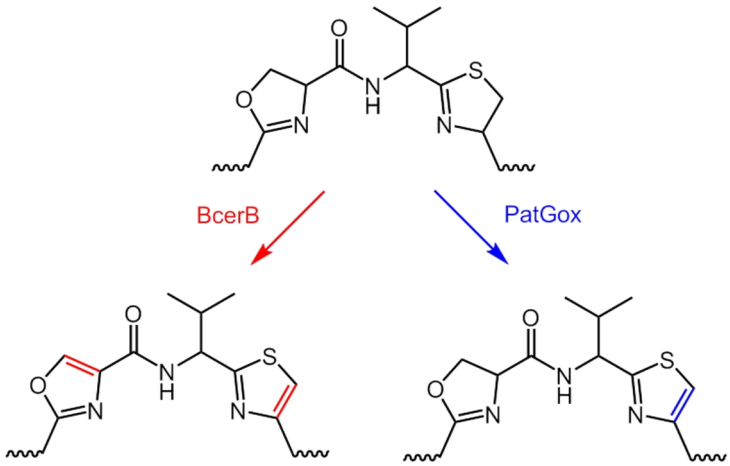


Figure 6.5: BcerB and PatGox Oxidation Reactions. Schematic of oxidation by enzymes BcerB and PatG oxidase domain highlighting the differences between the two. BcerB oxidises both thiazolines and oxazolines to thiazoles and oxazoles respectively. PatGox catalyses only the oxidation of thiazolines to thiazoles.

The study of the oxidase domain of PatG (PatGox) would assist in determining the mechanism and upon which substrate it acts; linear or macrocyclic peptide. Oxidation must occur after heterocyclisation and is most likely to occur after epimerisation as the extraction of the hydrogen from an aromatic ring is less energetically favourable than that of a azoline. Additionally, it would be intriguing to elucidate why PatGox only

oxidises thiazolines. It may be due to the enzyme only binding thiazolines or because of the chemistry of the reaction.

```
PatG-N-terminal Domain  MISIMITIDYPFTVSLNRDIQVTSTEDYYTLQVTESDPSAWLTFATTP   48
BcerB                   ..................................................

PatG-N-terminal Domain  AMDMAFDHLKAGTTTESLVQTLAELGGPAAREQFALTLQQLDERGWLS   96
BcerB                   ..................................................

PatG-N-terminal Domain  YAVLPLAEAIPMVESAELNLPGNPHWMETGVTLSRFAYQHPYEGTMVL  144
BcerB                   ..................................................

PatG-N-terminal Domain  ESPLSKFRVKLLDWRASALLAQLAQPQTLGTIAPPPYLGPETAYQFLN  192
BcerB                   ..................................................

PatG-N-terminal Domain  LLWATGFLASDHEPVSLQLWDFHNLLF.HSRSRLGRHDYPGTDLNVDN  239
BcerB                   .....KFEDHTEKYILSRTYDCNTRIFPKLLPRLSRHSTPSNYFTKIT   43

PatG-N-terminal Domain  WSDFPVVKPPMSDRIVPLPRPNLEALMSNDATLTEAIETRKSVREYDD  287
BcerB                   KKHYCI.......KSIPL.RPDKN.LL..RQEFYQVLNNRKSVEELEI   80

PatG-N-terminal Domain  DNPITIEQLGELLYRAARVTKLLSPEERFGKLWQQNKPVFEEAGVDEG  335
BcerB                   RTRIRFETLSNLLH........................F...SYGYI  100

PatG-N-terminal Domain  EFSHRPYPGGGAMYELEIYPVVRLCQGLSQGVYHYDPLNHQLEQIVES  383
BcerB                   NKPHSAAPSAGGKYPINIYIAVFNVENLEQGIYYYDREQDVLDMIRRG  148

PatG-N-terminal Domain  K.....DDIFAVSGSPLASKLGPHVLLVITARFGRLFRLYRSVAYALV  426
BcerB                   DFRESINNLYVDN.THI...HSSSFIMFHAANLDQTSSKYADRGYKLI  192

PatG-N-terminal Domain  LKHVGVLQQNLYLVATNMGLAP..CAGGAGDSDAFAQVTGIDYVEESA  472
BcerB                   HLDMGHLSQNLYLLSSAQQLGIRAIFGLYE......NKVNDFLELDG  233

PatG-N-terminal Domain  VGEFIL.....GSLASE......VESDVVEGEDEIESAGVSASEVES  508
BcerB                   ENEFVLLSHVFGGIKLSTPITMDTKFSDIYYENEETKSEG........  273

PatG-N-terminal Domain  SATKQKVAL   517
BcerB                   .........   273
```

Figure 6.6 – PatGox and BcerB Sequence Alignment – Sequence alignment of the N-terminal domain of PatG (PatGox) from *Prochloron sp*. and BcerB from *Bacillus cereus*.

This study set out to determine the structure of the N-terminal domain (oxidase) of PatG or related enzymes and to determine biochemically their substrates and potential position within the biosynthetic pathway.

6.2 Materials and Methods

6.2.1 DNA Cloning

The 5'-region of *patG*, encompassing the oxidase domain (residues 1-517, PatGox) was synthesised using optimised codons for *E. coli* into the pJexpress 411 vector (DNA 2.0) with an N-terminal His_6-tag and TEV protease site [105].

PatGox was sub-cloned by PCR to create two shorter constructs of residues 250-478 (PatGox2) and 271-481 (PatGox3) using the primers outlined in Table 6.1 giving *NcoI* and *EcoRI/SalI* restriction sites at the 5' and 3' ends respectively. The PCR products were digested with the appropriate restriction enzymes and ligated into the pHisTEV and pHisMBPTEV vectors. PatGox3 was additionally cloned into the pSUMO plasmid.

5' PatGox2NcoI	cttgccatgggcgatcgtattgtgccg
3' PatGox2EcoRI	cttgagctcttagcccaggatgaattcaccaac
5' PatGox3NcoI	ggcgccatgggcacgctgacggaagcgattgaa
3' PatGox3XhoI	tttctcgagttacagggagcccaggatgaattc

Table 6.1: PatGox Primer Sequences: Oligonucleotide sequences for PatGox2 and PatGox3 cloning experiments.

The 5'-regions of a selection of *patG* homologues were synthesised using optimised codons for *E. coli* into the pJexpress 401 vector (DNA 2.0) with an N-terminal His_6-tag and TEV protease site (Figure 6.7) as outlined in Table 6.2. The CyaGox PCC7425 was supplied by Dr Houssen (University of Aberdeen).

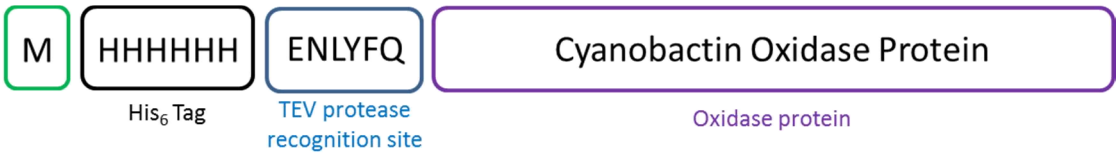


Figure 6.7: Schematic for Oxidase Protein Constructs: Oxidase constructs with N-terminal His_6-tag followed by a TEV protease site and the protein of interest.

Protein	Organism	Construct Name	Protein Residues
ArtG	*Arthrospira platensis*	ArtGox	2-481
CyaG	*Cyanothece sp*. PCC 7425	CyaGox PCC 7425	1-473
CyaG	*Cyanothece sp*. PCC 7822	CyaGox PCC 7822	2-458
MicG	*Microcystis aeruginosa* NIES-298	MicGox	2-517
TriG	*Trichodesmum erythraeum* IMS101	TriGox	2-475

Table 6.2: PatGox Homologues: Cyanobactin oxidase proteins, source organisms and construct lengths.

6.2.2 PatGox Expression and Purification

PatGox was expressed from the pJexpress 411, pHisTEV and pHisMBPTEV plasmids in a range of *E. coli* cell lines [BL21 DE3, BL21 Star, BL21 gold, Rosetta, C43, Origami, Tuner] on a small scale (10 ml cultures). The cells were grown in LB media to an optical density at 600 nm (OD_{600nm}) of 0.6 and then induced with 0.01, 0.05, 0.1 and 0.25 mM IPTG concentrations and expressed at either 37 °C for 3 hours, 30 °C for 5 hours, or 25 °C or 20 °C overnight.

Each culture was harvested by centrifugation (4,000 x g) with the cell pellet re-suspended in lysis buffer (Appendix A.9), sonicated at 10 microns (SoniPrep 150, MSE) and centrifuged at 30,000 x g with the soluble fraction then applied to a His-tag pull down assay (Qiagen BioSprint). The elution and flow-through samples were analysed by SDS-PAGE to determine any soluble expression.

PatGox2 and PatGox3 were expressed from the pHisTEV and pHisMBPTEV or pSUMO plasmids in BL21 (DE3) *E. coli* cells by growing in LB medium at 37 °C to an OD_{600nm} of 0.6 and inducing with 1 mM IPTG and growing overnight at 20 °C. The cultures were treated as for PatGox and applied to His-tag pull down assay. Samples were analysed by SDS-PAGE.

6.2.3 Homologue Expression Trials

Homologues of PatGox (ArtGox, CyaGox PCC 7425, CyaGox PCC 7822, MicGox and TriGox) were expressed separately from the pJexpress 401 plasmid using BL21 (DE3) *E. coli* cells grown on auto-induction medium using the Studier method [107] with the addition of 50 µM riboflavin (see Appendix A.10). The cultures were grown at 20 °C, 250 rpm for 48 hours before harvesting by centrifugation.

Each culture was harvested by centrifugation (4,000 x g) with the cell pellet re-suspended in lysis buffer, sonicated at 10 microns (SoniPrep 150, MSE) and centrifuged at 30,000 x g with the soluble fraction then applied to a His-tag pull down assay (Qiagen BioSprint). The elution and flow-through samples were analysed by SDS-PAGE to determine any soluble expression.

6.2.4 ArtGox Expression, Purification and Crystallisation

ArtGox was expressed from the pJexpress 401 plasmid as described in section 6.2.3.

Cell pellets of ArtGox were re-suspended in lysis buffer plus EDTA-free protease inhibitor tablets (Roche) and DNAse at 0.4 mg g^{-1} wet cell pellet. The re-suspension was lysed by passage through a cell disruptor at 30 kPsi (Constant Systems). The lysate was cleared by centrifugation (40,000 x g, 4 °C, 20 min) and then loaded onto a Ni-Sepharose 6 FF column (GE Healthcare) equilibrated in lysis buffer. The column was washed with lysis buffer and then ArtGox eluted with elution buffer. The protein was then passed over a desalting column (Desalt 16/10, GE Healthcare) in desalting buffer. TEV protease was added at a mass-to-mass ratio of 1:10 and the protein was digested for two hours at 20 °C to remove the His_6-tag. The sample was then loaded on to a second Ni-column, in desalting buffer. ArtGox was collected in the flow through. The protein was concentrated to 7.5 ml (Vivaspin concentrators, 30 kDa MWCO) and

applied to a Superdex 200 gel filtration column (GE Healthcare) equilibrated in gel filtration buffer. FMN was included in all stages of the purification with the exception of the final gel filtration step. The protein was concentrated to 7.5 mg ml^{-1} for crystallography. The purity of the protein was confirmed by SDS-PAGE analysis and its identity confirmed by mass spectrometry (MS).
Crystallisation trials were set up at varying concentrations from 5 - 10 mg ml^{-1} using the Gryphon crystallisation robot against a range of stochastic screens.

6.2.5 CyaGox (PCC 7425) Expression, Purification and Crystallisation

CyaGox (PCC 7425) was expressed from the pJexpress 401 plasmid as described previous in section 6.2.3. The protein was purified using the same methods used for ArtGox (section 6.2.4).

(L)-Selenomethionine-labelled CyaGox was expressed in BL21 (DE3) *E. coli* cells, cultures of which were grown in a minimal medium supplemented with glucose-free nutrient mix (Molecular Dimensions), 5 % glycerol and 50 µM riboflavin. The medium was inoculated with overnight culture grown in LB medium which was subsequently washed three times with minimal medium. After 15 min growth at 37 °C, 60 mg L^{-1} (L)-selenomethionine was added to the cultures. An amino acid mix (100 mg L^{-1} lysine, phenylalanine and threonine, 50 mg L^{-1} isoleucine and valine) was added to the cultures at an optical density at 600nm (OD_{600nm}) of 0.6. After 15 min further growth at 37 °C the cultures were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and grown for 30 hours at 20 °C before harvesting by centrifugation. SeMet CyaGox was purified using the same protocols as native CyaGox and ArtGox.

Crystallisation trials were set up at 5 - 12 mg ml^{-1} using the Gryphon crystallisation robot against a range of stochastic screens. Additionally, the protein was incubated

with 1 mM FMN and 1:1.1 equivalents of either PatE2-C50P or the peptide 'NILPQQGQPVIR' and set up in crystallisation trials as for the apo protein.

6.2.6 Phasing Trials

Native crystals were soaked in a drop of 10 - 100 mM heavy atom solutions made up in mother liquor and suspended over a fresh well of mother liquor at 20 °C for various time periods ranging from five minutes to 48 hours.

6.2.7 Oxidase Activity Assays

The activity of ArtGox was determined by incubation of the enzyme with either full length Cys-heterocyclised PatE-IMACIMAC or with a selection of macrocycles containing thiazolines (created in Section 4). 200 µM PatE / macrocycle was incubated with 20 µM ArtGox and 1 mM FMN at 37 °C overnight in the dark. (The reaction was carried out in the absence of light due to the potential for FMN to oxygenate His residues on light activation). The reactions were analysed by MALDI-MS or by LC-MS.

6.2.8 Structural Predictions

Structural predictions of PatGox, ArtGox and CyaGox (PCC 7425) were carried out using the Phyre2 server by inputting the amino acid sequence and searching against the PDB database of protein structures.

6.3 Results

6.3.1 PatGox Expression and Purification

PatGox was expressed from pJexpress 411 plasmid in various *E. coli* cell lines with a range of induction concentrations and induction temperatures. Each expression trial was analysed by applying the soluble fraction to a His-tag pull-down assay (Qiagen BioSprint) and running the appropriate eluted fraction on SDS-PAGE (Figure 6.8). No significant PatGox protein was found under any of the conditions/cell lines. A screen of the lysis buffer (Appendix A.11) also led to no isolated protein from nickel elution.

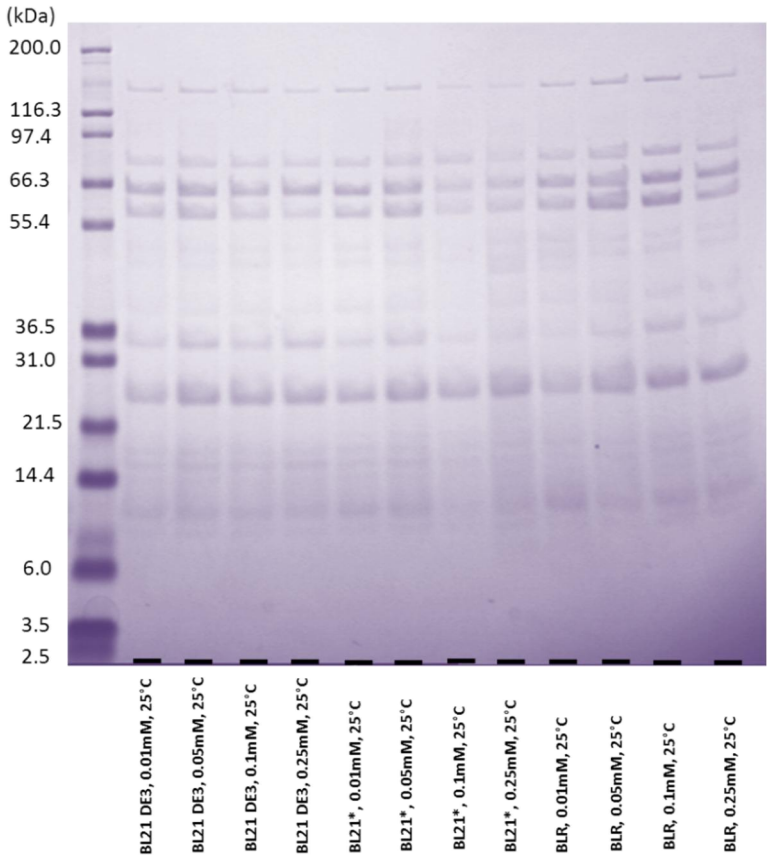


Figure 6.8: SDS-PAGE Analysis of PatGox Expression Trials: Representative SDS-PAGE analysis of PatGox expression trials under varying cell lines, induction concentrations and temperatures. There was no evidence of PatGox expression under any conditions. Gel bands in the size region of PatGox (57 kDa) were confirmed by peptide mass fingerprinting to be *E. coli* proteins.

Analysis of the His-tag pull-down flow-through (non-bound material) showed significant quantities of PatGox (Figure 6.9). Two feasible explanations for this are possible; firstly, the His-tag is somehow occluded in the PatGox structure or secondly, the protein is highly aggregated resulting in no tag being free to bind the nickel resin.

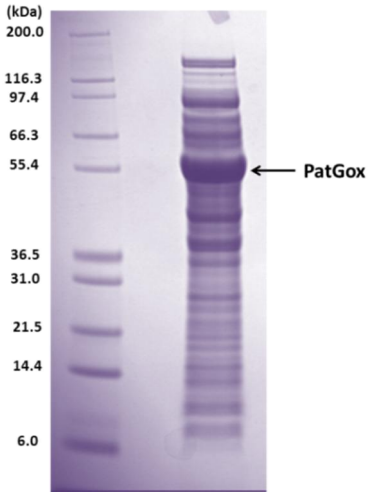


Figure 6.9: SDS-PAGE Analysis of PatGox Non-Bound Material: SDS PAGE analysis of the flowthrough (non-bound) sample of a His-tag pull-down experiment showing a significantly sized band corresponding to PatGox.

In order to assess this further, a large scale expression was applied first to a His-trap affinity column with the flow-through then passed over a Capto-Q column with the aim to separate proteins based on their charge. An elution gradient of 0.1 to 1 M NaCl was applied to the column after loading. However, multiple broad peaks were observed with PatGox present in all peaks, each with significant contaminants. The most abundant PatGox containing peak was pooled and applied to gel filtration (Superdex S200). However, the PatGox eluted in the void volume confirming that the protein is highly aggregated (Figure 6.10).

As a wide range of expression trials (differing cell lines, buffers, induction concentrations, temperatures) were trialled, but led to no improvement in expression or reduction in aggregation the full length PatGox construct was abandoned.

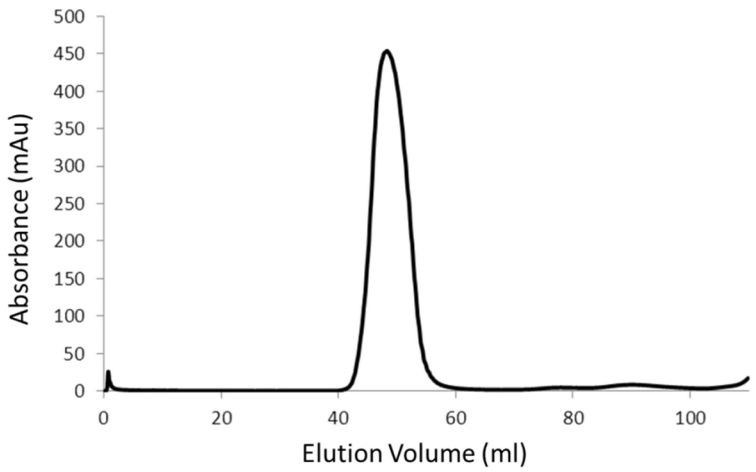


Figure 6.10: Gel Filtration of PatGox: UV spectrum from gel filtration of PatGox pooled from Capto-Q column. The PatGox is the predominant protein and elutes in the void volume (40 - 55 ml) indicating that the protein is highly aggregated.

PatGox2 and PatGox3 were expressed from the pHisMBPTEV and/or pSUMO plasmids in BL21 (DE3) *E. coli* cells however both constructs showed no level of soluble expression when applied to His-tag pull down assays. Minimal levels were found in the flow through and insoluble fractions.

6.3.2 Homologue Expression Trials

Five cyanobactin oxidase proteins were expressed from the pJexpress 401 plasmid in BL21 DE3 *E. coli* cells and analysed by His pull-down using the Qiagen Biosprint robot. The flow through and elution fractions were analysed by SDS-PAGE to assess soluble expression (Figure 6.11)

ArtGox and CyaGox PCC7425 showed significant levels of soluble protein in the elution fractions, while TriGox showed a minor level of soluble protein in the elution fraction. All three proteins were confirmed by MS. At this stage ArtGox and CyaGox PCC7425 were carried forward for large scale expression and purification trials.

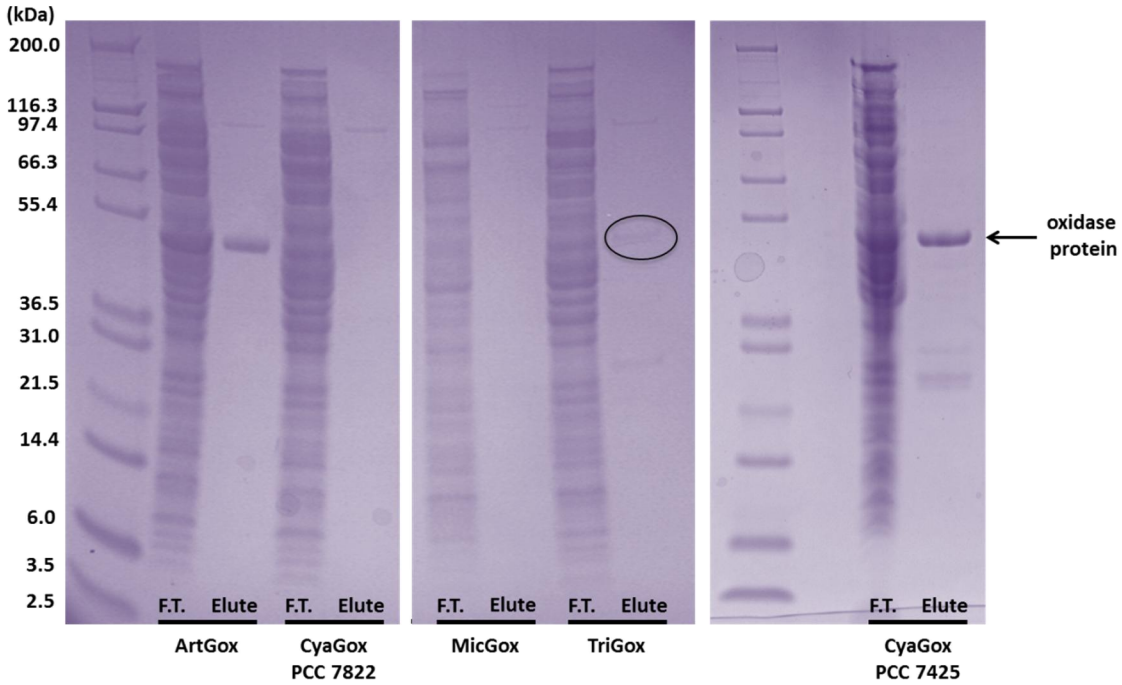


Figure 6.11: SDS-PAGE Analysis of Oxidase Homologues: SDS-PAGE analysis of six cyanobactin oxidase proteins (ArtGox, CyaGox PCC7822, MicGox, TriGox and CyaGox PCC7425) when analysed by His pull-down. Flow-through and elution fractions for each protein are shown. ArtGox and CyaGox PCC 7425 show significant levels of expression in the elution fraction. TriGox shows minor expression in the elution (gel band is not clear so has been circled).

6.3.3 ArtGox Expression and Purification

ArtGox was expressed from the pJexpress 401 plasmid in BL21 DE3 *E. coli* cells using the Studier method [158]. The protein was isolated by nickel chromatography using the His_6-tag. The tag was removed by treatment with TEV protease at > 95 % efficiency. The protein eluted off gel filtration as a monomer and with a distinct yellow colour indicating that FMN was bound (No FMN was included in the final gel filtration buffer). A final yield of 25 mg L^{-1} culture was achieved. SDS-PAGE analysis shows the protein to be > 95 % pure but with a shadow band directly under the main band (Figure 6.12). MSMS analysis of both gel bands confirmed that both contained ArtGox peptides suggesting that the lower band has been slightly cleaved. The protein was concentrated to 7.5 mg ml^{-1} and applied to crystallisation trials.

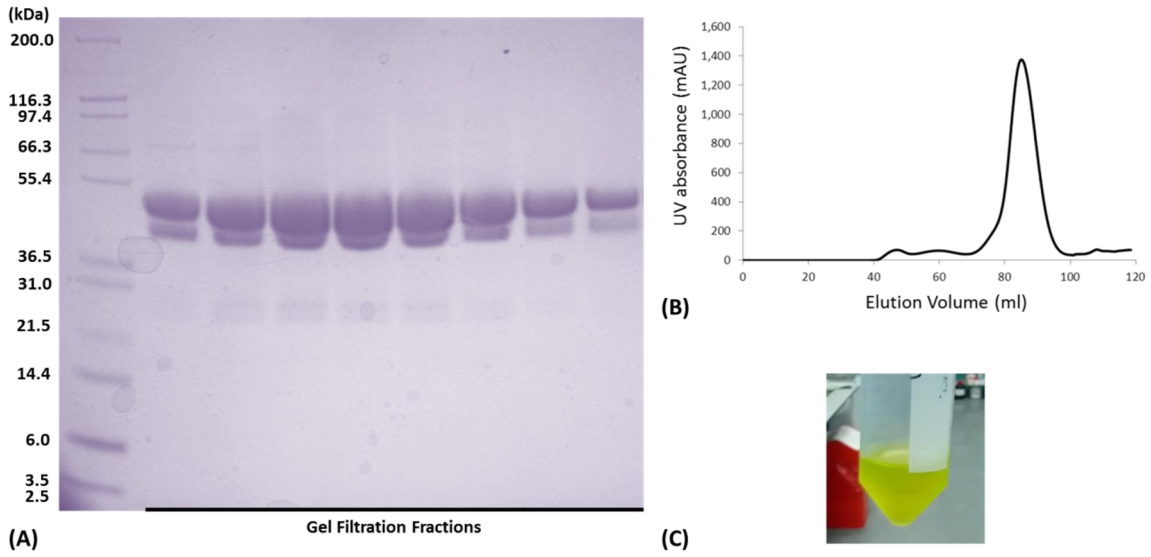


Figure 6.12: Purification of ArtGox: (A) SDS-PAGE analysis of ArtGox gel filtration fractions (B) Gel filtration UV spectra of ArtGox showing a single peak corresponding to a monomer (C) Concentrated ArtGox post gel-filtration showing a deep yellow colour indicating the presence of FMN.

6.3.4 CyaGox (PCC 7425) Expression and Purification

CyaGox (PCC 7425) was expressed from the pJexpress 401 plasmid in BL21 DE3 *E. coli* cells using the Studier method [158]. The protein was isolated by nickel chromatography using the His_6-tag. The tag was removed by treatment with TEV protease at > 90 % efficiency. The protein eluted off gel filtration as a monomer and with a distinct yellow colour indicating that FMN was bound (No FMN was included in the final gel filtration buffer) (Figure 6.13). A final yield of 5 mg L^{-1} culture was achieved. SDS-PAGE analysis shows the protein to be > 95 % pure. The protein was concentrated to 12 mg ml^{-1} and applied to crystallisation trials.

Purification of SeMet CyaGox occurs as with the native protein however with much reduced yields. A final yield of 5 mg L^{-1} cell culture with purity greater than 98 % was achieved and the protein confirmed by SDS-PAGE and MS.

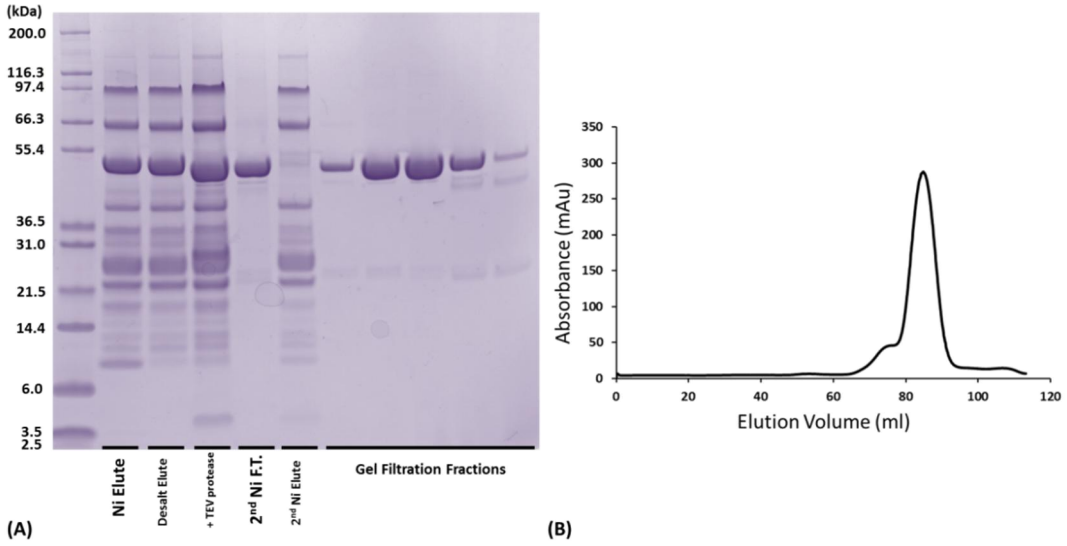


Figure 6.13: Purification of CyaGox: (A) SDS-PAGE analysis of CyaGox purification steps with CyaGox the primary band (B) Gel filtration UV of CyaGox showing a single peak corresponding to a monomer.

6.3.5 ArtGox Crystallisation

Crystallisation trials of ArtGox were carried out against stochastic screens (Appendix E) using the Gryphon nanolitre crystallisation robot. In the majority of cases, only clear or precipitated drops were observed under a range of protein concentrations. However, under one condition (Qiagen PEGS I Suite, condition 64 – 0.2 M potassium thiocyanate, 20 % PEG 3350) small, non-single crystals were observed. These were confirmed as protein due to their yellow colour (as result of bound FMN) and their high absorbance under ultraviolet light (Figure 6.14).

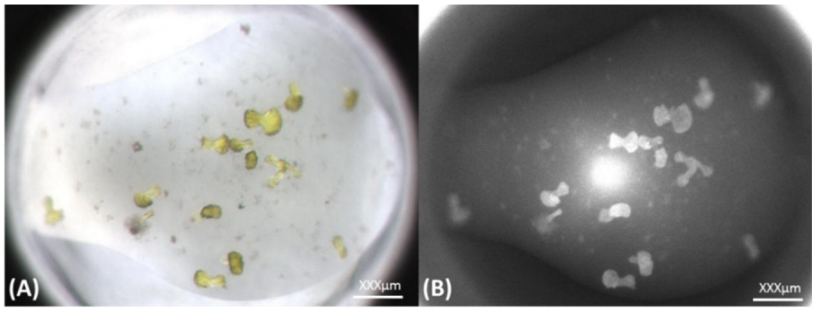


Figure 6.14: Crystallisation of ArtGox: (A) Crystals of ArtGox grown in 0.2 M potassium thiocyanate, 20 % PEG 3350 (B) The same crystals under ultraviolet light.

Several attempts were made to repeat and optimise these into X-ray diffraction quality crystals but to date this has been unsuccessful.

6.3.6 CyaGox (PCC 7425) Crystallisation

Crystallisation trials of CyaGox PCC 7425 resulted in no crystal hits, however a single crystal hit was observed when the protein was setup in complex with 1 : 1.1 equivalents of PatE2-ITACITFP and 1 mM FMN. The preliminary crystals were found in the Wizard Screen 2 (Rigaku Reagents) condition 16 (1.0 M sodium citrate, 0.1 M CHES pH 9.5). These crystals were analysed by UV light and subsequently soaked in Izit Dye (Hampton Research) to confirm that they were protein and not salt crystals (Figure 6.15).

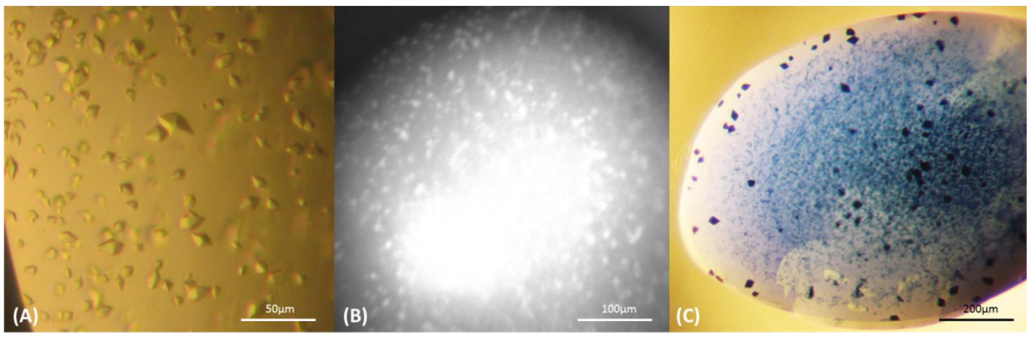


Figure 6.15: Crystallisation of CyaGox: Crystals of CyaGox grown in 1.2 M sodium citrate, 0.1 M CHES pH 8.5 (A) under visible light (B) under UV light and (C) when soaked with Hampton Izit dye.

Several rounds of optimisation were carried out with single crystals grown in a condition of 0.9 M sodium citrate, 0.1 M CHES pH 8.0 and with a protein concentration of 5 mg ml^{-1}. Crystals of size 50 x 50 x 50 µm were observed after three days (Figure 6.16.A). The crystals were cryo-protected in paraffin oil and analysed by synchrotron radiation at the Diamond Light Source, beamline I04-1. A dataset was collected (Figure 6.16 B, C) and processed to 2.97 Å in the space group $P4_12_12$ with a cell of 110 x 110 x 197 Å (Table 6.3). These crystals did not diffract reproducibly and often diffraction datasets were limited to 8.0 Å.

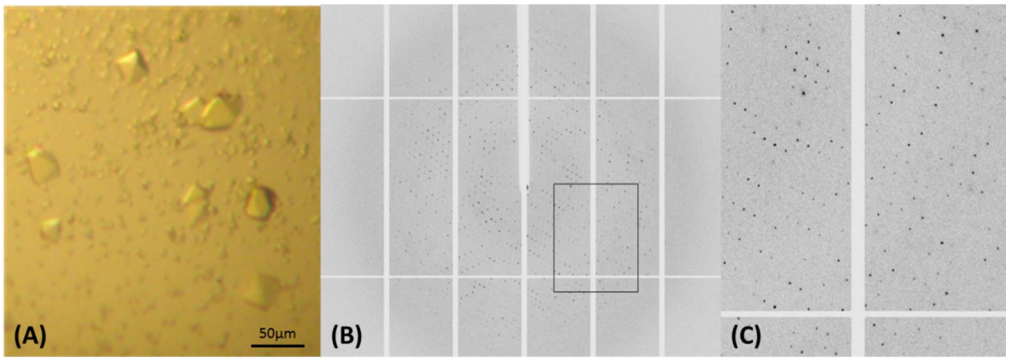


Figure 6.16: Optimisation and Analysis of CyaGox Crystals: (A) Optimised co-crystals of CyaGox and PatE2-ITACITFP grown in 0.9 M sodium citrate, 0.1 M CHES pH 8.0 (B) X-ray diffraction pattern of CyaGox complex with PatE2-ITACITFP with spots to 2.97 Å (C) Zoom view of diffraction pattern

Native CyaGox + PatE2-ITACITFP Dataset	
Wavelength (Å)	0.92
Space group	P 4_1 2_1 2
a, b, c (Å)	109.8, 109.8, 196.9
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	73.30 – 2.97 (3.05-2.97)
I/σI	27.9 (3.6)
R_{merge} (%)	7.6 (83.9)
Completeness (%)	100.0 (100.0)
Redundancy	14.3 (15.2)

Table 6.3: Data Collection Statistics of CyaGox in Complex with PatE-C50P: Values for the highest resolution shell are provided in parenthesis. The dataset was collected at the Diamond Light Source beamline I03 (Dectris Pilatus 6M detector).

Following the initial CyaGox:PatE-C50P native dataset, a second shorter peptide was trialled in co-crystallisation with CyaGox and 1mM FMN. The use of the PatE leader sequence fragment 'NILPQQGQPVIR' yielded significantly larger (200 x 80 x 80 µm) and visually better looking crystals (Figure 6.17 A). These crystals were cryo-protected in paraffin oil, analysed by X-ray and found to diffract more routinely in-house (to between three and five angstroms) than the previous crystals. Several crystals were

tested at the synchrotron with the best dataset being collected to a resolution of 2.65 Å (Figure 6.17 B, C and Table 6.4).

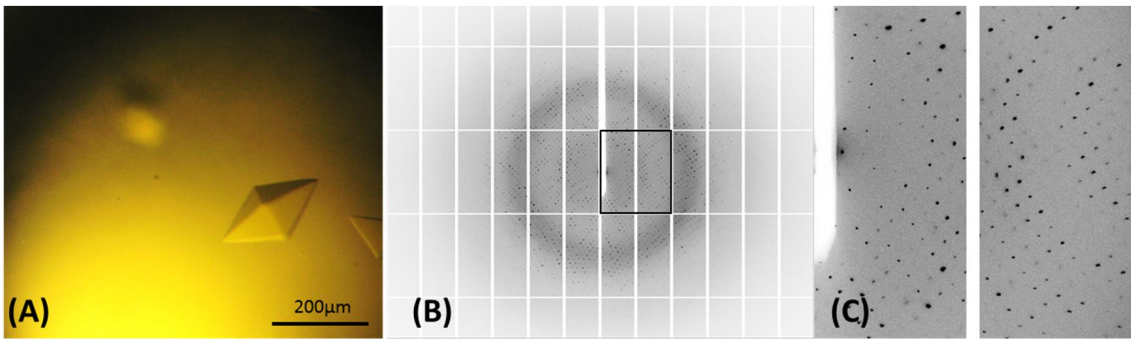


Figure 6.17: Co-crystallisation of CyaGox: (A) Optimised co-crystals of CyaGox and NILPQQGQPVIR peptide grown in 1.05 M sodium citrate, 0.1 M CHES pH 7.75 (B) Synchrotron X-ray diffraction patterns of CyaGox complex with 'NILPQQGQPVIR' peptide with spots to 2.65 Å (C) Zoom view of diffraction pattern.

Native CyaGox + NILPQQGQPVIR pep Dataset	
Wavelength (Å)	1.77
Space group	P 4_1 2_1 2
a, b, c (Å)	109.3, 109.3, 195.4
α, β, γ (°)	90.0, 90.0, 90.0
Resolution (Å)	72.83 – 2.65 (2.72-2.65)
I/σI	17.9 (2.7)
R_{merge} (%)	8.3 (83.6)
Completeness (%)	100.0 (100.0)
Redundancy	8.1 (8.5)

Table 6.4: Data Collection Statistics of CyaGox in Complex with Peptide NILPQQGQPVIR: Values for the highest resolution shell are provided in parenthesis. The dataset was collected at the DLS beamline I02 (Dectris Pilatus 6M detector).

In order to confirm fully that the protein crystals were indeed CyaGox and not a contaminant, several crystals were washed in four subsequent drops of fresh mother liquor (to remove any remaining soluble protein) and then finally dissolved in SDS

loading buffer. The sample was then run on SDS-PAGE (Figure 6.18). The gel band at 53 kDa was excised, subjected to tryptic digest (ProGest robot) and analysed by MALDI TOF MS with the resultant peptides further analysed by MSMS. The data was searched using the Mascot server (Matrix Science) against the St Andrews and NCBI databases with both searches giving strong peptide hits for CyaGox confirming the crystals were as expected.

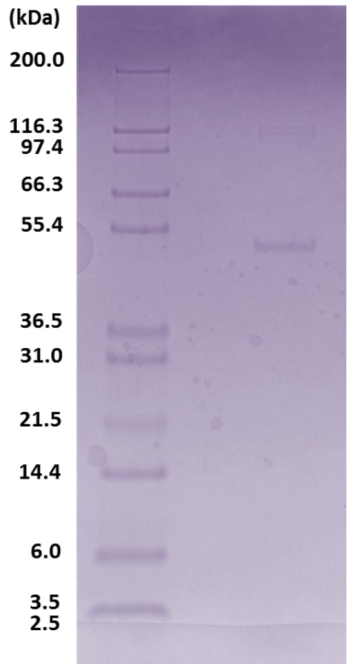


Figure 6.18: SDS-PAGE of CyaGox Crystals: SDS-PAGE analysis of CyaGox protein crystals dissolved in SDS loading buffer with a single band at 53 KDa.

6.3.7 CyaGox (PCC 7425) Phasing Trials

Molecular replacement, using Phaser [137] as part of the CCP4 suite [138], was initially attempted to determine the crystal structure with several known FMN oxidoreductases used as search models, in particular two putative nitroreductases from *anabena varibilis* (PDB code: 3EO7) and *ralstonia eutropha* (PDB code: 3HJ9) which showed the highest sequence homology of any PDB entry. Analysis of the Matthews coefficient strongly suggested two molecules in the asymmetric unit and as

a result two models were searched for in each trial. Unfortunately, no success was gained even after trimming side chains and flexible loops from these structures (summarised in Table 6.5). In general a minimum Z-score of 8.0 is the cut-off for proceeding with molecular replacement and both molecules must pack in the asymmetric unit. A model (Figure 6.19) was creating using the Robetta [159] and server and subsequently used in MR trials however this too failed to correctly determine the phases.

Model Protein	Residues	Notes	Top Z-Score	Packs?
PDB: 3EO7	1-510	Full length	6.0	Yes
PDB: 3E07	274-510	Aligned region	6.1	Yes
PDB: 3E07	274-510	Side chains removed	5.9	Yes
PDB: 3E07	274-510	Side chains & loops removed	6.1	Yes
PDB: 3HJ9	2-216	Full length	6.1	Yes
PDB: 3HJ9	19-204	Aligned region	6.3	No
PDB: 3HJ9	19-204	Side chains removed	5.7	No
PDB: 3HJ9	19-204	Side chains & loops removed	5.6	Yes
Rosetta Output	1-477	None	6.0	No

Table 6.5: CyaGox Molecular Replacement Trials. Z-scores from Phaser output of CyaGox dataset versus various models

The failure of MR led to a focusing of efforts on solving the phases via crystallisation of the selenomethionine-containing protein. SeMet CyaGox was placed into co-crystallisation trials using both the PatE-C50P and NILPQQGQPVIR peptide under the same conditions as for native CyaGox. No crystals were observed, however, under these conditions or subsequent optimisation around them. Additionally, the SeMet protein complexes were subjected to stochastic screening, but again no crystals were observed. This result suggests that the one or more of the SeMet residues may be causing an interference with the crystal packing.

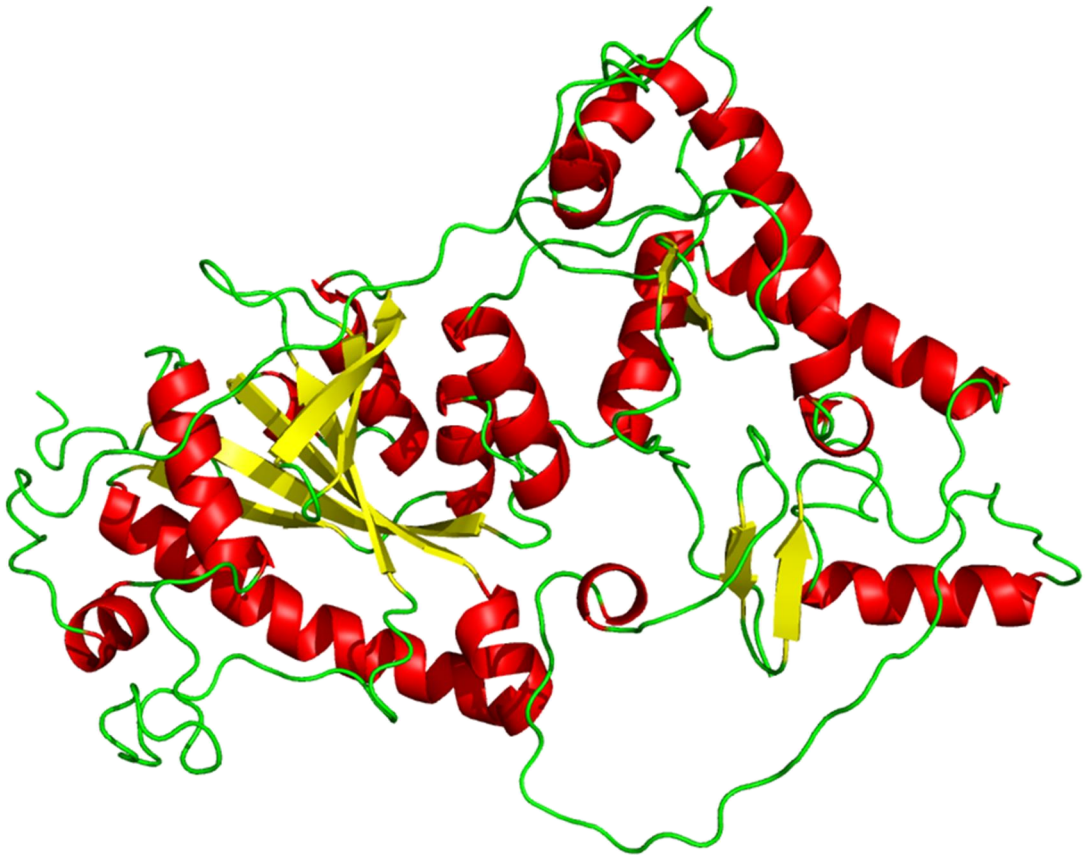


Figure 6.19: Robetta Molecular Model. Molecular model created from the Robetta server upon input of the CyaGox protein sequence (alpha helices - red, beta sheets - yellow) [159].

As SeMet crystallisation was not possible with the current protein construct, attempts were made to soak various heavy atoms into native crystals. Short, high concentration and also longer, lower concentration soaks were carried out with the following heavy atom derivatives; K_2PtCl_4, $KAu(CN)_2$, $(CH_3)_3PbCl$, NaBr, NaI, I_3C, KBr, KI, $SmOAc_2$, K_2PtBr_4, $HgCl_2$ and Thiomersal ($C_9H_9HgNaO_2S$).

Crystals were cryo-protected, frozen in at 100 K and assessed under synchrotron radiation at the appropriate wavelength (anomalous edge) to give the strongest anomalous signal for the heavy atom used. The crystals diffracted in general between 3.0 and 5.0 Å and a high redundancy dataset was collected on each. Unfortunately no anomalous signal was detected with any of the heavy atoms trialled to date. At the time of writing work is still on-going to determine phases for the CyaGox crystals.

6.3.8 Oxidase activity assay

Full length PatE (IMACIMAC) was heterocyclised by TruD to give two thiazolines from the cysteine residues as determined by a loss of 36 Da. The heterocyclised PatE was subsequently incubated with ArtGox and FMN overnight and MALDI-MS confirmed a further loss of 4 Da indicating the oxidation of the thiazolines to thiazoles (Figure 6.20)

In order to confirm the oxidation of the thiazoline to thiazole, PatE was N-terminally cleaved by trypsin and macrocyclised with PatGmac. The resultant macrocycle was confirmed by MS and shown by MSMS fragmentation to be the oxidised variant of IMACIMAC (Figure 6.21, 6.22)

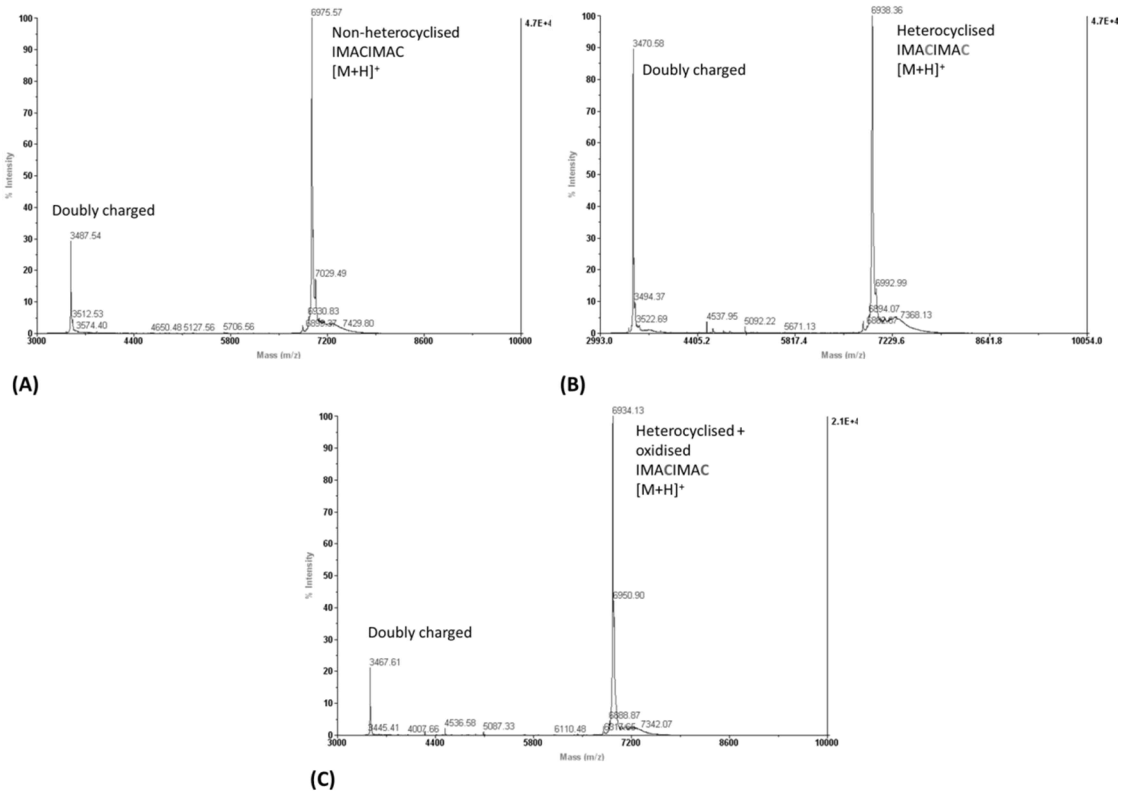


Figure 6.20: Linear Oxidation of PatE by ArtGox. MALDI-TOF-MS spectra of PatE-IMACIMAC (A) unprocessed (B) after heterocyclisation by TruD showing loss of 36 Da confirming production of two heterocycles (C) after heterocyclisation and subsequent oxidation by ArtGox with a further 4 Da loss confirming heterocycle oxidation.

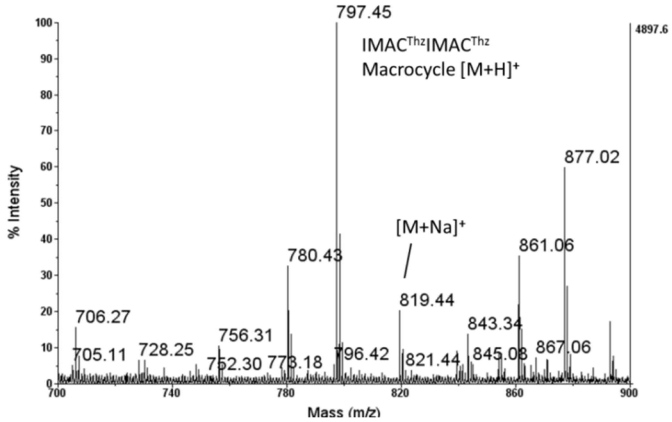


Figure 6.21: MS of cyclo[IMAC^{Thz}IMAC^{Thz}]. MALDI TOF MS spectrum of the cyclo[IMAC^{Thz}IMAC^{Thz}] macrocycle. The sample arose from the oxidation of linear TruD heterocyclised PatE (Figure 6.20) and subsequent cleavage by trypsin and macrocyclisation by PatGmac.

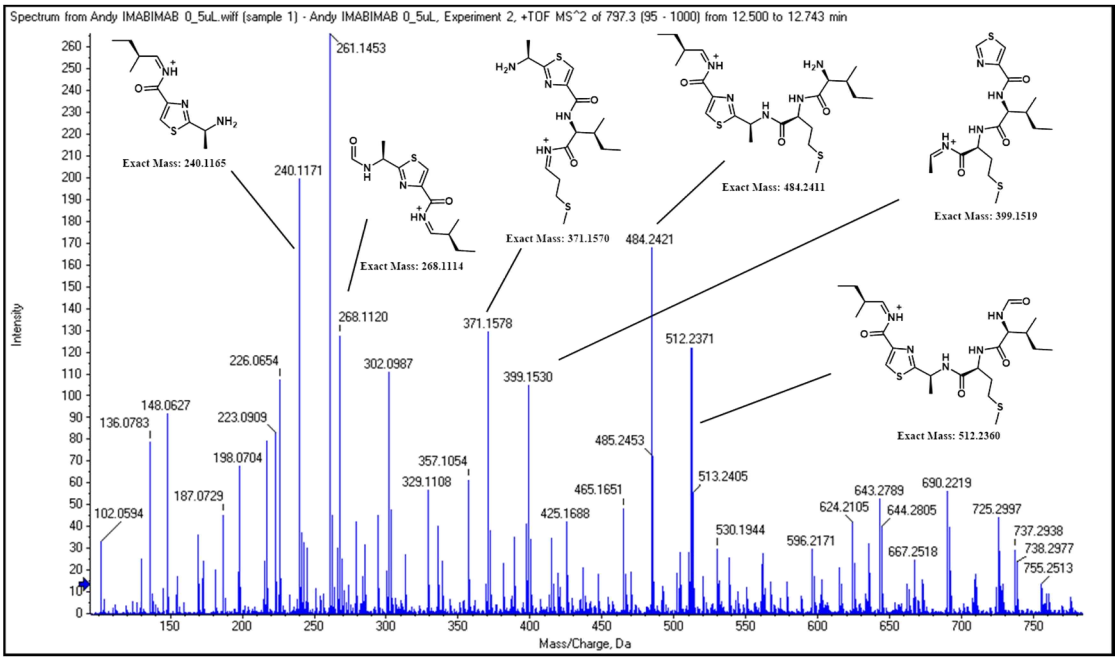


Figure 6.22: LC/MS/MS Analysis of cyclo[IMAC^{Thz}IMAC^{Thz}]. LC/MS/MS spectrum of cyclo[IMAC^{Thz}IMAC^{Thz}]. Several of the fragment ions have been assigned and correlate with the chemical structure.

To explore the diversity of the oxidase enzyme ArtGox, samples of non-oxidised cyclo[ATAC^{Thn}ITFC^{Thn}] and cyclo[ITAAITFC^{Thn}] (generated in Chapter 4) were incubated overnight with ArtGox and FMN and re-tested by MS. Losses of 4 and 2 Da were observed respectively corresponding to the oxidation of the thiazolines to thiazoles (Figures 6.23 and 6.24).

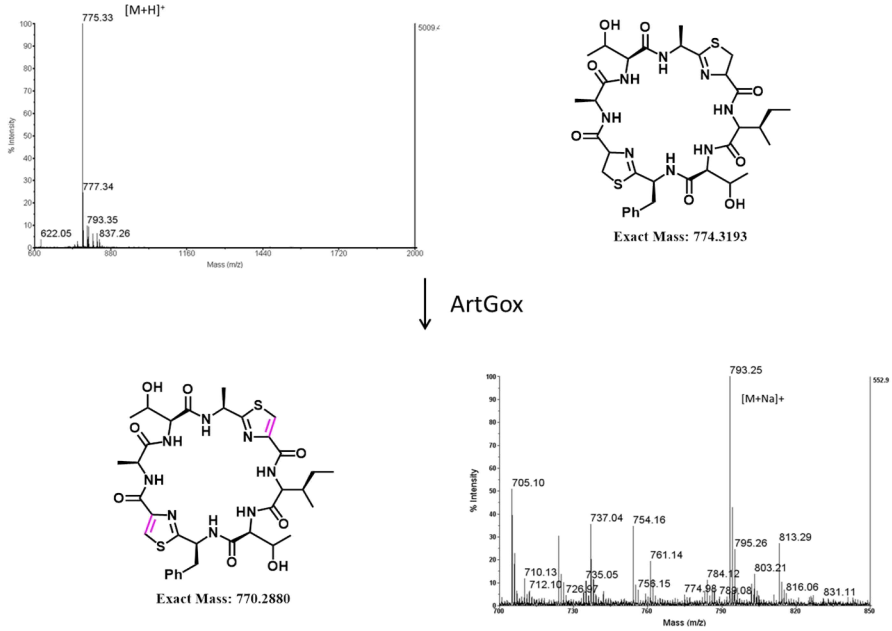


Figure 6.23: Oxidation of cyclo[ATAC^{Thn}ITFC^{Thn}]. MALDI-TOF-MS spectra of cyclo[ATAC^{Thn}ITFC^{Thn}]prior to ArtGox incubation and post ArtGox incubation resulting in a loss of 4 Da, when H^+/Na^+ ions are taken into account, generating cyclo[ATAC^{Thz}ITFC^{Thz}]. Chemical structures for both are shown with oxidation highlighted in magenta.

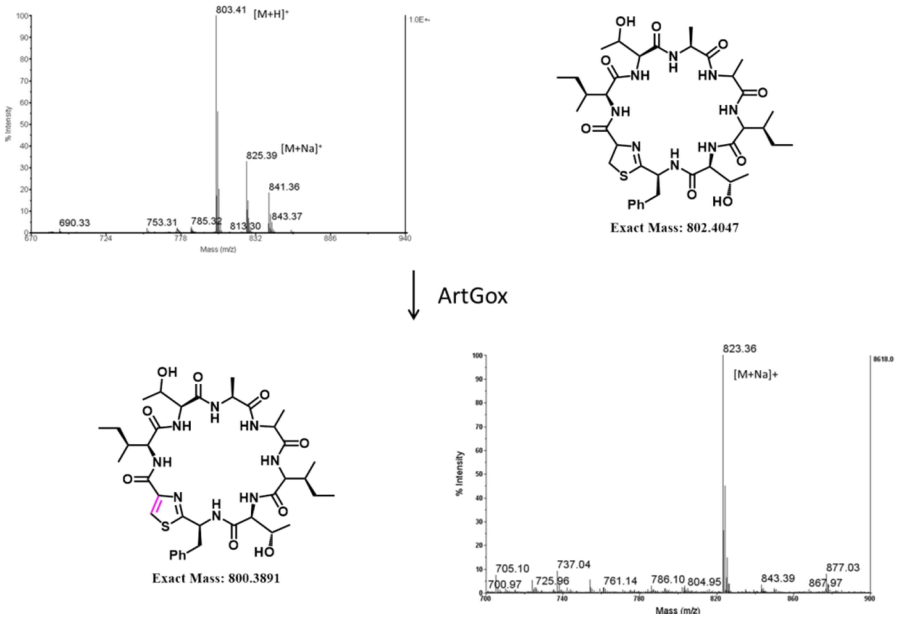


Figure 6.24: Oxidation of cyclo[ITAAITFC^{Thn}]. MALDI-TOF-MS spectra of cyclo[ITAAITFC^{Thn}] prior to ArtGox incubation and post ArtGox incubation resulting in a loss of 2 Da, when H^+/Na^+ ions are taken into account, generating cyclo[ITAAITFC^{Thz}]. Chemical structures for both are shown with oxidation highlighted in magenta.

6.3.9 Structure Predictions

Homology models of PatGox, ArtGox and CyaGox (PCC 7425) were created using Phyre2 structure prediction software [125] by supplying the protein sequences, with similar structural predictions observed for each (Figure 6.25). Predictions were also determined by supplying separately the N-terminal DUF and the oxidase portions of each protein. The N-terminal DUF portions of PatGox, CyaGox and ArtGox all gave moderate structural predictions (confidence 65 - 74 %) based on the peptide-binding domain of NisB (PDB: 4WD9, [30]) from the Nisin pathway (a member of the lanthipeptide class of RiPPs discussed in Chapter 1), whilst ArtGox and CyaGox additionally gave predictions (65 and 83 % confidence respectively) based on the peptide-binding domain of TruD (PDB 4BS9, [67]) from the trunkamide pathway (Chapter 1). The remaining oxidase portions of the proteins all gave strong predictions (confidence > 99 %) based on standard oxidoreductase domains such as those from *anabena varibilis* (PDB code: 3EO7) and *ralstonia eutropha* (PDB code: 3HJ9) (Table 6.6).

Protein	**Domain**	**Residues Searched**	**Residues Aligned**	**Model**	**Residues Aligned**	**Prediction Confidence**
PatGox	N-term DUF	1 - 197	14 - 101	NisB	140 - 229	67 %
ArtGox	N-term DUF	1 - 196	7 - 101	NisB	141 - 229	65 %
ArtGox	N-term DUF	1 - 196	5 - 95	TruD	3 - 80	65 %
CyaGox (PCC 7425)	N-term DUF	1 - 194	7 - 92	NisB	140 - 229	74 %
CyaGox (PCC 7425)	N-term DUF	1 - 194	7 - 87	TruD	6 - 81	83 %
PatGox	Oxidase	198 - 486	19 - 304	Flavoprotein (*a. variabilis*)	9 - 258	100 %
PatGox	Oxidase	198 - 486	46 - 285	Oxidoreductase (*r. eutropha*)	16 - 214	100 %
ArtGox	Oxidase	197 - 481	19 - 285	Flavoprotein (*a. variabilis*)	9 - 239	100 %
ArtGox	Oxidase	197 - 481	48 - 283	Oxidoreductase (*r. eutropha*)	16 - 211	100 %
CyaGox (PCC 7425)	Oxidase	195 - 473	19 - 275	Flavoprotein (*a. variabilis*)	9 - 241	100 %
CyaGox (PCC 7425)	Oxidase	195 - 473	48 - 274	Oxidoreductase (*r. eutropha*)	16 - 214	100 %

Table 6.6: Phyre2 Structural Predictions.

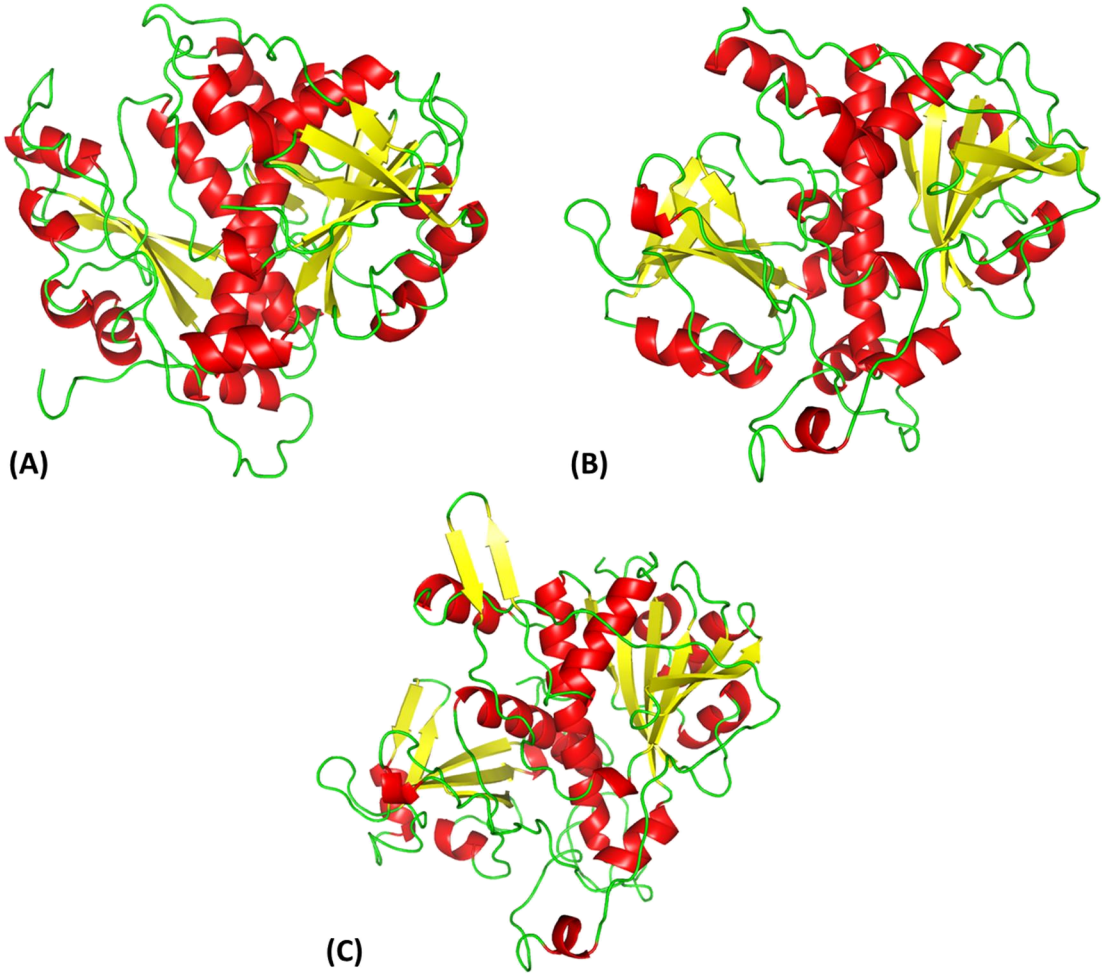


Figure 6.25: Homology Models of Oxidase Proteins: Phyre2 derived homology models of (A) PatGox, (B) ArtGox and (C) CyaGox (PCC7425).

6.4 Discussion

Isolation of the oxidase domain of the PatG protein has not been possible to date as difficulties with aggregation or no expression at all has hampered efforts. Attempts to improve soluble expression by varying cell lines, induction conditions and also the domain boundaries failed.

As a result of the lack of PatGox protein, efforts were shifted to related cyanobactin oxidases ArtGox from *Arthrospira platensis* and CyaGox from *Cyanothece sp*. PCC 7425 and both of these have been expressed and purified. Crystals have been obtained for both proteins, with optimised crystals and native datasets in complex with two substrate peptides for CyaGox. At present, no phases have been calculated to allow structure determination however efforts are on-going.

Homology models of PatGox, ArtGox and CyaGox (PCC 7425) have been created using Phyre2 prediction software [125]. The proteins predominantly show structural homology to known oxidoreductases, however this analysis has determined that the N-terminal DUF of the proteins has structural homology to the peptide-clamp domains of NisB and TruD proteins. This finding could provide a theory for why the oxidase protein can function on both the macrocyclic and linear substrates, because the full length linear peptide can be held in place by the clamp domain.

Finally, activity of the ArtGox enzyme on both the final macrocycle and the full length PatE has been determined by MS. The MS analysis suggests that ArtGox processes the macrocycle faster than the linear peptide which would imply that in nature the enzyme works on the macrocycle. However, detailed kinetic experiments will have to be carried out to fully ascertain this. The preference for the macrocycle over the linear peptide would correlate with the hypothesis that epimerisation (which must occur before oxidation) occurs on the macrocycle [94] [95].

6.5 Conclusions and Future Work

The oxidase enzyme, PatGox, from the patellamide pathway could not be expressed in a soluble state and therefore cyanobactin homologues ArtGox and CyaGox which are soluble have been purified and studied instead.

It has been determined that ArtGox is an active oxidase for thiazolines when incorporated in both a full length linear peptide and a macrocycle. Our collaborators (University of Aberdeen) reported that CyaGox (PCC 7425) is active on the macrocycle, [69] however studies by Greg Mann (Naismith Group, University of St Andrews) have since shown activity on the linear peptide too. At present, no kinetic experiments have been carried out on the oxidase enzymes and this is work to be undertaken. The ability to oxidise thiazolines to thiazoles now allows us to add yet more diversity to our *in vitro* compound library. There are cyanobactins from other pathways which also contain oxazoles, oxidised oxazolines (e.g. - Tenuecyclamides [36]), and the oxidase enzymes responsible for this reaction will be explored, which could enable us to introduce oxazoles to further diversify our macrocycle library.

For both enzymes, crystallisation trials were attempted and small initial crystals were created in each case, with the CyaGox optimised with two substrate peptides to give native datasets to 2.97 and 2.65 Å. As yet phases have not been found to allow the crystal structure to be determined and work on this is on-going. SeMet protein did not crystallise, but we have plans to make SeCys protein for Se-SAD experiments and also to send native crystals to the long wavelength beamline I23 at the Diamond Light Source to attempt sulfur-SAD. Should neither of these techniques work we will mutate the protein to selectively remove each methionine in turn and attempt SeMet crystal growth or alternatively mutate in a free cysteine which should improve heavy atom phasing with for example mercury derivatives.

Finally, the absence of a crystal structure has led us to structure predictions which appear to suggest a standard FMN binding domain responsible for oxidation but with an additional N-terminal DUF domain. This small N-terminal DUF has some homology to the peptide-binding clamp in TruD, which could explain the activity on the linear peptide. A future crystal structure should elucidate this further.

7. Unnatural Amino Acid Incorporation into Cyclic Peptides

7.1 Introduction

It has been demonstrated that patellamide-like cyclic peptides can be derived from the reconstitution of the patellamide biosynthetic pathway *in vitro* (Chapter 4). There are strict limitations in the composition of these peptides to amino acids which can be incorporated *in vivo* to the precursor peptide PatE. We have explored the incorporation of selenium variants with a some success, but selenium is known to be accepted into cells (Chapter 5). Taniero *et al.* (2012) [103] have exploited the pEVOL system pioneered by the group of Peter Schultz (Scripps Research Institute) [160], [161] to introduce non-proteinogenic amino acids into *E. coli* cells expressing the patellamide biosynthetic pathway *in vivo*. This *in vivo* approach has highlighted the potential in adapting enzymatic pathways to yield diversified products however the quantities of the target compound isolated in this study remain low (5 – 174 µg compound L^{-1} cell culture).

We sought to establish a method of producing macrocycles containing unnatural amino acids in higher yields (mg quantities or above) and as such we explored two different approaches; intein-mediated peptide ligation and the use of the pEVOL system to generate substrate for *in vitro* processing.

7.1.1 Intein Technology

Semi synthetic approaches have been employed in a variety of biotechnologies. Examples include large scale drug production from a natural product starting material expressed in cells (e.g. Yondelis®, an anti-tumour drug [162] [163]) or the engineering of the protein backbone of HIV protease by ligating two peptides by a thioester bond [164].

One class of semi-synthetic method is the exploitation of intein proteins. Inteins, first identified in 1990 [165], are protein elements which self-excise themselves from a longer protein and catalyse the subsequent ligation of their N- and C- terminal flanking proteins (Figure 7.1) [166]. This is conceptually similar to the well-known pre-mRNA splicing where RNA introns are removed and the exons are ligated together; to make the analogy explicit, the protein equivalent is termed 'protein splicing' with inteins and exteins. [167].

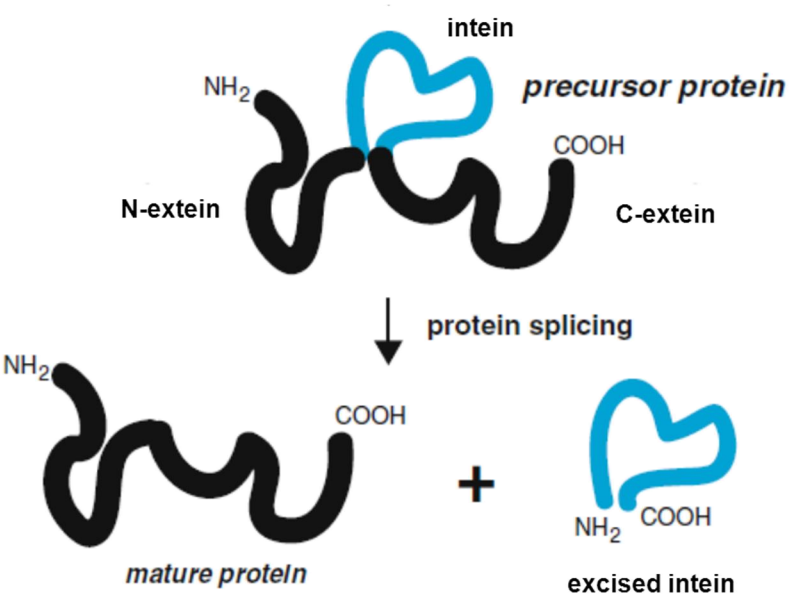


Figure 7.1: Intein Protein Splicing. Schematic of intein protein splicing where the central intein protein element self excises resulting in the two flanking extein segments ligating together to form the mature peptide while releasing the intein. Figure adapted from Elleuche and Poggeler (2010) [166].

Inteins have been exploited in biotechnology for multiple purposes including cleavage of affinity tags in protein purifications [168], semi-synthesis of cytotoxic proteins [169], semi-synthetic protein ligation [170] and peptide cyclisation [171].

Of particular interest to this project is the intein-mediated protein ligation (IPL) method which uses the intein to chemically ligate two peptides to form a larger single peptide with a standard peptide bond linking the two (Figure 7.2) [172]. This occurs by

expressing a protein-intein fusion and purifying it using standard protein purification techniques. The intein is then removed using a free thiol reagent, the most common being sodium 2-sulfanylethanesulfonate (MESNA) or 4-mercaptophenylacetic acid (MPAA) (Figure 7.3), which reacts at the carbonyl carbon to form an activated peptide while releasing the intein. The activated peptide is then reacted with a C-terminal peptide with a cysteine at its N-terminus to form the complete peptide. The C-terminal peptide can be expressed heterologously or synthesised by chemical means.

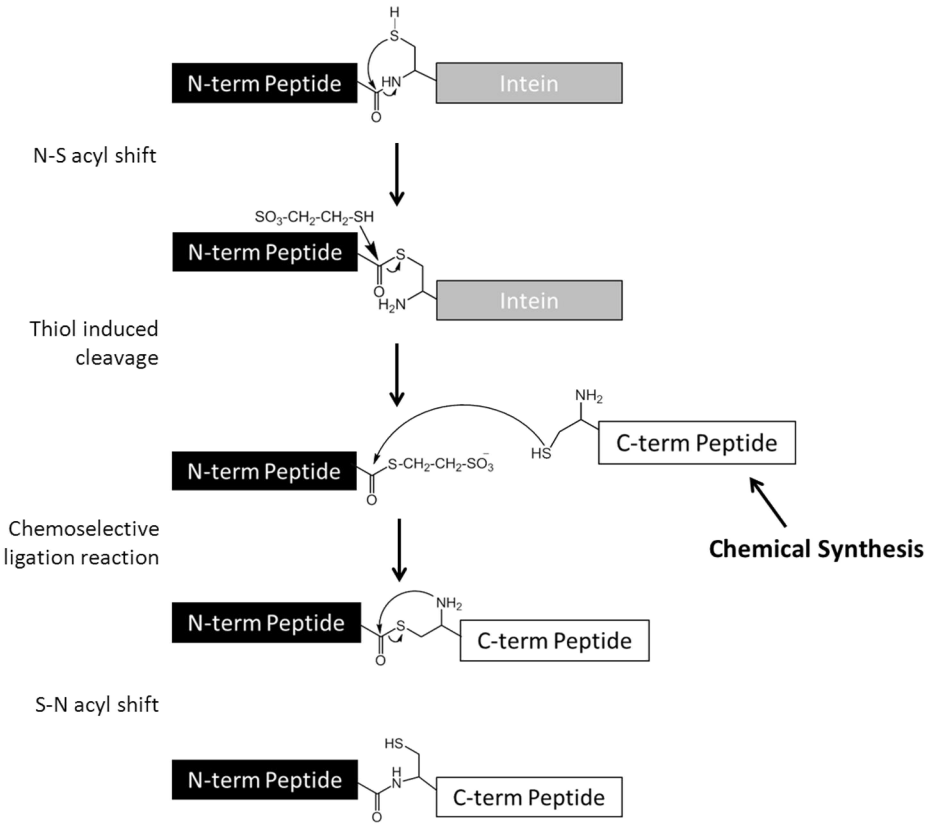


Figure 7.2: Intein-mediated Protein Ligation (IPL). Schematic of intein-mediated protein ligation where the N-terminal peptide-intein fusion undergoes an N-S acyl shift followed by cleavage with MESNA resulting in activated N-terminal peptide and removal of intein. A secondary peptide (which can be chemically synthesised) with an N-terminal Cys is added and reacts with the activated peptide to form a combined peptide. Finally, the longer peptide undergoes an S-N acyl shift to yield a new peptide with native peptide bond. (Adapted from Xu and Evans, 2001) [172]

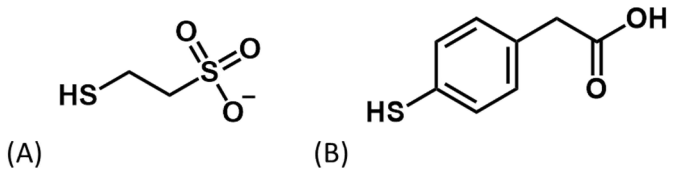


Figure 7.3: Intein Cleavage Thiols. Chemical structures of two thiols commonly used in the cleavage step of intein-mediated peptide ligation – (A) sodium 2-sulfanylethanesulfonate (MESNA) (B) 4-mercaptophenylacetic acid (MPAA)

It was planned to generate an intein-mediated protein ligation system to fuse the PatE leader sequence to chemically synthesised core peptides containing unnatural amino acids plus the C-terminal recognition site 'AYDG' (Figure 7.4). The ligated peptides would then be processed *in vitro* to create macrocyclic peptides containing unnatural amino acids.

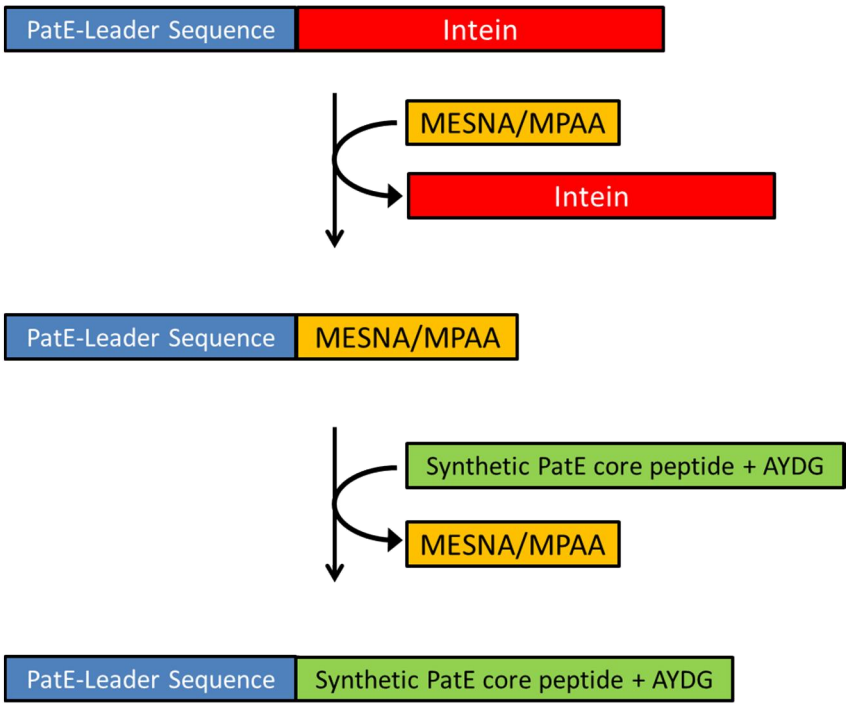


Figure 7.4: Semi Synthetic PatE Production. Schematic of the production of a semi-synthetic PatE by thiol-mediated cleavage of PatE-Intein fusion protein to generate an activated PatE leader sequence. The activated peptide is then ligated by a free cysteine on the N-terminal of a synthetic PatE core peptide.

7.1.2 pEVOL Amber Stop Codon Technology

Genetically encoding unnatural amino acids directly into proteins *in vivo* using heterologous systems opens up a wide potential for the use in biochemical applications. The ability to reassign codons to achieve this was first recognised by Normanly *et al* (1990) [173], however it was the laboratory of Peter Schultz (The Scripps Research Institute) which pioneered this field with the development of technology using evolved M*ethanocaldococcus jannaschii* aminoacyl-tRNA synthetase (aaRS)/suppressor tRNA pairs to incorporate unnatural amino acids in response to the amber stop codon 'TAG' [174]. The pairs are constructed in a fashion that means they are not recognised by endogenous aaRSs or endogenous tRNAs of the host organism respectively, thus are orthogonal. The synthetase specificity is altered so that the tRNA is only loaded with the unnatural amino acid of interest [175] [176] (Figure 7.5)

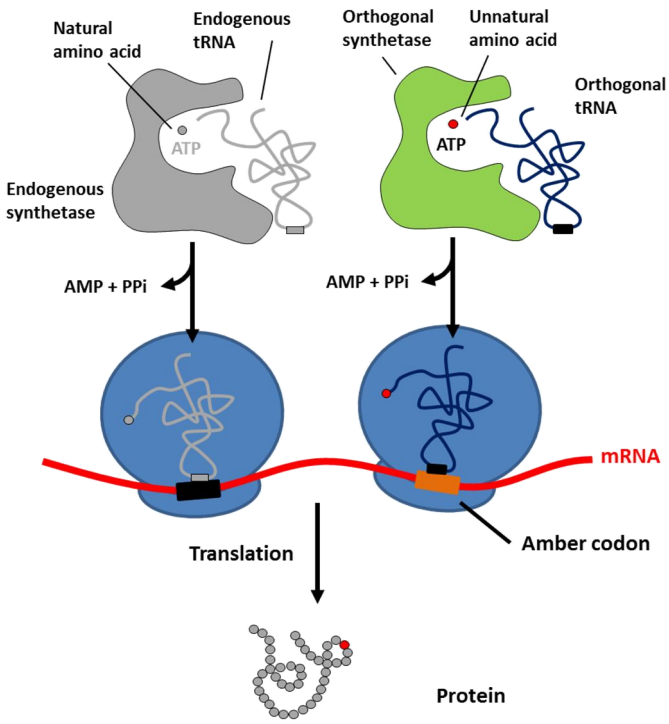


Figure 7.5: Unnatural Amino Acid Incorporation. Schematic of orthogonal aminoacyl-tRNA synthetase (aaRS)/suppressor tRNA pairs incorporating unnatural amino acids in response to the amber stop codon during protein synthesis. Figure adapted from Lin and Wang, 2008 [176] .

More than seventy unnatural amino acids have been genetically encoded in bacteria, yeast, and mammalian cells using this method and a selection of the various amino acids is represented in Figure 7.6 [175] [177].

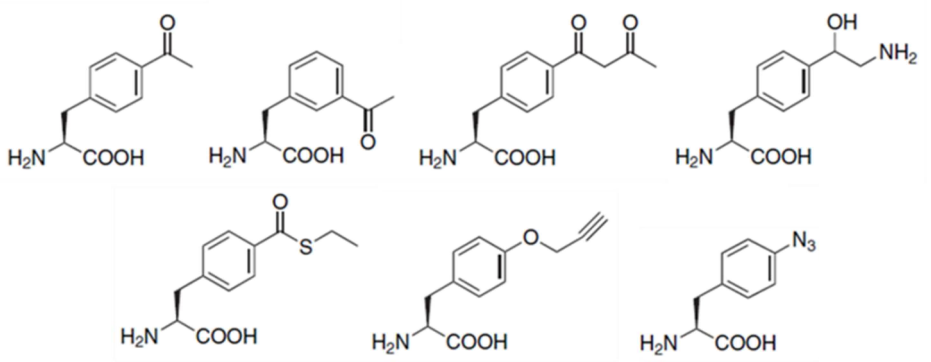


Figure 7.6: Genetically Encoded Unnatural Amino Acids. A small representation of the many unnatural amino acids which have been genetically encoded in *E. coli*, yeast or mammalian cells. Figure adapted from Wang *et al*, 2006 [175].

The development of this technology has been implemented successfully in many areas of biochemistry such as the addition of fluorophores [178], the addition of photo-cross-linkers [179], site directed spin-labelling for Electron Pulse Resonance (EPR) [180], introduction of glycosylated residues in *E. coli* derived proteins [181] and heavy atom phasing for X-ray crystallography [182].

Following on from their initial successes, the Schultz group developed a single *E. coli* vector, pEVOL, containing two copies of the *M. jannaschii* aaRS gene along with an optimised suppressor tRNA [160]. The vector can be simply co-expressed heterologously in common *E. coli* expression cells (e.g. BL21 DE3) with a standard vector containing the gene of interest (e.g. pET system [158]) complete with the amber stop codon 'TAG' in the specified position for the unnatural amino acid [160]. The vectors are expressed under the influence of different inducers (L-arabinose vs IPTG) with the expression resulting in the protein of interest with the unnatural amino acid incorporated. A range of different aaRS pEVOL plasmids that incorporate different

unnatural amino acids have been developed with a small selection being made available commercially through Addgene [160].

This study proposed to create a PatE containing the unnatural amino acid p-benzoyl-L-phenylalanine (pBpa) (Figure 7.7) using the pEVOL system [160] [183] and to process it *in vitro* to form a macrocycle containing an unnatural amino acid.

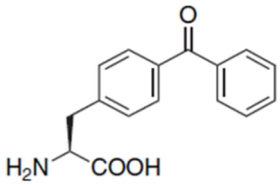


Figure 7.7: p-Benzoyl-L-Phenylalanine. Chemical structure of the amino acid p-benzoyl-L-phenylalanine (pBpa).

7.2 Intein-Mediated Approach

7.2.1 Materials and Methods

7.2.1.1 Construct Design and DNA Cloning

A DNA construct of full length PatE leader sequence, lysine and the first three residues of the PatE2 core peptide were combined at the C-terminus with the Mxe GyrA Intein (txb1, New England Biosciences, [184] [185]) using fusion PCR with the primers listed in Table 7.1

The construct was created by a three step PCR process. Firstly, the PatE leader section was amplified by PCR using primers to give an *Nco1* site at the 5' end and the first eight amino acids from the cleavable cysteine of the intein on the 3' end. Next, the Intein sequence was amplified by PCR using primers to give the final nine residues of the PatE to be included at the 5' end, and an *Xho1* site at the 3' end. Finally, the two newly formed DNA fragments were fused by PCR using the 5' primer of the first reaction and the 3' primer of the second to give the full length DNA insert to be used. (Figure 7.8)

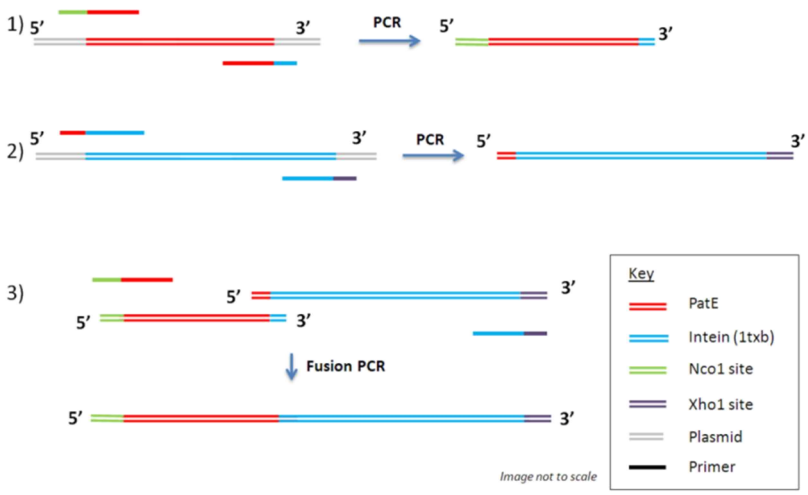


Figure 7.8: PatE Intein Fusion PCR Schematic. 3 Step PCR schematic to produce a fusion of PatE leader sequence and intein (1txb). 1) N-terminal PatE leader sequence production 2) C-terminal Intein (1txb) production 3) PCR to fuse the two sequences together to from the final DNA product.

5′PatE	cttccatggacaaaaaaaacattcta
3′PatEInt	aactagtgcatctcccgtgatgcaagcagttattttagatgcttccaaacc
5′PatEint	ggtttggaagcatctaaaataactgcttgcatcacgggagatgcactagtt
3′Int	cctctcgagggtagcgtgagatacgaaccc
5′PatEIntSmall	cttccatggatgctggtttggaagcatct
5′PatEIntLarge	cttccatgggcttggctgaactgtctgaggaa
5′PatEIntGLEA	ttggaagcatgcatcacgggagatgcactagttgcccta
3′PatEIntGLEA	cgtgatgcatgcttccaaaccagcatcgccgagagcttc
3′PatE-CKITAPITWP	cttctcgagttcaccatcataagccggccaggtaatcggcgcggtaattttg cactgaaaatacag

Table 7.1: PatE - Intein Mutagenesis Primers. Oligonucleotide sequences for PatE – Intein fusion, truncation and deletion PCR and for model C-terminal peptide CKITAPITWP.

The new DNA insert was gel purified, extracted and ligated into the pBMS23CHis plasmid using the methods described in Section 2.2.1 to give the protein construct "PatE-Intein Full Length" (Figure 7.9 A)

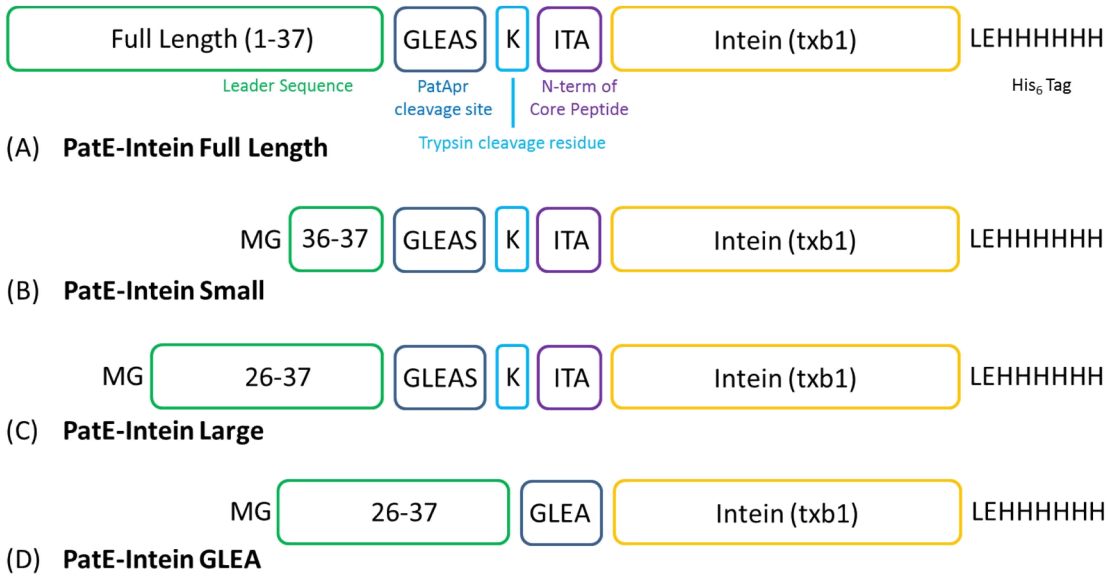


Figure 7.9: PatE-Intein Construct Schematic. Construct diagrams for A) PatE-Intein Full Length, B) PatE-Intein Small C) PatE-Intein Large and D) PatE-Intein GLEA.

Two shorter variants of the full length PatE-Intein fusion were created using standard PCR methods (primers - Table 7.1) in which the PatE leader sequence was truncated from the N-terminal end. The two new constructs were designated 'PatE-Intein-Small' (Figure 7.9 B) and 'PatE-Intein-Large' (Figure 7.9 C). A final construct based on 'PatE-Intein Large' where the intein section of the DNA starts directly after the 'GLEA' of the GLEAS motif was created using PCR mutagenesis with the primers listed in Table 7.1, 'PatE-Intein-GLEA' (Figure 7.9 D).

The test peptide 'CITAPITWPAYDGELEHHHHHH' was created from a full length PatE based on the PatE2TEV construct. The full length PatE sequence was created by mutagenesis of PatE2TEV (section 4.2.1) using the primers listed in Table 7.1 and the PCR method described previously (Section 2.2.1).

7.2.1.2 Expression and Purification

Full length PatE-Intein was expressed from the pBMS23CHis plasmid in BL21 (DE3) *E. coli* cells grown in LB medium at 37 °C to an optical density at 600 nm (OD_{600}) of 0.6. The cultures were then induced with 1 mM IPTG and grown for 24 hours at 37 °C. PatE-Intein Small, Large and GLEA were expressed from the pBMS23CHis plasmid in BL21 (DE3) *E. coli* cells grown in LB medium at 37 °C to an OD_{600} of 0.6. The cultures were then induced with 0.5 mM IPTG and grown for three hours at 37 °C. All cells were harvested by centrifugation (4000 x g).

Full length PatE-Intein was purified by re-suspending cell pellets in urea lysis buffer (Appendix A.12) which were then lysed by sonication at 15 microns (SoniPrep 150, MSE), cleared by centrifugation (40,000 x g, 4 °C, 20 min) and loaded onto a Ni-Sepharose 6 FF column (GE Healthcare) equilibrated in urea lysis buffer. The column was washed with lysis buffer and the protein eluted with elution buffer. The protein was refolded by step dialysis from 8 M to 6 M, 4 M, 2 M, 1 M and final 0 M Urea Lysis

Buffer. Each step was carried out over four hours or overnight at 20 °C with the exception of the final step from 1 M to 0 M urea which occurred overnight at 4 °C.

PatE-Intein Small, Large and GLEA were purified by re-suspending whole cells in lysis buffer (Appendix A.13). The cells were lysed by passage through a cell disruptor at 30 kPSI (Constant Systems) and the lysate cleared by centrifugation (40,000 x g, 4 °C, 20 min) and then loaded onto a Ni-Sepharose 6 FF column (GE Healthcare) equilibrated with lysis buffer. The column was washed with lysis buffer and the protein was eluted with elution buffer. The eluted fraction was subjected to size-exclusion chromatography (Superdex 75, GE Healthcare) in gel filtration buffer where the protein eluted as a single monomer peak. The samples were visualised by SDS-PAGE and their identity confirmed by MS.

7.2.1.3 C-terminal Peptide Production

The precursor PatE of the test peptide 'CKITAPITWPAYDGELEHHHHHH' was expressed and purified as for PatE2K (Chapter 4.2.2). The mature peptide was created by cleavage at the TEV protease recognition site using 1 mg TEV protease per 10 mg of PatE and incubated at room temperature for 4 hours. The cleaved protein was purified on a Superdex S30 column (GE Healthcare) in gel filtration buffer. The peak fractions were confirmed by MS and concentrated to 1 mM.

Two synthetic peptides 'CIS[Bpa]CAYDGE' and 'CKI[Thi]ACI[Hyp]APAYDG' containing the unnatural amino acids Benzophenylalanine (Bpa), Thienylalanine (Thi) and Hydroxyproline (Hyp) were synthesised by the laboratory of Tom Muir (Princeton University) and purchased from Peptide Protein Research respectively (Figure 7.10).

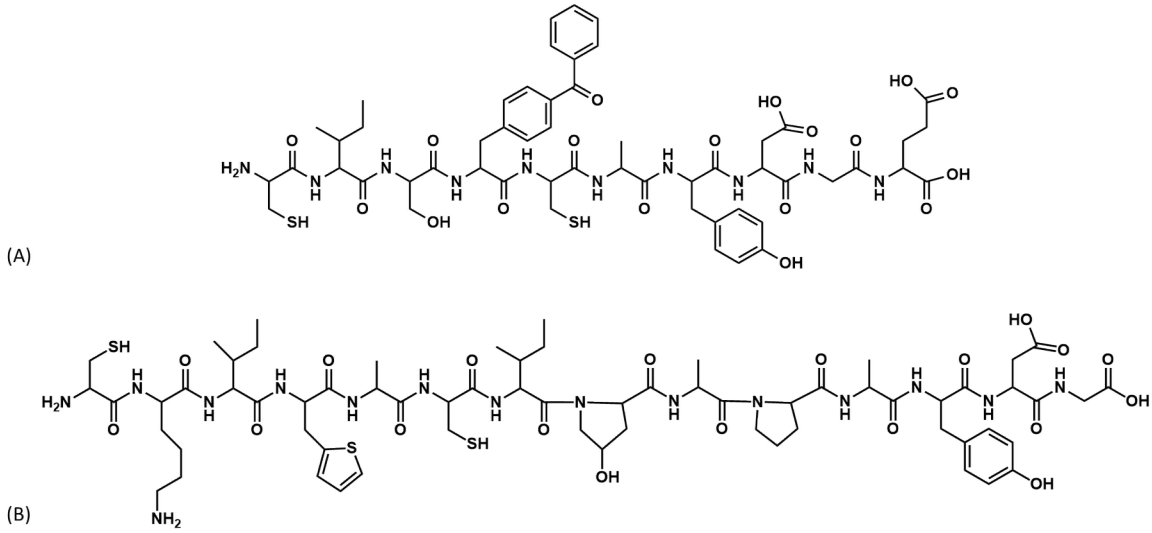


Figure 7.10: Chemical Structures of Synthetic Peptides. Chemical structures of chemically synthesised peptides with amino acid sequences A) CIS[Bpa]CAYDGE and B) CKI[Thi]ACI[Hyp]APAYDG.

7.2.1.4 Thiol-mediated Cleavage

For full length PatE-Intein, PatE-Intein-Small and PatE-Intein-Large (Figure 7.9); thiol-mediated cleavage was achieved by the addition of 200 mM MESNA (sodium 2-sulfanylethanesulfonate) or MPAA (4-mercaptophenylacetic acid). The cleavage was carried out in the PatE-Intein gel filtration buffer and incubated at room temperature for 24 hours. Reactions were purified on a Superdex S30 column (GE Healthcare) and the identities of the thiol-activated PatE leader sequences were confirmed by MS.

7.2.1.5 Peptide Ligation

Ligation of PatE-Intein Small was abandoned after we showed that TruD could not process fully with such a small leader sequence [67]. It was, however, confirmed that TruD would process a leader sequence of the same size as PatE-Intein-Large. PatE-Intein-Large was therefore prioritised for ligation reactions by incubating 50 µM PatE-

Large-MESNA/MPAA with 100 µM C-terminal peptide in 10 mM HEPES pH 7.4, 150 mM NaCl, 5 mM TCEP for 48 hours at 37 °C

7.2.1.6 One Step Thiol Cleavage/Peptide Ligation

One step thiol cleavage and peptide ligation was carried out using PatE-intein-large and PatE-intein-GLEA with the various C-terminal peptides. This was achieved by incubating the PatE-Intein and C-terminal peptide each at a concentration of 100 µM in a buffer of 150 mM NaCl, 10 mM HEPES pH 7.4, 20 mM TCEP and 200 mM MPAA for 24 hours at room temperature. The ligated peptide was purified on a Superdex S30 column (GE Healthcare) in gel filtration buffer and its identity was confirmed by MS.

7.2.1.7 *In Vitro* Processing

The newly formed ligated peptides were processed by the heterocyclase TruD to assess activity. The peptide produced by ligation of PatE-Intein-GLEA and the C-terminal peptide CKI[Thi]ACI[Hyp]APAYDG was subsequently treated with trypsin and the macrocyclase enzyme PatGmac to yield a macrocycle. These steps were achieved as per the protocol used for PatE2K outlined in Sections 4.2.4 – 4.2.6

7.2.1.8 Mass Fragmentation

The final macrocycle, containing unnatural amino acids, was subjected to LC-MSMS using an AB SCIEX Triple TOF 5600 system with the fragment peaks assigned to confirm the structure. This work was carried out by Dr. Sally Shirran (University of St Andrews).

7.2.2 Results

7.2.2.1 Expression and Purification

Full length PatE-Intein was overexpressed with a C-terminal His_6-tag in BL21 (DE3) *E. coli* cells. The protein was solubilised from inclusion bodies in urea lysis buffer, isolated by Ni-NTA chromatography and refolded using step dialysis into lysis buffer containing no urea. The final protein yield was 30 - 40 mg L^{-1} culture.

PatE-Intein Small, Large and GLEA were overexpressed with a C-terminal His_6-tag in BL21 (DE3) *E. coli* cells. The peptide was purified by Ni-NTA chromatography and elutes off gel filtration as a single peak (Figure 7.11 A). The final protein yield is 40 - 60 mg L^{-1} culture. The purity of all proteins was determined by SDS-PAGE (Figure 7.11 B)

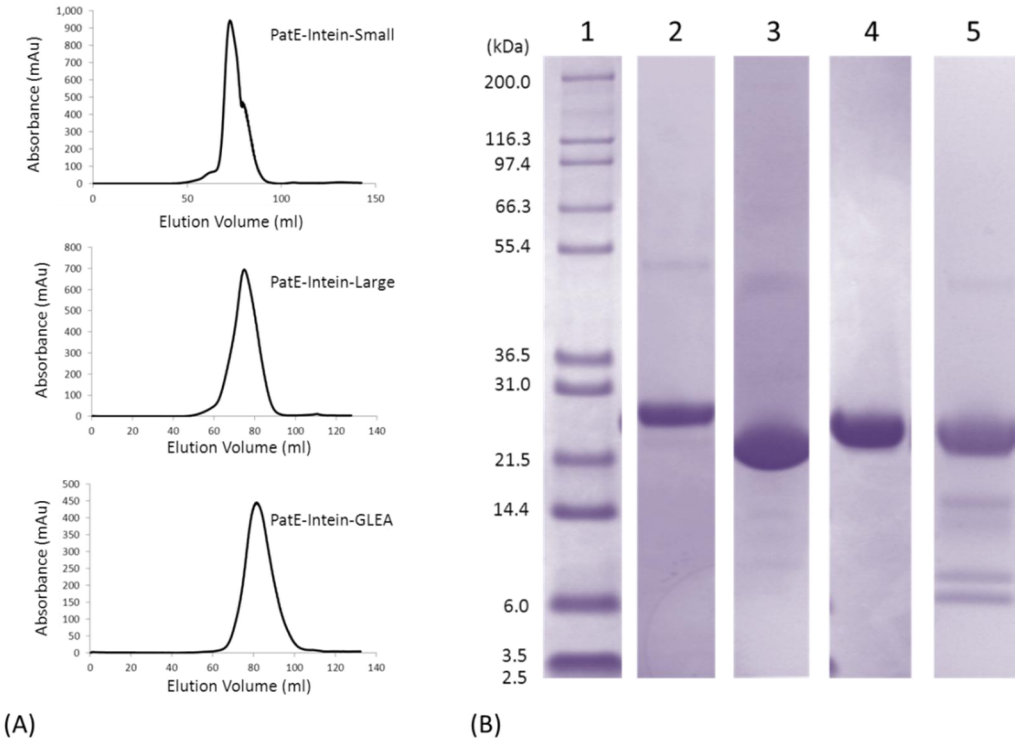


Figure 7.11: Gel Filtration and SDS-PAGE Analysis of PatE-Intein. A) Gel filtration UV traces of PatE-Intein-Small, PatE-Intein-Large and PatE-Intein-GLEA B) SDS-PAGE of: 1) Molecular Markers 2) PatE-Intein Full Length 3) PatE-Intein Small 4) PatE-Intein Large and 5) PatE-Intein-GLEA. All protein samples are > 95 % pure.

7.2.2.2 C-terminal Peptide Production

C-terminal peptides 'CIS[Bpa]CAYDGE' and 'CKI[Thi]ACI[Hyp]APAYDG' were synthesised and HPLC purified by collaborators/suppliers.

The C-terminal peptide 'CKITAPITWPAYDGELEHHHHHH' was expressed as a PatE precursor overexpressed with a C-terminal His_6-tag in BL21 (DE3) *E. coli* cells using the Studier auto-induction method [107]. The peptide was solubilised from inclusion bodies in urea lysis buffer, isolated by Ni-NTA chromatography and refolded using DTT. The peptide elutes off gel filtration as a single peak (Figure 7.12 A, B). The final protein yield is 50 mg L^{-1} culture. The mature peptide was isolated by cleavage with TEV protease and purified on a Superdex S30 with a single main peak corresponding to the cleaved peptide (Figure 7.12 C). A final mature peptide yield of 10 mg L^{-1} original culture was achieved and confirmed by MS (Figure 7.13).

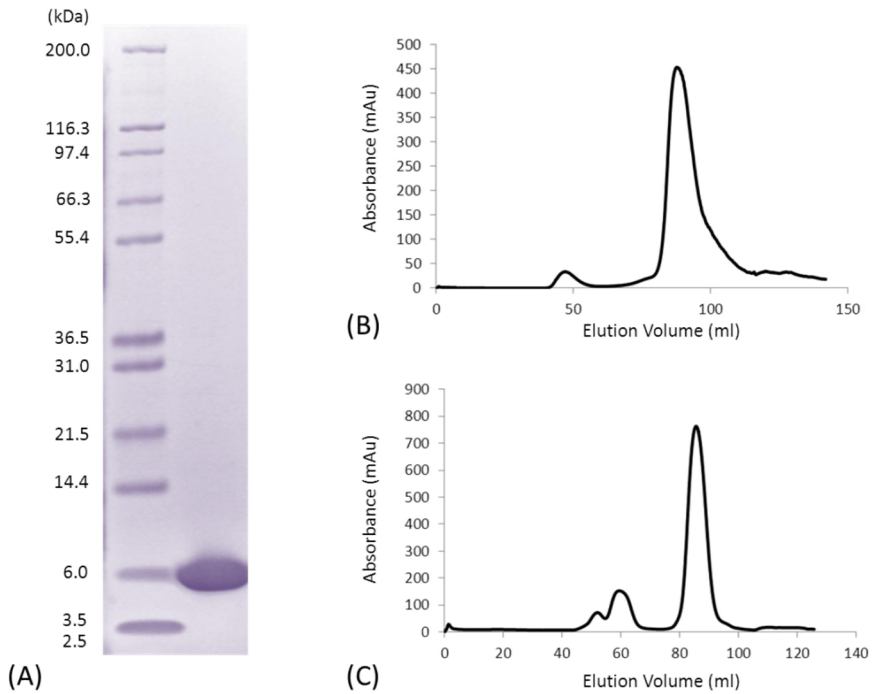


Figure 7.12: SDS-PAGE and Gel Filtration Analysis of CKITAPITWPAYDGELEHHHHHH. A) SDS-PAGE of purified PatE precursor peptide B) Superdex 75 gel filtration UV trace of the precursor PatE C) Superdex 30 gel filtration UV trace of the mature peptide.

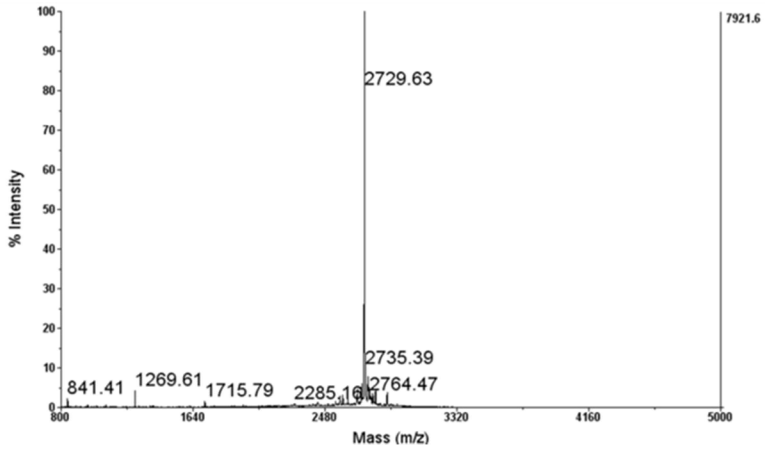


Figure 7.13: MALDI MS of CKITAPITWPAYDGELEHHHHHH. MALDI-TOF-MS of the mature peptide CKITAPITWPAYDGELEHHHHHH derived from full length PatE (Expected M+H = 2729.4).

7.2.2.3 Thiol-mediated Cleavage

PatE-intein-Full Length, PatE-intein-Small and PatE-intein-Large were all subjected to thiol cleavage by MESNA and MPAA. The full length PatE variant precipitated immediately on addition of the thiol and so work on this was stopped. PatE-intein-Small and PatE-intein-Large were cleaved from the intein portion yielding an activated PatE leader sequence. These samples were purified on a Superdex S30 column with great difficulty due to a lack of absorbing residues at Abs_{280nm}. A small bump on the gel filtration trace (Figure 7.14) was determined by MS to be the activated peptide.

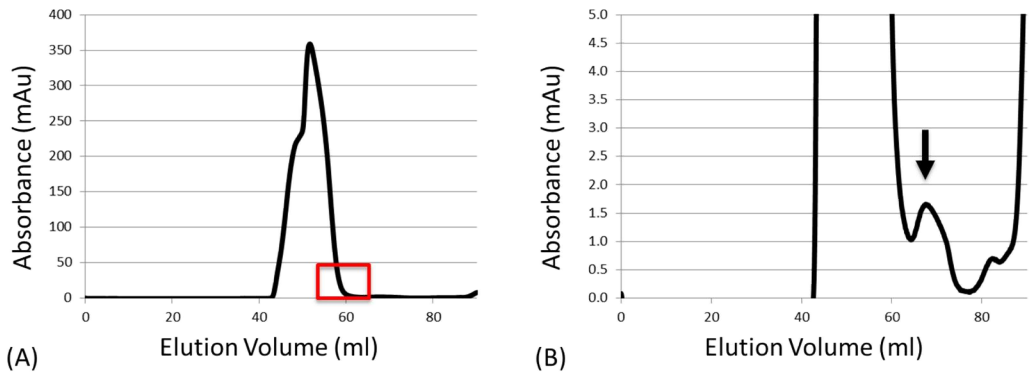


Figure 7.14: Gel Filtration of PatE-Large-MESNA. Gel filtration UV trace of PatE-Intein-Large post treatment with MESNA. (A) Full trace (B) Zoom view (red box). The small bump at an elution volume of 65 ml (indicated by black arrow) corresponds to activated PatE-Large-MESNA. The large peak in the trace is the excised intein protein.

MSMS was carried out on both PatE-Small-MESNA and PatE-Large-MESNA complexes which confirmed the presence of the MESNA group (Figure 7.15).

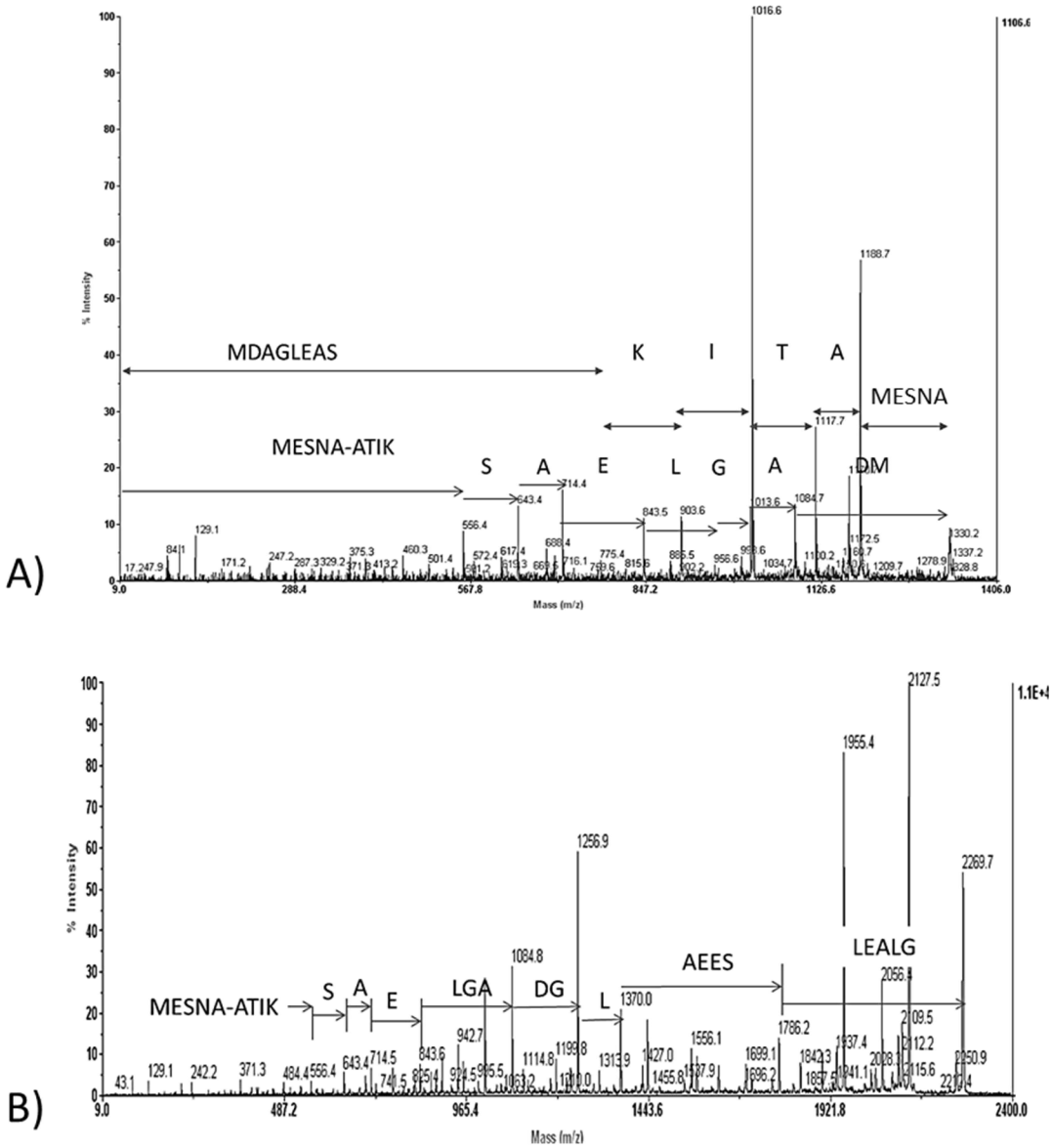


Figure 7.15: MSMS of Activated PatE Constructs. MALDI TOF MSMS and amino acid peak assignment of A) PatE-Small-MESNA and B) PatE-Large-MESNA.

At this stage work on PatE-small-MESNA was halted as it was discovered at the same time that TruD could not process efficiently using the small leader sequence. TruD could however process efficiently peptides containing the larger leader sequence of PatE-large-MESNA and so this was carried forward to peptide ligation trials.

7.2.2.4 Peptide Ligation

PatE-Large-MESNA was incubated separately with two fold excess of the C-terminal peptide 'CIS[Bpa]CAYDGE' and incubated at 37 °C for 48 hours. The reaction mixture was analysed by MS to identify peptide ligation. Very small levels of ligation were evident (Figure 7.16) however there remained significant amounts of the starting material (as judged by MS). The reaction was left for one week. However, no further ligation was evident suggesting that the reaction conditions require optimisation.

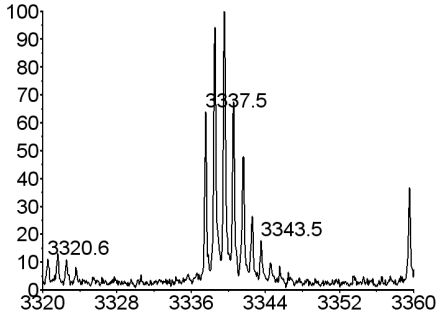


Figure 7.16: MS of Ligated PatE-Large-CIS[Bpa]CAYDGE Peptide. MALDI-MS analysis of the product from the ligation of PatE-Large-MESNA plus CIS[Bpa]CAYDGE synthetic peptide showing $[M+H]^+$ of 3337.5 at very low intensity.

7.2.2.5 One Step Thiol Cleavage / Peptide Ligation

Ligation as a two-step process was very inefficient giving low yields of ligated peptide. In order to assess a one-step ligation, a C-terminal peptide CKITAPITWPAYDGELEHHHHHH was created using a variant of the full length PatETEV created in Section 4.2.1. The addition of 200 mM MPAA and 20 mM TCEP yielded the most efficient ligation when peptides were in a 1:1 ratio (Figure 7.17).

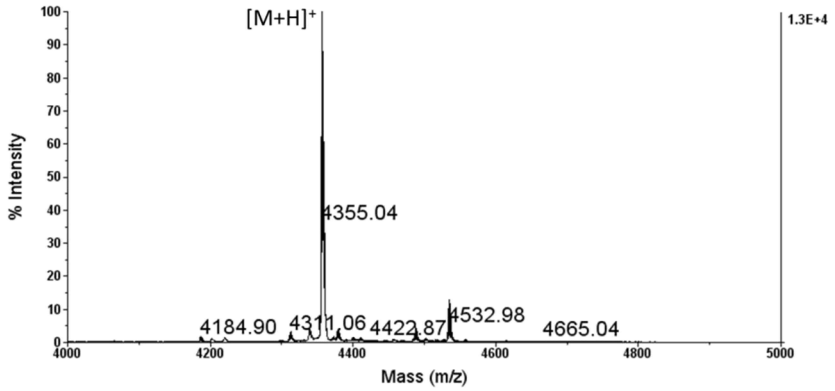


Figure 7.17: MALDI MS of PatE-Intein-GLEA Ligation with CKITAPITWPAYDGELEHHHHHH. MALDI TOF MS of the resultant ligated peptide PatE-GLEA-CKITAPITWPAYDGELEHHHHHH (Expected $M+H^+$ = 4355.05).

In order to demonstrate that this procedure could be carried out on more than one peptide and with varying lengths of PatE-Leader/C-terminal peptide. PatE-Intein-Large was reacted with the C-terminal peptide CIS[Bpa]CAYDGE and analysed by MALDI MS (Figure 7.18). PatE-Intein-GLEA was reacted with the C-terminal peptide CKI[Thi]ACI[Hyp]APAYDG and analysed by MALDI MS (Figure 7.19). In both cases the predominant MS peaks corresponded to ligated peptide. These samples were purified from the remaining intein portion and any non-ligated C-terminal peptide by passing over a Superdex 30 column.

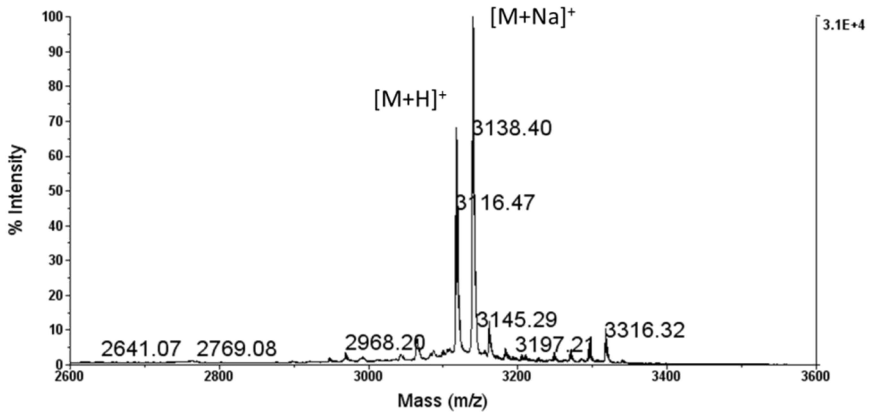


Figure 7.18: MALDI MS of PatE-Intein-Large Ligation with CIS[Bpa]CAYDGE. MALDI TOF MS of the resultant ligated peptide PatE-Large-CIS[Bpa]CAYDGE with $[M+H]^+$ and $[M+Na]^+$ peaks observed.

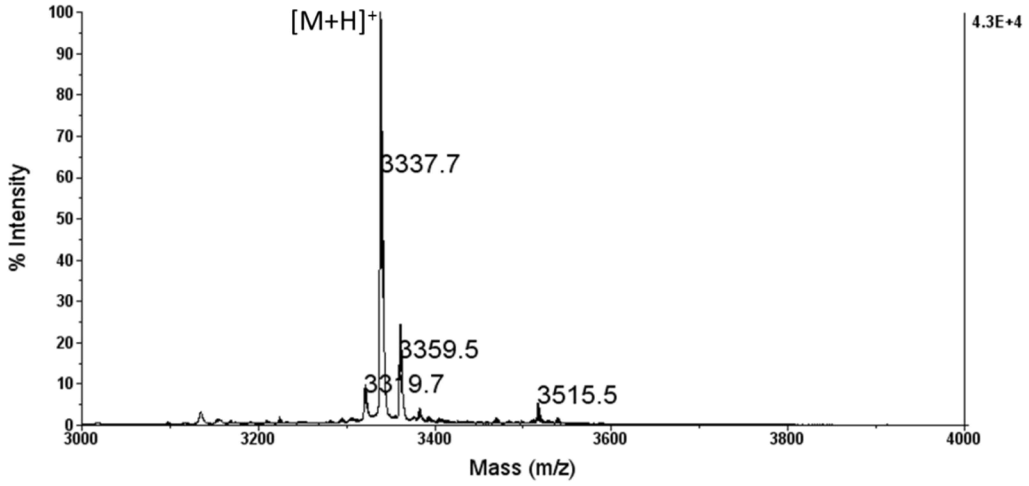


Figure 7.19: MALDI MS of PatE-Intein-GLEA Ligation with CKI[Thi]ACI[Hyp]APAYDG. MALDI TOF MS of the resultant ligated peptide PatE-GLEA-CKIT[Thi]ACI[Hyp]APAYDG.

7.2.2.6 *In Vitro* Processing

The ligation of PatE-GLEA with CKITAPITWPAYDGELEHHHHHH was instrumental in determining the correct conditions for the ligation. This peptide however does not contain any cysteines for heterocyclase treatment by TruD, therefore the two other ligated peptides were explored.

PatE-Large was ligated with CIS[Bpa]CAYDGE and treated with TruD overnight and then analysed by MALDI MS (Figure 7.20). The loss of 36 Da (corresponding to the loss of two waters) confirms that two thiazolines have been made from the cysteines and that the ligated peptide is an active substrate. Unfortunately, due to the Large Bpa amino acid in the 7th position of the core peptide, macrocyclisation of this sample would be challenging due to a steric clash with the PatGmac enzyme (See Section 3) and so this peptide was not taken further.

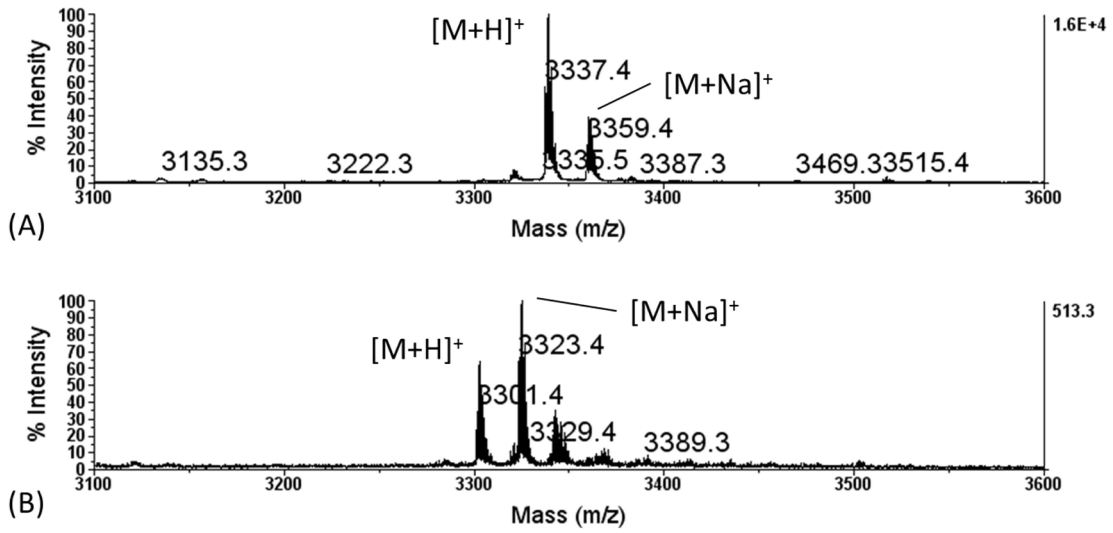


Figure 7.20: MALDI MS of PatE-Large-CIS[Bpa]CAYDGE and TruD. MALDI TOF MS of (A) the ligated peptide PatE-Large-CIS[Bpa]CAYDGE and (B) PatE-Large-CIS[Bpa]CAYDGE following TruD treatment. The loss of 36 Da corresponds to two water losses and the formation of two thiazoline heterocycles derived from the cysteine residues.

The final ligated peptide, PatE-GLEA-CKI[Thi]ACI[Hyp]APAYDG was subjected to TruD treatment overnight and a single water loss was observed in MALDI MS (Figure 7.21). The heterocyclised peptide was purified and cleaved with trypsin (Figure 7.22) and then incubated with PatGmac for one week to yield the final macrocyclic compound (Figure 7.23).

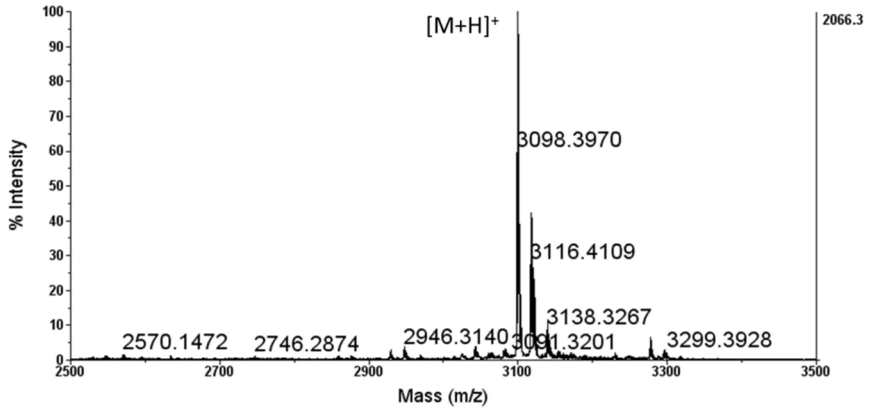


Figure 7.21: MALDI MS of PatE-GLEA-CKI[Thi]ACI[Hyp]APAYDG and TruD. MALDI TOF MS of the ligated peptide PatE-GLEA-CKI[Thi]ACI[Hyp]APAYDG following TruD treatment. The loss of 18 Da corresponds to one water loss and the formation of a single heterocycle.

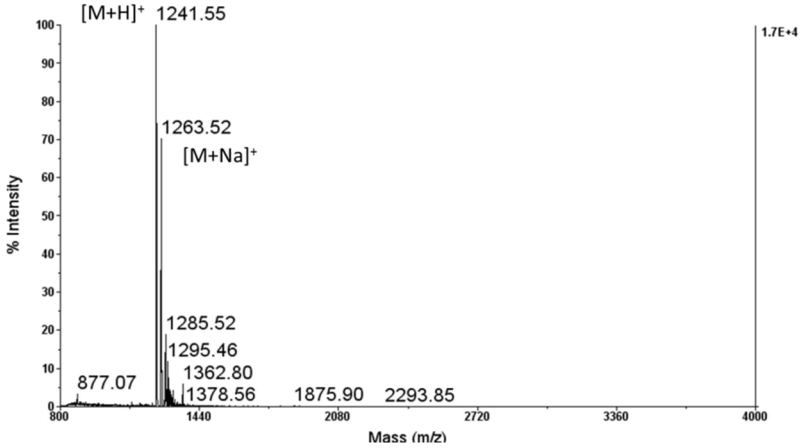


Figure 7.22: MALDI MS of Heterocyclised I[Thi]ACI[Hyp]APAYDG. MALDI TOF MS of heterocyclised PatE-GLEA-CKI[Thi]ACI[Hyp]APAYDG following trypsin treatment to remove the leader peptide. The mass corresponds only to the core peptide and C-terminal 'AYDG' recognition sequence.

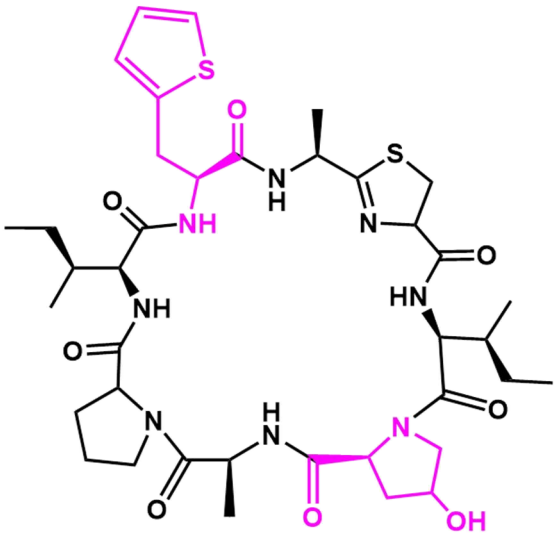


Figure 7.23: Chemical Structure of cyclo[I[Thi]AC^{Thn}I[Hyp]AP]. Expected chemical structure of cyclo[I[Thi]AC^{Thn}I[Hyp]AP] with the unnatural amino acids Thienylalanine and Hydroxproline highlighted in magenta.

7.2.2.7 Mass Fragmentation

Mass fragmentation analysis of cyclo[I[Thi]AC^{Thn}I[Hyp]AP] was carried out by Dr Sally Shirran (Figure 7.24). The fragments were assigned based on the hypothesised structure and were confirmed to correspond to the expected macrocycle.

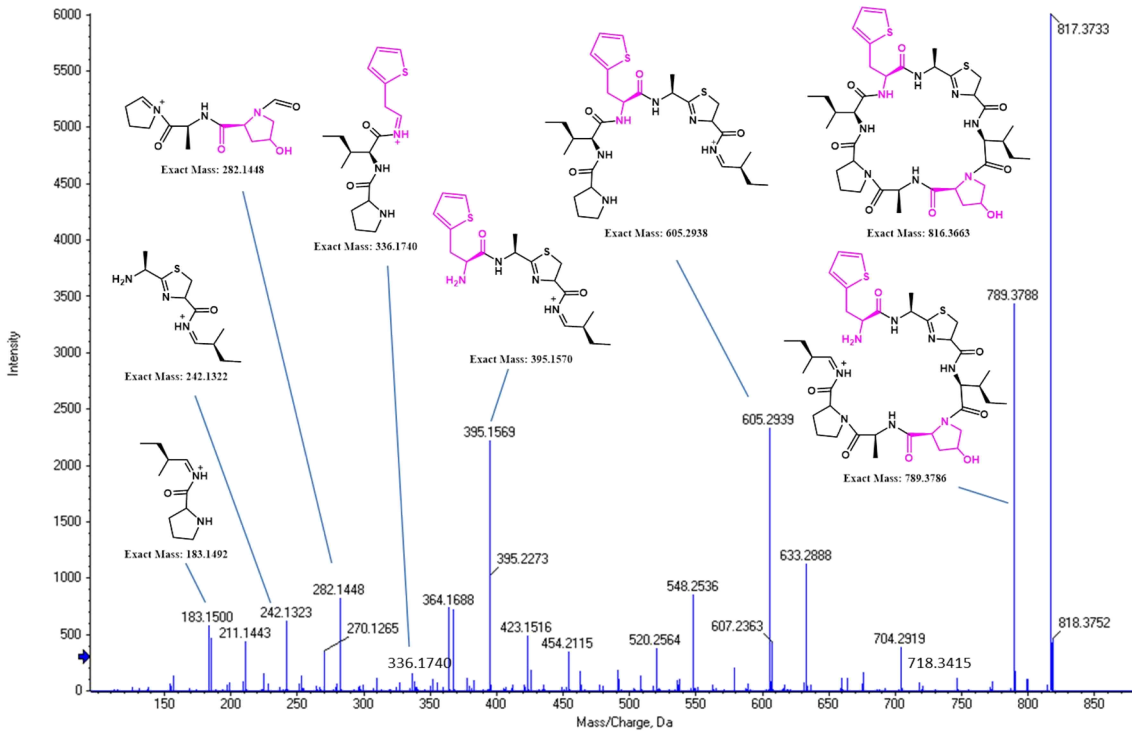


Figure 7.24: LC/MS/MS of cyclo[I[Thi]AC^{Thn}I[Hyp]AP]. LC/MS/MS spectrum of cyclo[I[Thi]AC^{Thn}I[Hyp]AP] with structures assigned to the primary peaks.

7.2.3 Discussion

Intein-mediated protein ligation offers a method of generating semi-synthetic libraries of peptides containing unnatural entities. Four PatE-Intein constructs have been developed using the Mxe GyrA intein (1txb, NEB) with varying lengths of PatE leader sequence. Two of these have been ruled out; small (leader residues 36-43) and full length. The small PatE-intein has limited heterocyclase processing and the full length precipitates upon thiol cleavage. The remaining two constructs PatE-Intein Large and PatE-Intein GLEA have leader sequences of size which have been shown previously to be processed efficiently by TruD [67] and it has been demonstrated that they can be ligated in a one-step reaction to chemically synthesised C-terminal peptides.

The heterologously expressed C-terminal peptide CKITAPITWPAYDGELEHHHHHH was used as a model substrate to allow the intein ligation step to be optimised. This was important as we could generate large quantities of the peptide at relatively low costs compared to chemically synthesised C-terminal peptides. The presence of the His_6-tag allows for efficient isolation and the incorporation of the tryptophan residue enables ligation efficiency to be monitored by $UV_{280nm.}$ We estimate that in our system the ligation efficiency was greater than 80 %.

The optimised one step ligation reaction was carried out using PatE-Intein-Large with the C-terminal peptide CIT[Bpa]CAYDG and PatE-Intein-GLEA with the C-terminal peptide CKI[Thi]ACI[Hyp]APAYDG and the resultant ligated peptides were shown to be fully active substrates for the heterocyclase enzyme TruD. The former peptide at this stage was abandoned as the Bpa amino acid in the 7th core peptide position doesn't macrocyclise due to the bulky nature of the side chain. The latter peptide was, however, taken through the rest of the *in vitro* process with cleavage by trypsin and finally macrocyclisation with PatGmac. The final macrocycle cyclo[I[Thi]AC^{Thn}I[Hyp]AP] was characterised by MSMS, confirming that the intein-mediated peptide ligation (IPL) method is amenable to creating unnatural amino acid containing macrocycles. This

opens up a wide range of chemical space with the only restriction in the core peptide being a C-terminal heterocycle or proline residue.

7.3 pEVOL technology

7.3.1 Materials and methods

7.3.1.1 Construct Design and DNA Cloning

A PatE precursor peptide was designed and cloned based upon the PatE2K sequence to generate a PatE with a core peptide of ITA[AmberStopCodon]ITAC using methods described in section 4.2.1 with the primers listed in Table 7.2 (Figure 7.25).

5' PatE	cttccatggacaaaaaaaacattcta
3' PatEAmberStop	cttctcgagttcgccatcatacgcgcacgcggtaatctacgcggtgatttta gatgcttccaaaccagc

Table 7.2: PatE Amber Stop Codon Primers: PCR primer sequences used to generate PatE containing amber stop in the core peptide sequence.

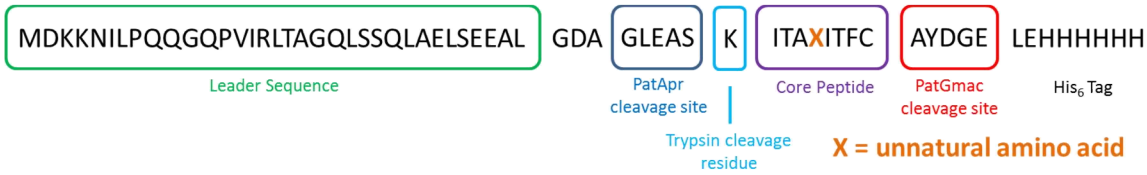


Figure 7.25: Schematic of PatE Amber Stop. Schematic of the PatE Amber Stop construct created with X representing the position where the unnatural amino acid is incorporated in response to the amber codon. This construct was derived from PatE2K (Section 4.2.1) with a lysine for efficient N-terminal cleavage and C-terminal His_6-tag for isolation.

The pEVOL-pBpF plasmid (expressing *p*-benzoyl-L-phenylalanyl-tRNA synthetase and tRNA pair), used to load the amino acid p-benzoyl-L-phenylalanine in response to the amber stop codon '*TAG*', was a gift from Peter Schultz (Scripps Research Institute, Addgene plasmid # 31190) [183].

7.3.1.2 Expression and Purification

The two plasmids PatE-Amber-Stop and pEVOL-pBpF were used to co-transform (selected using two antibiotics) BL21 (DE3) *E. coli* cells, which were then grown at 37 °C to an OD_{600nm} of 0.6 where they were then induced with 1 mM IPTG (PatE induction), 0.02% L-arabinose (pEVOL-pBpF induction) and the addition of 0.3 mM Bpa amino acid (0.3 M stock solubilised in 1 M NaOH). The culture was them grown overnight at 37 °C, 200 rpm before harvesting by centrifugation.

Cell pellets of PatE-Bpa were purified using the same refolding technique described previously for PatE2/PatE2K (Section 4.2.2). The protein was visualised by SDS-PAGE and Bpa incorporation was confirmed by MALDI-TOF-MS analysis.

7.3.1.3 *In Vitro* Processing

The PatE-Bpa peptide was processed *in vitro* using the same methods as outlined for PatE2K (Sections 4.2.4 - 4.2.6), with cysteine heterocyclisation by TruD, N-terminal cleavage with bovine trypsin and finally macrocyclisation with PatGmac.

7.3.1.4 Mass Fragmentation

The final macrocycle containing the Bpa amino acid was subjected to LC/MS/MS using an AB SCIEX Triple TOF 5600 instrument and the fragment peaks were assigned confirming the structure. This work was carried out by Dr. Sally Shirran (University of St Andrews).

7.3.2 Results

7.3.2.1 Expression and Purification

PatE2K(ITA[Bpa]ITAC) was expressed from the pBMSCHis plasmid in BL21 DE3 with the p-benzoyl-l-phenylalanine (Bpa) amino acid loaded using the pEVOL-pBpF tRNA synthase/tRNA pair plasmid. The peptide was solubilised from inclusion bodies in denaturing buffers, isolated by Ni-NTA chromatography and refolded using DTT. The peptide elutes off gel filtration as a single peak with a final yield of 10 - 15 mg L^{-1} culture (Figure 7.26 A, B). The peptide was analysed by MALDI-TOF-MS confirming 100 % incorporation of the Bpa amino acid (Figure 7.26 C).

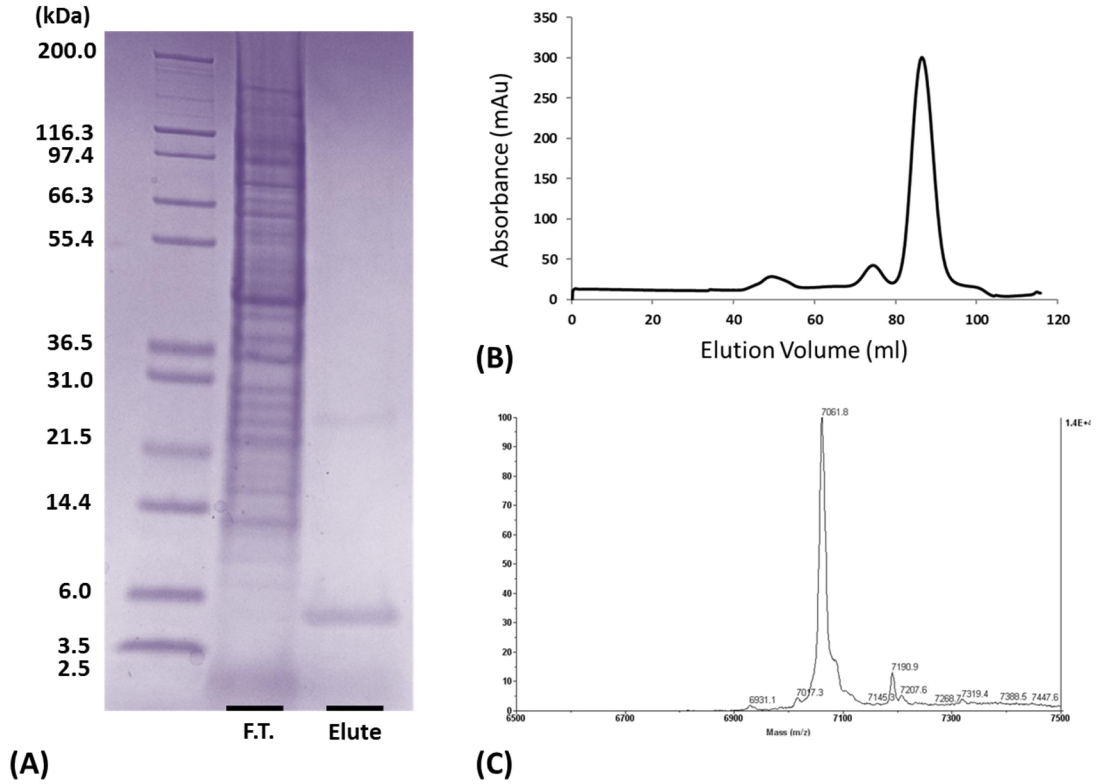


Figure 7.26: Purification of Bpa-containing PatE. (A) SDS-PAGE analysis of Ni Elution showing F.T (with no PatE present) and elution with PatE. (B) Gel filtration UV spectra of PatE-Bpa (C) MALDI TOF MS of final purified PatE-Bpa showing full incorporation of the Bpa amino acid into the peptide.

7.3.2.2 *In vitro* Processing

The purified PatE was treated with TruD in order to heterocyclise the cysteine of the core peptide as described previously (Section 4.2.4) and then purified from the enzyme. MALDI-MS confirmed the majority (> 90 %) was heterocyclised (Figure 7.27). The sample was then cleaved by trypsin and purified on a Superdex S30 column.

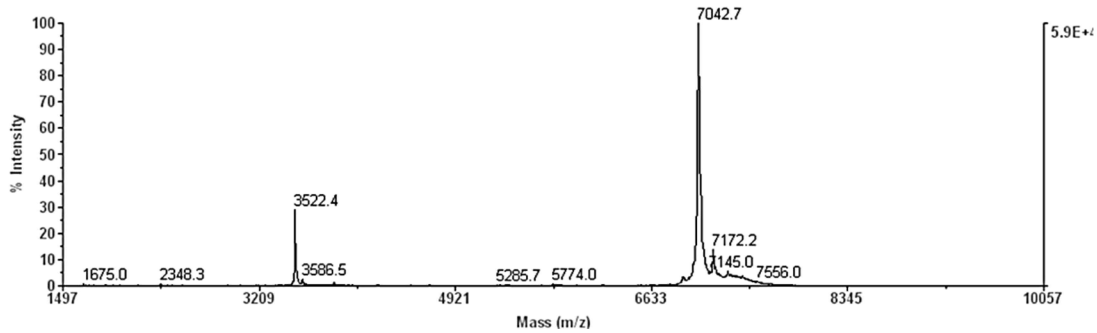


Figure 7.27: MS of TruD-treated PatE-ITA[Bpa]ITAC. MS spectra for PatE-ITA[Bpa]ITAC post heterocyclisation by TruD. A mass loss of 18 Da [7043 $M+H^+$] from starting material confirms the formation of a thiazoline from the single cysteine residue.

The sample was then cleaved by trypsin, purified on a Superdex S30 column and then incubated for one week with PatGmac in order to form the corresponding macrocycle (Figures 7.28, 7.29).

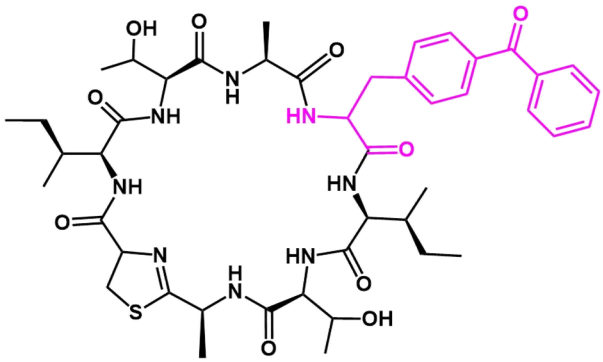


Figure 7.28: Chemical Structure of cyclo[ITA[Bpa]ITAC^{Thn}]. Expected chemical structure of cyclo[ITA[Bpa]ITAC^{Thn}] with the unnatural amino acid Benzophenylalanine highlighted in magenta.

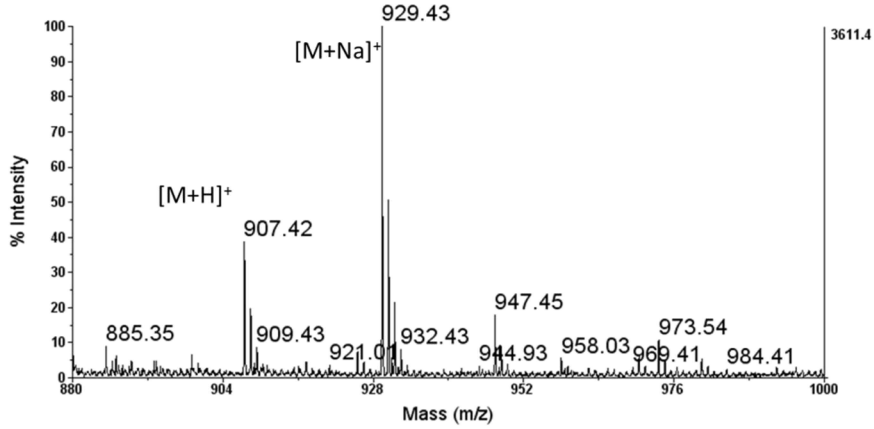


Figure 7.29 – MALDI TOF MS of cyclo[ITA[Bpa]ITAC^{Thn}]. MALDI-TOF-MS of cyclo[ITA[Bpa]ITAC^{Thn}] with peaks of $[M+H]^+$ at 907.4, $[M+Na]^+$ at 929.4 and $[M+K]^+$ = 947.4.

7.3.2.3 Mass Fragmentation

Mass fragmentation analysis of cyclo[ITA[Bpa]ITAC^{Thn}] was carried out by Dr Sally Shirran (Figure 7.30). The fragments were assigned based on the hypothesised structure and were confirmed to correspond to the expected macrocycle.

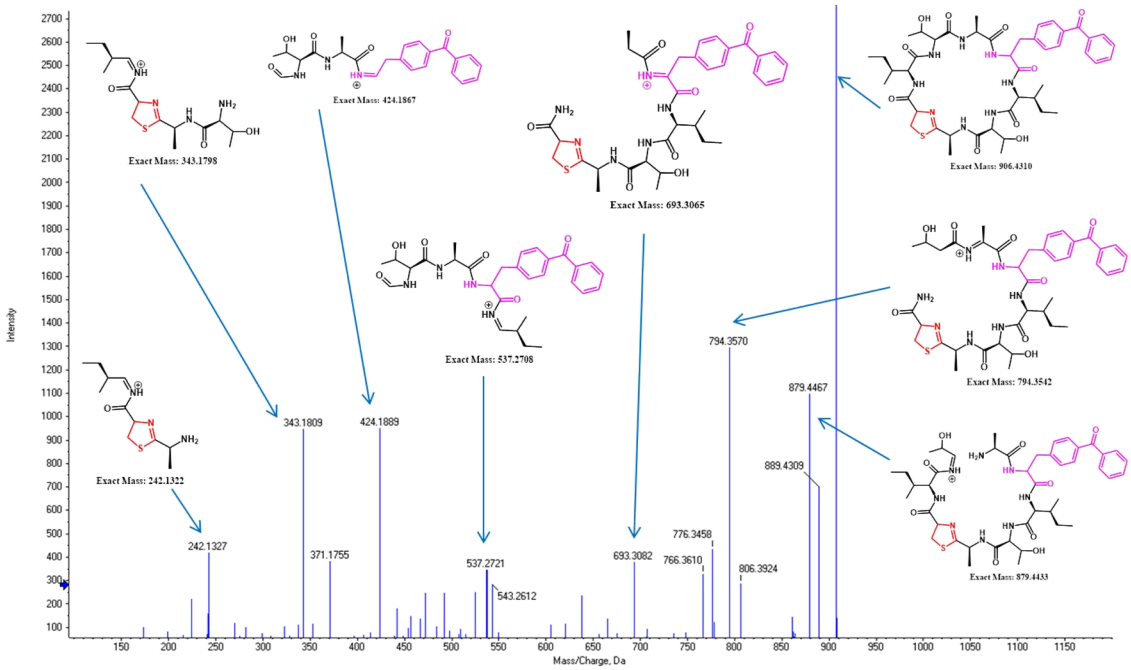


Figure 7.30: LC/MS/MS Analysis of Bpa-containing macrocycle. LC/MS/MS spectrum of cyclo[ITA[Bpa]ITAC^{Thn}] with peaks assigned based on the chemical structure.

7.3.3 Discussion

It has been demonstrated that macrocyclic peptides containing an unnatural amino acid can be derived *in vitro* from precursor peptides created *in vivo* using the pEVOL amber stop codon system which directly incorporates the unnatural amino acid on to the precursor peptide.

The co-expression of the pEVOL and pBMS plasmids containing aminoacyl-tRNA synthase/suppressor tRNA pairs and PatE-AmberStop genes, respectively, efficiently allows the incorporation *in vivo* of the pBpa amino acid into the core peptide of PatE. The resulting PatE variant was then purified and processed *in vitro* using the same enzymes and processes as for native PatE peptides (Chapter 4). This approach offers a very flexible method of producing cyclic peptides with unnatural amino acids as only the initial *in vivo* expression of the PatE is modified. Alternate unnatural amino acids could be used by exchanging the pEVOL co-expression plasmid enabling a range of amino acids to be incorporated into PatE proteins.

The main disadvantage of this technique is that the expression levels with pBpa are between 10 and 20 % of a native PatE and as such expression conditions may have to be further optimised to improve yields.

7.4 Conclusions and Future Work

Two methods have been presented which allow the generation of precursor peptides containing unnatural amino acids, which are further processed *in vitro* to yield the corresponding macrocycles. Each of the methods has its own advantages and disadvantages. The intein method offers wide diversity in the final product as it can incorporate several different unnatural amino acids per macrocycle. However, it requires the synthetic C-terminal peptide to be soluble in aqueous buffer (up to 10 % DMSO can be tolerated). The pEVOL system allows a straight substitution for the standard PatE expression system. However, unlike the intein approach, only a single unnatural amino acid type can be incorporated per PatE (although, it can be incorporated at multiple sites) and at present only a small selection of unnatural amino acid tRNA/synthase pairs are commercially available. It should be possible to use both techniques in a complementary fashion to create an array of macrocycles containing a wide range of unnatural amino acids.

The ability to make unnatural amino acid-containing macrocycles has been demonstrated by MSMS analyses. To further confirm the structures of the products, we aim to scale up to levels which will allow NMR spectroscopic analysis to be run on the final products and their structures to be fully assigned.

Since the development of these protocols, the Naismith laboratory has developed the heterocyclase LynD (a homologue of PatD capable of heterocyclising cysteine residues only) with the PatE leader fused directly to its N-terminus [78]. This LynD fusion protein heterocyclises chemically synthesised peptides efficiently without the requirement of the leader sequence. While this new enzyme may reduce the requirement for the intein and pEVOL approaches, it is always important to have multiple approaches to a problem and the experience of using these techniques developed in this project could facilitate their use in other applications.

8. Bibliography

[1] F. P. Tally and M. F. DeBruin, "Development of daptomycin for Gram-positive infections," *J. Antimicrob. Chemother.,* vol. 46, no. 4, pp. 523-526, 2000.

[2] L.-J. Ball, C. M. Goult, J. A. Donarski, J. Micklefield and V. Ramesh, "NMR structure determination and calcium binding effects of lipopeptide antibiotic daptomycin," *Org. Biomol. Chem.,* vol. 2, pp. 1872-1878, 2004.

[3] A. R. Hamel, F. Huber, A. Carrupt, R. M. Wenger and M. Mutter, "Cyclosporin A prodrugs: design, synthesis and biophysical properties," *J. Peptide Res.,* vol. 63, pp. 147-154, 2004.

[4] L. C. Sun and D. H. Coy, "Somatostatin receptor-targeted anti-cancer therapy," *Curr. Drug Deliv.,* vol. 8, pp. 2-10, 2011.

[5] H. Ma, B. Zhou, Y. Kim and K. D. Janda, "A cyclic peptide-polymer probe for the detection of Clostridium botulinum neurotoxin serotype A," *Toxicon,* vol. 47, pp. 901-8, 2006.

[6] P. Wipf, "Synthetic Studies of Biologically Active Marine Cyclopeptides," *Chem. Rev.,* vol. 95, pp. 2115-2134, 1995.

[7] T. Ishida, H. Ohishi, M. Inoue, M. Kamigauchi, M. Sugiura, N. Takao, S. Kato, Y. Hamada and T. Shioiri, "Conformational properties of ulithiacyclamide, a strongly cytotoxic cyclic peptide from a marine tunicate, determined by 1H nuclear magnetic resonance and energy minimization calculations," *J. Org. Chem.,* vol. 54, pp. 5337-5343, 1989.

[8] C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Adv. Drug Deliv. Rev.,* vol. 46, pp. 3-26, 2001.

[9] A. T. Bockus, C. M. McEwen and R. S. Lokey, "Form and function in cyclic Peptide natural products: a pharmacokinetic perspective," *Curr. Top. Med. Chem.,* vol. 13, pp. 821-836, 2013.

[10] A. Ehrlich, H.-U. Heyne, R. Winter, M. Beyermann, H. Haber, L. A. Carpino and M. Bienert, "Cyclization of all-L-Pentapeptides by Means of 1-Hydroxy-7-azabenzotriazole-Derived Uronium and Phosphonium Reagents," *J. Org. Chem.,* vol. 61, pp. 8831-8838, 1996.

[11] L. Zhang and J. P. Tam, "Synthesis and Application of Unprotected Cyclic Peptides as Building Blocks for Peptide Dendrimers," *J. Am. Chem. Soc.,* vol. 119, pp. 2363-2370, 1997.

[12] C. T. Wong, D. K. Rowlands, C.-H. Wong, T. W. C. Lo, G. K. T. Nguyen, H.-Y. Li and J. P. Tam, "Orally active peptidic bradykinin B1 receptor antagonists," *Angew. Chem. Int. Ed.,* vol. 51, pp. 5620-5624, 2012.

[13] P. Hoogerhout, W. Kamphuis, H. F. Brugghe, J. A. Sluijs, H. A. Timmermans, J.

Westdijk, G. Zomer, C. J. Boog, E. M. Hol and G. P. van den Dobbelsteen, "A cyclic undecamer peptide mimics a turn in folded Alzheimer amyloid β and elicits antibodies against oligomeric and fibrillar amyloid and plaques," *PLoS One,* vol. 6, p. e19110, 2011.

[14] G. F. Gause and M. G. Brazhnikova, "Gramicidin S and its use in the treatment of infected wounds," *Nature,* vol. 154, p. 703, 1944.

[15] J. W. Blunt, B. R. Copp, R. A. Keyzers and M. H. G. Munro, "Marine natural products," *Nat. Prod. Rep.,* vol. 29, pp. 144-222, 2012.

[16] P. A. Arnison, M. J. Bibb, G. Bierbaum, A. A. Bowers, T. S. Bugni, G. Bulaj, J. A. Camarero, D. J. Campopiano, G. L. Challis, J. Clardy, P. D. Cotter, D. J. Craik, M. Dawson, E. Dittmann, S. Donadio, P. C. Dorrestein, K.-D. Entian, M. A. Fischbach, J. S. Garavelli, U. Goransson, C. W. Gruber, D. H. Haft, T. K. Hemscheidt, C. Hertweck, C. Hill, A. R. Horswill, M. Jaspars, W. L. Kelly, J. P. Klinman, O. P. Kuipers, A. J. Link, W. Liu, M. A. Marahiel, D. A. Mitchell, G. N. Moll, B. S. Moore, R. Muller, S. K. Nair, I. F. Nes, G. E. Norris, B. M. Olivera, H. Onaka, M. L. Patchett, J. Piel, M. J. T. Reaney, S. Rebuffat, R. P. Ross, H.-G. Sahl, E. W. Schmidt, M. E. Selsted, K. Severinov, B. Shen, K. Sivonen, L. Smith, T. Stein, R. D. Sussmuth, J. R. Tagg, G.-L. Tang, A. W. Truman, J. C. Vederas, C. T. Walsh, J. D. Walton, S. C. Wenzel, J. M. Willey and W. A. van der Donk, "Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature," *Nat. Prod. Rep.,* vol. 30, pp. 108-160, 2013.

[17] P. J. Knerr and W. A. van der Donk, "Discovery, biosynthesis, and engineering of lantipeptides," *Annu. Rev. Biochem.,* vol. 81, pp. 479-505, 2012.

[18] J. M. Waisvisz, M. G. van der Hoeven, J. van Peppen and W. C. M. Zwennis, "Bottromycin. I. a new sulfur-containing antibiotic," *J. Am. Chem. Soc.,* vol. 79, pp. 4520-4521, 1957.

[19] S. Duquesne, D. Destoumieux-Garzon, J. Peduzzi and S. Rebuffat, "Microcins, gene-encoded antibacterial peptides from enterobacteria," *Nat. Prod. Rep.,* vol. 24, pp. 708-34, 2007.

[20] J. O. Melby, M. J. Nard and D. A. Mitchell, "Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates," *Curr. Opin. Chem. Biol.,* vol. 15, pp. 369-378, 2011.

[21] J. D. Walton, H. E. Hallen-Adams and H. Luo, "Ribosomal biosynthesis of the cyclic peptide toxins of Amanita mushrooms," *Biopolymers,* vol. 94, pp. 659-664, 2010.

[22] K. Sivonen, N. Leikoski, D. P. Fewer and J. Jokela, "Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria," *Appl. Microbiol. Biotechnol.,* vol. 86, pp. 1213-1225, 2010.

[23] D. J. Craik, N. L. Daly, T. Bond and C. Waine, "Plant cyclotides: A unique family of cyclic and knotted proteins that defines the cyclic cystine knot structural motif," *J. Mol. Biol.,* vol. 294, pp. 1327-1336, 1999.

[24] J. Lubelski, R. Rink, R. Khusainov, G. N. Moll and O. P. Kuipers, "Biosynthesis,

immunity, regulation, mode of action and engineering of the model lantibiotic nisin," *Cell. Mol. Life Sci.,* vol. 65, pp. 455-476, 2008.

[25] K. R. Gustafson, T. C. McKee and H. R. Bokesch, "Anti-HIV cyclotides," *Curr. Protein. Pep. Sci.,* vol. 5, pp. 331-340, 2004.

[26] B. Kaylon, S. E. Helaly, R. Scholz, J. Nachtigall, J. Vater, R. Borriss and R. D. Sussmuth, "Plantazolicin A and B: structure elucidation of ribosomally synthesized thiazole/oxazole peptides from Bacillus amyloliquefaciens FZB42," *Org. Lett.,* vol. 13, pp. 2996-2999, 2011.

[27] "Structure determination and interception of biosynthetic intermediates for the Plantazolicin class of highly discriminating antibiotics," *ACS Chem. Biol.,* vol. 6, pp. 1307-1313, 2011.

[28] H. Shimamura, H. Gouda, K. Nagai, T. Hirose, M. Ichioka, Y. Furuya, Y. Kobayashi, S. Hirono, T. Sunazuka and S. Omura, "Structure determination and total synthesis of Bottromycin A2: a potent antibiotic against MRSA and VRE," *Angew. Chem. Int. Ed.,* vol. 48, pp. 914-917, 2009.

[29] T. Wieland, H. Faulstich and L. Fiume, "Amatoxins, Phallotoxins, Phallolysin, and Antamanide: The biologically active components of poisonous Amanita mushroom".

[30] M. A. Ortega, Y. Hao, Q. Zhang, M. C. Walker, W. A. van der Donk and S. K. Nair, "Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB," *Nature,* vol. 517, pp. 509-512, 2015.

[31] A. Karakas-Sen, A. Narbad, N. Horn, H. M. Dodd, A. J. Parr, I. Colquhoun and M. J. Gasson, "Post-translational modification of nisin. The involvement of NisB in the dehydration process," *Eur. J. Biochem.,* vol. 261, pp. 524-532, 1999.

[32] N. Garg, L. M. Salazar-Ocampo and W. A. van der Donk, "In vitro activity of the nisin dehydratase NisB," *Proc. Natl. Acad. Sci. USA,* vol. 110, pp. 7258-7263, 2013.

[33] Schrodinger, LLC, "The PyMOL Molecular Graphics System," Version 1.5.0.4.

[34] N. Leikoski, D. P. Fewer, J. Jokela, M. Wahlsten, L. Rouhiainen and K. Sivonen, "Highly Diverse Cyanobactins in Strains of the Genus Anabaena," *Appl. Environ. Microbiol.,* vol. 76, pp. 701-709, 2010.

[35] M. S. Donia and E. W. Schmidt, "Linking chemistry and genetics in the growning cyanobactin natural products family," *Chem. Biol.,* vol. 18, pp. 508-519, 2011.

[36] R. Banker and S. Carmeli, "Tenuecyclamides A-D, cyclic hexapeptides from the cyanobacterium Nostoc spongiaeforme var. tenue," *J. Nat. Prod.,* vol. 61, pp. 1248-51, 1998.

[37] S. Sudek, M. G. Haygood, D. T. Youssef and E. W. Schmidt, "Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium Trichodesmium erythraeum, predicted from the genome sequence," *Appl. Environ. Microbiol.,* vol. 72, pp. 4382-4387, 2006.

[38] A. R. Carroll, J. C. Coll, D. J. Bourne, J. K. Macleod, T. M. Zabriskie, C. M. Ireland and B. F. Bowden, "Patellins 1-6 and Trunkamide A: Novel Cyclic Hexa-, Hepta-

and Octa-peptides From Colonial Ascidians, Lissoclinum sp.," *Aus. J. Chem.,* vol. 49, pp. 659-667, 1996.

[39] C. Ireland and P. J. Scheuer, "Ulicyclamide and ulithiacyclamide, two new small peptides from a marine tunicate," *J. Am. Chem. Soc.,* vol. 102, pp. 5688-5691, 1980.

[40] C. M. Ireland, A. R. Durso Jr, R. A. Newman and M. P. Hacker, "Antineoplastic cyclic peptides from the marine tunicate lissoclinum patella," *J. Org. Chem.,* vol. 47, pp. 1807-1811, 1982.

[41] K. Kohda, Y. Ohta, Y. Yokoyama, Y. Kawazoe, T. Kato, Y. Suzumura, Y. Hamada and T. Shioiri, "Mechanistic aspects of the cytocidal action of ulithiacyclamide on mouse leukemia L1210 cells in vitro," *Biochem. Pharmacol.,* vol. 38, pp. 4497-4500, 1989.

[42] B. M. Degnan, C. J. Hawkins, M. F. Lavin, E. J. McCaffrey, D. L. Parry, A. L. van den Brenk and D. J. Watters, "New cyclic peptides with cytotoxic activity from the ascidian lissoclinum patella," *J. Med. Chem.,* vol. 32, pp. 1349-1354, 1989.

[43] X. Salvatella, J. M. Caba, F. Albericio and E. Giralt, "Solution structure of the antitumor candidate trunkamide A by 2D NMR and restrained simulated annealing methods," *J. Org. Chem.,* vol. 24, pp. 211-215, 2003.

[44] R. G. Linington, J. Gonzalez, L. D. Urena, L. I. Romero, E. Ortega-Barria and W. H. Gerwick, "Venturamides A and B: antimalarial constituents of the panamanian marine Cyanobacterium Oscillatoria sp.," *J. Nat. Prod,* vol. 70, pp. 397-401, 2007.

[45] J. Ogino, R. E. Moore, G. M. Patterson and C. D. Smith, "Dendroamides, new cyclic hexapeptides from a blue-green alga. Multidrug-resistance reversing activity of dendroamide A," *J. Nat. Prod.,* vol. 59, pp. 581-586, 1996.

[46] S. F. Brady, L. Simmons, J. H. Kim and E. W. Schmidt, "Metagenomic approaches to natural products from free-living and symbiotic organisms," *Nat. Prod. Rep.,* vol. 26, pp. 1488-1503, 2009.

[47] E. W. Schmidt, J. T. Nelson, D. A. Rasko, S. Sudek, J. A. Eisen, M. G. Haygood and J. Ravel, "Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of lissoclinum patella," *Proc. Natl. Acad. Sci. USA,* vol. 17, pp. 7315-7320, 2005.

[48] T. S. Bibby, J. Nield, M. Chen, A. W. D. Larkum and J. Barber, "Structure of a photosystem II supercomplex isolated from Prochloron didemni retaining its chlorophyll a/b light-harvesting system," *Proc. Natl. Acad. Sci. USA,* vol. 100, pp. 9050-9054, 2003.

[49] W. E. Houssen and M. Jaspars, "Azole-Based Cyclic Peptides from the Sea Squirt Lissoclinum Patella: Old Scaffolds, New Avenues," *ChemBioChem,* vol. 11, pp. 1803-1815, 2010.

[50] F. J. Schmitz, M. B. Ksebati, C. J. L. Wang, M. B. Hossain and D. van der Helm, "Cyclic peptides from the ascidian Lissoclinum patella: conformational analysis of patellamide D by X-ray analysis and molecular modeling," *J. Org. Chem.,* vol. 54, pp. 3463-3472, 1989.

[51] L. A. McDonald and C. M. Ireland, "Patellamide E: a New Cyclic Peptide from the

Ascidian Lissoclinum patella," *J. Nat. Prod.,* vol. 55, pp. 376-379, 1992.

[52] M. A. Rashid, K. R. Gustafson, J. H. Cardellina II and M. R. Boyd, "Patellamide F, a New Cytotoxic Cyclic Peptide from the Colonial Ascidian Lissoclinum patella," *J. Nat. Prod.,* vol. 58, pp. 594-597, 1995.

[53] X. Fu, T. Do, F. J. Schmitz, V. Andrusevich and M. H. Engel, "New Cyclic Peptides from the Ascidian Lissoclinum patella," *J. Nat. Prod.,* vol. 61, pp. 1547-1551, 1998.

[54] C. J. Hawkins, M. F. Lavin, K. A. Marshall, A. L. van den Brenk and D. J. Watters, "Structure-activity relationships of the lissoclinamides: cytotoxic cyclic peptides from the ascidian Lissoclinum patella," *J. Med. Chem.,* vol. 33, pp. 1634-1638, 1990.

[55] A. B. Williams and R. S. Jacobs, "A marine natural product, patellamide D, reverses multidrug resistance in a human leukemic cell line," *Cancer Lett.,* vol. 71, pp. 97-102, 1993.

[56] G. Szakacs, J. K. Paterson, J. A. Ludwig, C. Booth-Genthe and M. M. Gottesman, "Targeting multidrug resistance in cancer," *Nat. Rev. Drug Discov.,* vol. 5, pp. 219-234, 2006.

[57] S. G. Aller, J. Yu, A. Ward, Y. Weng, S. Chittaboina, R. Zhuo, P. M. Harrell, Y. T. Trinh, Q. Zhang, I. L. Urbatsch and G. Chang, "Structure of P-glycoprotein Reveals a Molecular Basis for Poly-Specific Drug Binding," *Science,* vol. 323, pp. 1718-1722, 2009.

[58] J. Li, K. F. Jaimes and S. G. Aller, "Refined structures of mouse P-glycoprotein," *Protein Sci.,* vol. 23, pp. 34-36, 2014.

[59] P. Garcia-Reynaga and M. S. VanNieuwenhze, "A new total synthesis of patellamide A," *Organic Letters,* vol. 10, pp. 4621-4623, 2008.

[60] S.-L. You, H. Razavi and J. W. Kelly, "A Biomimetic Synthesis of Thiazolines Using Hexaphenyloxodiphosphonium Trifluoromethanesulfonate," *Angew. Chem. Int. Ed.,* vol. 42, pp. 83-85, 2003.

[61] M. S. Donia, B. J. Hathaway, S. Sudek, M. G. Haygood, M. J. Rosovitz, J. Ravel and E. W. Schmidt, "Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians," *Nat. Chem. Biol.,* vol. 2, pp. 729-735, 2006.

[62] M. S. Donia, J. Ravel and E. W. Schmidt, "A global assembly line for cyanobactins," *Nat. Chem. Biol.,* vol. 4, pp. 341-343, 2008.

[63] J. Lee, J. McIntosh, B. J. Hathaway and E. W. Schmidt, "Using Marine Natural Products to Discover a Protease that Catalyzes Peptide Macrocyclization of Diverse Substrates," *J. Am. Chem. Soc.,* vol. 131, pp. 2122-2124, 2009.

[64] J. A. McIntosh, M. S. Donia and E. W. Schmidt, "Insights into Heterocyclization from Two Highly Similar Enzymes," *J. Am. Chem. Soc.,* vol. 132, pp. 4089-4091, 2010.

[65] J. A. McIntosh, M. S. Donia, S. K. Nair and E. W. Schmidt, "Enzymatic Basis of Ribosomal Peptide Prenylation in Cyanobacteria," *J. Am. Chem. Soc.,* vol. 133, pp. 13698-13705, 2011.

[66] V. Agarwal, E. Pierce, J. McIntosh, E. W. Schmidt and S. K. Nair, "Structures of cyanobactin maturation enzymes define a family of transamidating proteases," *Chem. Biol.,* vol. 19, pp. 1411-1422, 2012.

[67] J. Koehnke, A. F. Bent, D. Zollman, K. Smith, W. E. Houssen, X. Zhu, G. Mann, T. Lebl, R. Scharff, S. Shirran, C. H. Botting, M. Jaspars, U. Schwarz-Linek and J. H. Naismith, "The cyanobactin heterocyclase enzyme: a processive adenylase that operates with a defined order of reaction," *Angewandte Chemie,* vol. 52, pp. 13991-13996, 2013.

[68] Y. Goto, Y. Ito, Y. Kato, T. Tsunoda and H. Suga, "One-pot synthesis of azoline-containing peptides in a cell-free translation system integrated with a posttranslational cyclodehydratase," *Chemistry & Biology,* vol. 21, pp. 766-744, 2014.

[69] W. E. Houssen, A. F. Bent, A. R. McEwan, N. Pieiller, J. Tabudravu, J. Koehnke, G. Mann, R. I. Adaba, L. Thomas, U. W. Hawas, H. Liu, U. Schawrz-Linek, M. C. Smith, J. H. Naismith and M. Jaspars, "An efficient method for the in vitro production of azol(in)e-based cyclic peptides," *Angewandte Chemie,* vol. 126, pp. 14395-14398, 2014.

[70] C. A. Regni, R. F. Roush, D. J. Miller, A. Nourse, C. T. Walsh and B. A. Schulman, "How the MccB bacterial ancestor of ubiquitin E1 initiates biosynthesis of the microcin C7 antibiotic," *EMBO,* vol. 28, pp. 1953-1964, 2009.

[71] J. Wang and Y. Chen, "Role of the Zn2+ motif of E1 in SUMO adenylation," *J. Biol. Chem.,* vol. 285, pp. 23732-23738, 2010.

[72] M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, "Clustal W and Clustal X version 2.0," *Bioinformatics,* vol. 23, pp. 2947-2948, 2007.

[73] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWililam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins, "Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega," *Molecular Systems Biology,* vol. 7, p. 539, 2011.

[74] C. S. Bond and A. W. Schuttelkopf, "ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments," *Acta Cryst.,* vol. D65, pp. 510-512, 2009.

[75] W. E. Houssen, S. H. Wright, A. P. Kalverda, G. S. Thompson, S. M. Kelly and M. Jaspars, "Solution structure of the leader sequence of the patellamide precursor peptide, PatE1-34," *ChemBioChem,* vol. 11, pp. 1867-1873, 2010.

[76] R. S. Roy, S. Kim, J. D. Beleja and C. T. Walsh, "Role of the microcin B17 propeptide in substrate recognition: solution structure and mutational analysis of McbA1−26," *Chemistry & Biology,* vol. 5, pp. 217-228, 1998.

[77] Y.-M. Li, J. C. Milne, L. L. Madison, R. Kolter and C. T. Walsh, "From Peptide Precursors to Oxazole and Thiazole-Containing Peptide Antibiotics: Microcin B17 Synthase," *Science,* vol. 274, pp. 1188-1193, 1996.

[78] J. Koehnke, G. Mann, A. F. Bent, H. Ludewig, S. Shirran, C. Botting, T. Lebl, W. E.

Houssen, M. Jaspars and J. H. Naismith, "Structural analysis of leader peptide binding enables leader-free cyanobactin processing," *Nat. Chem. Biol.,* vol. 11, pp. 558-563, 2015.

[79] J. A. McIntosh and E. W. Schmidt, "Marine molecular machines: heterocyclization in cyanobactin biosynthesis," *ChemBioChem,* vol. 11, pp. 1413-1421, 2010.

[80] K. L. Dunbar, J. O. Melby and D. A. Mitchell, "YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations," *Nat. Chem. Biol.,* vol. 8, pp. 569-575, 2012.

[81] K. L. Dunbar and D. A. Mitchell, "Insights into the Mechanism of Peptide Cyclodehydrations Achieved," *J. Am. Chem. Soc.,* vol. 135, pp. 8692-8701, 2013.

[82] K. L. Dunbar, J. R. Chekan, C. L. Cox, B. J. Burkhart, S. K. Nair and D. A. Mitchell, "Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis," *Nat. Chem. Biol.,* vol. 10, pp. 823-829, 2014.

[83] K. L. Dunbar, J. I. Tietz, C. L. Cox, B. J. Burkhart and D. A. Mitchell, "Identification of an Auxiliary Leader Peptide-Binding Protein," *J. Am. Chem. Soc.,* vol. 137, pp. 7672-7677, 2015.

[84] R. J. Siezen, B. Renckens and J. Boekhorst, "Evolution of prokaryotic subtilases: Genome-wide analysis reveals novel subfamilies with different catalytic residues," *Proteins: Structure, Function & Bioinformatics,* vol. 67, pp. 681-694, 2007.

[85] G. Dodson and A. Wlodawer, "Catalytic triads and their relatives," *Trends in Biochem. Sci.,* vol. 23, pp. 347-352, 1998.

[86] J. M. Berg, J. L. Tymoczko and L. Stryer, Biochemistry, 5th ed., New York: W.H. Freeman and Co., 2002.

[87] W. E. Houssen, J. Koehnke, D. Zollman, J. Vendome, A. Raab, M. C. Smith, J. H. Naismith and M. Jaspars, "The discovery of new cyanobactins from Cyanothece PCC 7425 defines a new signature for processing of patellamides," *ChemBioChem,* vol. 21, pp. 2683-2689, 2012.

[88] J. A. McIntosh, C. R. Roberston, V. Agarwal, S. K. Nair, G. W. Bulaj and E. W. Schmidt, "Circular Logic: Nonribosomal Peptide-like Macrocyclization with a Ribosomal Peptide Catalyst," *J. Am. Chem. Soc.,* vol. 132, pp. 15499-15501, 2010.

[89] J. W. Trauger, R. M. Kohli, H. D. Mootz, M. A. Marahiel and C. T. Walsh, "Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase," *Nature,* vol. 407, pp. 215-218, 2000.

[90] D. E. Cane and C. T. Walsh, "The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases," *Chemistry & Biology,* vol. 6, pp. R319-R325, 1999.

[91] C. J. S. Barber, P. T. Pujara, D. W. Reed, S. Chiwocha, H. Zhang and P. S. Covello, "The two-step biosynthesis of cyclic peptides from linear precursors in a member of the plant family caryophyllaceae involves cyclisation by a serine protease-like enzyme," *J. Biol. Chem.,* vol. 288, pp. 12500-12510, 2013.

[92] H. Luo, S.-Y. Hong, R. M. Sgambelluri, E. Angelos, X. Li and J. D. Walton, "Peptide

macrocyclization catalyzed by a prolyl oligopeptidase involved in α-amanitin biosynthesis," *Chemistry & Biology,* vol. 21, pp. 1610-1617, 2014.

[93] G. K. T. Nguyen, S. Wang, Y. Qiu, X. Hemu, Y. Lian and J. P. Tam, "Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis," *Nat. Chem. Biol.,* vol. 10, pp. 732-738, 2014.

[94] B. F. Milne, P. F. Long, A. Starcevic, D. Hranueli and M. Jaspars, "Spontaneity in the patellamide biosynthetic pathway," *Org. Biomol. Chem.,* vol. 4, pp. 631-638, 2006.

[95] G. Mann, J. Koehnke, A. F. Bent, R. Graham, W. E. Houssen, M. Jaspars and J. H. Naismith, "Structural studies of the cyanobactin DUF domains," *Acta. Cryst.,* vol. F70, pp. 1597-1603, 2014.

[96] J. O. Melby, K. L. Dunbar, N. Q. Trinh and D. A. Mitchell, "Selectivity, Directionality and Promiscuity in Peptide Processing from a Bacillus sp. Al Hakam Cyclodehydratase," *J. Am. Chem. Soc.,* vol. 134, pp. 5309-5316, 2012.

[97] T. L. Schneider, B. Shen and C. T. Walsh, "Oxidase Domains in Epothilone and Bleomycin Biosynthesis: Thiazoline to Thiazole Oxidation during Chain Elongation," *Biochemistry,* vol. 42, pp. 9722-9730, 2003.

[98] P. J. Casey and M. C. Seabra, "Protein prenyltransferases," *J. Biol. Chem.,* vol. 271, pp. 5289-5292, 1996.

[99] N. Leikoski, D. P. Fewer, J. Jokela, P. Alakoski, M. Wahlsten and K. Sivonen, "Analysis of an inactive cyanobactin biosynthetic gene cluster leads to discovery of new natural products from strains of the genus Microcystis," *PLoS One,* vol. 7, pp. 1-9, 2012.

[100] F. L. Zhang and P. J. Casey, "Protein Prenylation: Molecular Mechanisms and Functional Consequences," *Ann. Rev. Biochem.,* vol. 65, pp. 241-269, 1996.

[101] S. A. Holstein and R. A. Hohl, "Isoprenoids: remarkable diversity of form and function," *Lipids,* vol. 39, pp. 293-309, 2004.

[102] J. A. McIntosh, Z. Lin, M. D. B. Tianero and E. W. Schmidt, "Aestuaramides, a natural library of cyanobactin cyclic peptides resulting from isoprene-derived Claisen rearrangements," *ACS Chem. Biol.,* vol. 8, pp. 877-883, 2013.

[103] M. D. B. Tianero, M. S. Donia, T. S. Young, P. G. Schultz and E. W. Schmidt, "Ribosomal Route to Small-Molecule Diversity," *J. Am. Chem. Soc.,* vol. 134, pp. 418-425, 2012.

[104] J. D. Majmudar and R. A. Gibbs, "Pericyclic prenylation: peptide modification through a Claisen rearrangement," *ChemBioChem,* vol. 12, pp. 2723-2726, 2011.

[105] S. Zhang, G. Zubay and E. Goldman, "Low-usage codons in Escherichia coli, yeast, fruit fly and primates," *Gene,* vol. 105, pp. 61-72, 1991.

[106] H. Liu and J. H. Naismith, "An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol," *BMC Biotechnol.,* vol. 8, pp. 1-10, 2008.

[107] F. W. Studier, "Protein production by auto-induction in high density shaking cultures," *Protein Expression & Purification,* vol. 41, pp. 207-234, 2005.

[108] J. Jancarik and S.-H. Kim, "Sparse matrix sampling: a screening method for crystallization of proteins," *J. Appl. Cryst.,* vol. 24, pp. 409-411, 1991.

[109] A. McPherson, C. Nguyen, R. Cudney and S. B. Larson, "The Role of Small Molecule Additives and Chemical Modification in Protein Crystallization," *Cryst. Growth Des.,* vol. 11, pp. 1469-1474, 2011.

[110] G. Winter, "xia2: an expert system for macromolecular crystallography data reduction," *J. Appl. Cryst.,* vol. 43, pp. 186-190, 2010.

[111] W. Kabsch, "XDS," *Acta Cryst.,* vol. D66, pp. 125-132, 2010.

[112] P. Evans, "Scaling and assessment of data quality," *Acta Cryst.,* vol. D62, pp. 72-82, 2006.

[113] G. M. Sheldrick, "Experimental phasing with SHELXC/D/E: combining chain tracing with density modification," *Acta Cryst.,* vol. D66, pp. 479-485, 2010.

[114] G. M. Sheldrick, "A short history of SHELX," *Acta Cryst.,* vol. A64, pp. 112-122, 2007.

[115] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Jung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, "PHENIX: a comprehensive Python-based system for macromolecular structure solution," *Acta Cryst.,* vol. D66, pp. 213-221, 2010.

[116] P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, "Features and Development of Coot," *Acta Cryst.,* vol. D66, pp. 486-501, 2010.

[117] G. N. Murshudov, A. A. Vagin and E. J. Dodson, "Refinement of Macromolecular Structures by the Maximum-Likelihood Method," *Acta Cryst.,* vol. D53, pp. 240-255, 1997.

[118] M. D. Winn, C. C. Ballard, K. D. Cowtan, E. J. Dodson, P. Emsley, P. R. Evans, R. M. Keegan, E. B. Krissinel, A. G. W. Leslie, A. McCoy, S. J. McNicholas, G. N. Murshudov, N. S. Pannu, E. A. Potterton, H. R. Powell, R. J. Read, A. Vagin and K. S. Wilson, "Overview of the CCP4 suite and current developments," *Acta Cryst.,* vol. D67, pp. 235-242, 2011.

[119] E. Krissinel and K. Henrick, "Inference of Macromolecular Assemblies from Crystalline State," *J. Mol. Biol.,* vol. 372, pp. 774-797, 2007.

[120] E. Krissinel, "Crystal contacts as nature's docking solutions," *J. Comput. Chem.,* vol. 31, pp. 133-143, 2010.

[121] J. Painter and E. A. Merritt, "Optimal description of a protein structure in terms of multiple groups undergoing TLS motion," *Acta Cryst.,* vol. D62, pp. 439-450, 2006.

[122] J. Painter and E. A. Merritt, "TLSMD web server for the generation of multi-group TLS models," *J. Appl. Cryst.,* vol. 39, pp. 109-111, 2006.

[123] M. D. Winn, M. N. Isupov and G. N. Murshudov, "Use of TLS parameters to model anisotropic displacements in macromolecular refinement," *Acta Cryst.,* vol. D57, pp. 122-133, 2001.

[124] V. B. Chen, W. B. Arendall III, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J.

Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, "MolProbity: all-atom structure validation for macromolecular crystallography," *Acta Cryst.,* vol. D66, pp. 12-21, 2010.

[125] L. A. Kelley and M. J. Sternberg, "Protein structure prediction on the web: a case study using the Phyre server," *Nat. Protoc.,* vol. 4, pp. 363-371, 2009.

[126] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, "The Protein Data Bank," *Nucl. Acids Res.,* vol. 28, pp. 235-242, 2000.

[127] R. B. Kapust, J. Tozser, T. D. Copeland and D. S. Waugh, "The P1' specificity of tobacco etch virus protease," *Biochem. Biophys.,* vol. 294, pp. 949-955, 2002.

[128] N. Chayen, "Turning protein crystallisation from an art into a science," *Curr. Opin. Struc. Biol.,* vol. 14, pp. 577-583, 2004.

[129] G. L. Taylor, "Introduction to phasing," *Acta Cryst.,* vol. D66, pp. 325-338, 2010.

[130] E. E. Lattman and P. J. Loll, Protein crystallography; a concise guide, John Hopkins University Press, 2008.

[131] G. Rhodes, Crystallography made clear, 3rd Edt., Academic Press, 2006.

[132] H. Walden, "Selenium incorporation using recombination techniques," *Acta Cryst.,* vol. D66, pp. 352-357, 2010.

[133] P. D. Sun, S. Radaev and M. Kattah, "Generating isomorphous heavy-atom derivatives by a quick-soak method. Part I: test cases," *Acta Cryst.,* vol. D58, pp. 1092-1098, 2002.

[134] B. W. Matthews, "Solvent content of protein crystals," *J. Mol. Biol.,* vol. 33, pp. 491-497, 1968.

[135] T. Pape and T. R. Schneider, "HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs," *J. Appl. Crystallogr.,* vol. 37, pp. 843-844, 2004.

[136] D. W. Banner, A. Bloomer, G. A. Petsko, D. C. Philliips and I. A. Wilson, "Atomic coordinates for triose phosphate isomerase from chicken muscle," *Biochem. Biophys. Res. Commun.,* vol. 72, pp. 146-155, 1976.

[137] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, "Phaser crystallographic software," *J. Appl. Cryst.,* vol. 40, pp. 658-674, 2007.

[138] E. Potterton, S. McNicholas, E. Krissinel, K. Cowtan and M. Noble, "The CCP4 molecular-graphics project," *Acta Cryst.,* vol. D58, pp. 1955-1957, 2002.

[139] L. Potterton, S. McNicholas, E. Krissinel, J. Gruber, K. Cowtan, P. Emsley, G. N. Murshudov, S. Cohen, A. Perrakis and M. Noble, "Developments in the CCP4 molecular-graphics project," *Acta Cryst.,* vol. D60, pp. 2288-2294, 2004.

[140] U. Metzger, C. Schall, G. Zocher, I. Unsold, E. Stec, S. M. Li, L. Heide and T. Stehle, "The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria," *Proc. Natl. Acad. Sci. USA,* vol. 106, pp. 14309-14314, 2009.

[141] J. Koehnke, A. Bent, W. E. Houssen, D. Zollman, F. Morawitz, S. Shirran, J.

Vendome, A. F. Nneoyiegbe, L. Trembleau, C. H. Botting, M. C. Smith, M. Jaspars and J. H. Naismith, "The machanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain," *Nat. Struct. Mol. Biol.,* vol. 19, pp. 767-772, 2012.

[142] C. A. Smith, H. S. Toogood, H. M. Baker, R. M. Daniel and E. N. Baker, "Calcium-mediated thermostability in the subtilisin superfamily: the crystal structure of Bacillus Ak.1 protease at 1.8 Å resolution," *J. Mol. Biol.,* vol. 294, pp. 1027-1040, 1999.

[143] H. Liu and J. H. Naismith, "A simple and efficient expression and purification system using two newly constructed vectors," *Protein Expr. Purif.,* vol. 63, pp. 102-111, 2009.

[144] R. A. Laskowski and M. B. Swindells, "LigPlot+: multiple ligand-protein interaction diagrams for drug discovery," *J. Chem. Inf. Model.,* vol. 51, pp. 2778-2786, 2011.

[145] G. Ramachandran and V. Sasisekharan, "Conformation of polypeptides and proteins," *Adv. Protein. Chem,* vol. 23, pp. 283-438, 1968.

[146] D. E. Stewart, A. Sarkar and J. E. Wampler, "Occurence and role of cis peptide bonds in protein structures," *J. Mol. Biol.,* vol. 214, pp. 253-260, 1990.

[147] J. Koehnke, A. F. Bent, W. E. Houssen, G. Mann, M. Jaspars and J. H. Naismith, "The structural biology of patellamide biosynthesis," *Current Opinion in Structural Biology,* vol. 29, pp. 112-121, 2014.

[148] A. C. Braisted, J. K. Judice and J. A. Wells, "Synthesis of proteins by subtiligase," *Methods Enzymol.,* vol. 289, pp. 298-313, 1997.

[149] T. K. Chang, D. Y. Jackson, J. P. Burnier and J. A. Wells, "Subtiligase: a tool for semisynthesis of proteins," *Proc. Natl. Acad. Sci. USA,* vol. 91, pp. 12544-12548, 1994.

[150] D. Y. Jackson, J. P. Burnier and J. A. Wells, "Enzymic cyclization of linear peptide esters using subtiligase," *J. Am. Chem. Soc.,* vol. 117, pp. 819-820, 1995.

[151] S. H. Joo, "Cyclic peptides as therapeutic agents and biochemical tools," *Biomol. Ther. (Seoul).,* vol. 20, pp. 19-26, 2012.

[152] U. Schreiber, R. Gademann, P. J. Ralph and A. W. D. Larkum, "Assessment of photosynthetic performance of prochloron in Lissoclinum patella in hospite by chlorophyll fluorescence measurements," *Plant Cell Physiol.,* vol. 38, pp. 945-951, 1997.

[153] E. J. Kim, D. C. Love, E. Darout, M. Abdo, B. Rempel, S. G. Withers, P. R. Rablen, J. A. Hanover and S. Knapp, "OGA inhibition by GlcNAc-selenazoline," *Bioorg. Med. Chem.,* vol. 18, pp. 7058-7064, 2010.

[154] A. Bodtke, M. Kandt, W.-D. Pfeiffer and P. Langer, "Synthesis of 4-Aryl-2-imino-2H-selenazolines by a Reaction of α α-(Selenocyanato)acetophenones With Anilines," *Phosphorus, Sulfur, and Silicon and the Related Elements,* vol. 182, pp. 209-217, 2007.

[155] M. Koketsu, H. Tanaka and H. Ishihara, "A Facile Preparation of 2-Amino-1,3-selenazoles by Reactions of N,N-Unsubstituted Selenoureas with

Chloroacetonitrile," *Chemistry Letters,* vol. 34, pp. 1260-1261, 2005.

[156] H. Tao, Y. Weng, R. Zhuo, G. Chang, I. L. Urbatsch and Q. Zhang, "Design and Synthesis of Selenazole-Containing Peptides for Cocrystallization with P-Glycoprotein," *ChemBioChem,* vol. 12, pp. 868-873, 2011.

[157] P. S. Salgado, J. D. Taylor, E. Cota and S. J. Matthews, "Extending the usability of the phasing power of diselenide bonds: SeCys," *Acta Cryst.,* vol. D67, pp. 8-13, 2011.

[158] F. W. Studier and B. A. Moffat, "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes," *J. Mol. Biol.,* vol. 189, pp. 113-130, 1986.

[159] D. E. Kim, D. Chivian and D. Baker, "Protein structure prediction and analysis using the Robetta server," *Nucleic Acids Res.,* vol. 32, pp. W526-W531, 2004.

[160] T. S. Young, I. Ahmad, J. A. Yin and P. G. Schultz, "An enhanced system for unnatural amino acid mutagenesis in E. coli," *J. Mol. Biol.,* vol. 395, pp. 361-374, 2010.

[161] Y. Ryu and P. G. Schultz, "Efficient incorporation of unnatural amino acids into proteins in Escherichia coli," *Nature Methods,* vol. 3, pp. 263-265, 2006.

[162] C. Cuevas and A. Francesch, "Development of Yondelis® (trabectedin, ET-743). A semisynthetic process solves the supply problem," *Nat. Prod. Rep.,* vol. 26, pp. 322-337, 2009.

[163] C. Cuevas, M. Perez, M. J. Martin, J. L. Chicharro, C. Fernandez-Rivas, M. Flores, A. Francesch, P. Gallego, M. Zarzuelo, F. de la Calle, J. Garcia, C. Polanco, I. Rodriguez and I. Manzanares, "Synthesis of Ecteinascidin ET-743 and Phthalascidin Pt-650 from Cyanosafracin B," *Org. Lett.,* vol. 2, pp. 2545-2548, 2000.

[164] M. Schnolzer and S. B. Kent, "Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease," *Science,* vol. 256, pp. 221-225, 1992.

[165] P. M. Kane, C. T. Yamashiro, D. F. Wolczyk, N. Neff, M. Goebl and T. H. Stevens, "Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase," *Science,* vol. 250, pp. 651-657, 1990.

[166] S. Elleuche and S. Poggeler, "Inteins, valuable genetic elements in molecular biology and biotechnology," *Appl. Microbiol. Biotechnol.,* vol. 87, pp. 479-489, 2010.

[167] F. B. Perler, E. O. Davis, G. E. Dean, F. S. Gimble, W. E. Jack, N. Neff, C. J. Noren, J. Thorner and M. Belfort, "Protein splicing elements: inteins and exteins--a definition of terms and recommended nomenclature," *Nucleic Acids Res.,* vol. 22, pp. 1125-1127, 1994.

[168] S. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus and M.-Q. Xu, "Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element," *Gene,* vol. 192,

pp. 271-281, 1997.

[169] T. C. Evans, J. Benner and M.-Q. Xu, "Semisynthesis of cytotoxic proteins using a modified protein splicing element," *Protein Sci.,* vol. 7, pp. 2256-2264, 1998.

[170] T. W. Muir, D. Sondhi and P. A. Cole, "Expressed protein ligation: A general method for protein engineering," *Proc. Natl. Acad. Sci. USA,* vol. 95, pp. 6705-6710, 1998.

[171] H. Sancheti and J. A. Camarero, ""Splicing up" drug discovery. Cell-based expression and screening of genetically-encoded libraries of backbone-cyclized polypeptides," *Adv. Drug Deliv. Rev.,* vol. 61, pp. 908-917, 2009.

[172] M.-Q. Xu and T. C. Evans, "Intein-Mediated Ligation and Cyclization of Expressed Proteins," *Methods,* vol. 24, pp. 257-277, 2001.

[173] J. Normanly, L. G. Kleina, J.-M. Masson, J. Abelson and J. H. Miller, "Construction of escherichia coli amber suppressor tRNA genes," *J. Mol. Biol.,* vol. 213, pp. 719-726, 1990.

[174] L. Wang, T. J. Magliery, D. R. Liu and P. G. Schultz, "A new functional suppressor tRNA/aminoacyl−tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins," *J. Am. Chem. Soc.,* vol. 122, pp. 5010-5011, 2000.

[175] L. Wang, J. Xie and P. G. Schultz, "Expanding the genetic code," *Annu. Rev. Biophys. Biomol. Struct.,* vol. 35, pp. 225-249, 2006.

[176] M. Z. Lin and L. Wang, "Selective labeling of proteins with chemical probes in living cells," *Physiology,* vol. 23, pp. 131-141, 2008.

[177] C. C. Liu and P. G. Schultz, "Adding new chemistries to the genetic code," *Annu. Rev. Biochem.,* vol. 79, pp. 413-444, 2010.

[178] J. Wang, J. Xie and P. G. Schultz, "A genetically encoded fluorescent amino acid," *J. Am. Chem. Soc.,* vol. 128, pp. 8738-8739, 2006.

[179] N. Hino, Y. Okazaki, A. Kobayashi, K. Sakamoto and S. Yokoyama, "Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid," *Nat. Methods,* vol. 2, pp. 201-206, 2005.

[180] M. R. Fleissner, E. M. Brustad, T. Kalai, C. Altenbach, D. Cascio, F. B. Peters, K. Hideg, S. Peuker, P. G. Schultz and W. L. Hubbell, "Site-directed spin labeling of a genetically encoded unnatural amino acid," *Proc. Natl. Acad. Sci. USA,* vol. 106, pp. 21637-21642, 2009.

[181] R. Xu, S. R. Hanson, Z. Zhang, Y. Y. Yang, P. G. Schultz and C. H. Wong, "Site-specific incorporation of the mucin-type N-acetylgalactosamine-alpha-O-threonine into protein in Escherichia coli," *J. Am. Chem. Soc.,* vol. 126, pp. 15654-15655, 2004.

[182] J. Xie, L. Wang, N. Wu, A. Brock, G. Spraggon and P. G. Schultz, "The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination," *Nat. Biotechnol.,* vol. 22, pp. 1297-1301, 2004.

[183] J. W. Chin, A. B. Martin, D. S. King, L. Wang and P. G. Schultz, "Addition of a photocrosslinking amino acid to the genetic code of Escherichia coli," *Proc. Natl. Acad. Sci. USA,* vol. 99, pp. 11020-11024, 2002.

[184] M. W. Southworth, K. Amaya, T. C. Evans, M.-Q. Xu and F. B. Perler, "Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein," *Biotechniques,* vol. 27, pp. 110-120, 1999.

[185] A. Telenti, M. Southworth, F. Alcaide, S. Daugelat, W. R. Jacobs Jr and F. B. Perler, "The Mycobacterium xenopi GyrA protein splicing element: characterization of a minimal intein," *J. Bacteriol.,* vol. 179, pp. 6378-6382, 1997.

9. Appendices

Appendix A. Media and Buffer Compositions

A. 1 PatF

Auto-Induction Medium	ZY (10 g N-Z-amine AS, 5 g yeast extract, 925 ml H_2O / litre culture) media supplemented with 1 mM $MgSO_4$, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, 25 mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4
Minimal Medium	19 mM NH_4Cl, 22 mM KH_2PO_4, 22 mM Na_2HPO_4, 9 mM NaCl
Lysis buffer	150 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 0.1 % Triton X-100, 3 mM BME
Elution Buffer	150 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0, 0.1 % Triton X-100, 3 mM BME
Desalt/Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP, 10 % glycerol.

A. 2 LynF

Auto-Induction Medium	ZY (10 g N-Z-amine AS, 5 g yeast extract, 925 ml H_2O / litre culture) media supplemented with 1 mM $MgSO_4$, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, 25 mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4
Lysis Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME
Elution Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0, 3 mM BME

Desalt Buffer	150 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME.
Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP.

A.3 PatGmac

Auto-Induction Medium	ZY (10 g N-Z-amine AS, 5 g yeast extract, 925 ml H_2O) media supplemented with 1 mM $MgSO_4$, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, 25 mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4.
Lysis Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME
Elution Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0, 3 mM BME
Desalt Buffer	100 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME
Mono-Q Buffer A	20 mM Tris-HCl pH 8.0, 3 mM BME
Mono-Q Buffer B	1 M NaCl, 20 mM Tris-HCl pH 8.0, 3 mM BME
Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP

A.4 PatE

Auto-Induction Medium	ZY (10 g N-Z-amine AS, 5 g yeast extract, 925 ml H_2O) media supplemented with 1 mM $MgSO_4$, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, 25 mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4.
Lysis Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME, 8 M Urea
Elution Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0, 3 mM BME, 8 M Urea

Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP

A.5 TruD

Auto-Induction Medium	ZY (10 g N-Z-amine AS, 5 g yeast extract, 925 ml H_2O) media supplemented with 1 mM $MgSO_4$, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, 25 mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4.
Lysis Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME
Elution Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0, 3 mM BME
Desalt Buffer	100 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME
Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP

A.6 PatApr

Auto-Induction Medium	ZY (10 g N-Z-amine AS, 5 g yeast extract, 925 ml H_2O) media supplemented with 1 mM $MgSO_4$, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, 25 mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4.
Lysis Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME
Elution Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0, 3 mM BME
Desalt Buffer	100 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME
Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP

A.7 PatD

Auto-Induction Medium	ZY (10 g N-Z-amine AS, 5 g yeast extract, 925 ml H_2O) media supplemented with 1 mM $MgSO_4$, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, 25 mM $(NH_4)2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4.
Lysis Buffer	500 mM NaCl, 20 mM Bis-Tris pH 6.8, 20 mM imidazole pH 8.0, 3 mM BME
Elution Buffer	500 mM NaCl, 20 mM Bis-Tris pH 6.8, 250 mM imidazole pH 8.0, 3 mM BME
Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP

A.8 PatE SeCys

Minimal Medium	19 mM NH_4Cl, 22 mM KH_2PO_4, 22 mM Na_2HPO_4, 9mM NaCl
Lysis Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME, 8 M Urea
Elution Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0, 3 mM BME, 8 M Urea
Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP

Amino Acid Mixture for Selenocysteine Growth per 1 L SeCys Media:

1.6 g	serine
1 g	leucine
0.4 g	alanine, glutamate, glutamine, arginine, glycine
0.25 g	aspartate
0.1 g	lysine, threonine, phenylalanine, asparagine, histidine, proline, tyrosine, tryptophan, methionine
0.05 g	isoleucine and valine

A.9 PatGox

Lysis Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME
Elution Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0, 3mM BME
Capto-Q Buffer A	100 mM NaCl, 20 mM Tris-HCl pH 8.0, 3 mM BME
Capto-Q Buffer B	1M NaCl, 20 mM Tris-HCl pH 8.0, 3 mM BME
Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP

A.10 PatGox Homologues (including ArtGox and CyaGox (PCC7425))

Auto-Induction Medium	ZY (10 g N-Z-amine AS, 5 g yeast extract, 925 ml H_2O) media supplemented with 1 mM MgSO4, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose, 25 mM $(NH_4)_2SO_4$, 50 mM KH_2PO_4, 50 mM Na_2HPO_4, 50 µM FMN.
Lysis Buffer	150 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 50 µM FMN, 3 mM BME
Elution Buffer	150 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0, 50 µM FMN, 3 mM BME
Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP

A.11 Fourteen Buffer Screen

For each lysis buffer, the equivalent elution buffer is the same with the exception of 250 mM imidazole rather than 20 mM.

Lysis Buffer 1:	150 mM NaCl, 20 mM Bis-Tris pH 6.8, 20 mM imidazole pH 8.0, 3 mM BME

Lysis Buffer 2:	150 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 3:	500 mM NaCl, 20 mM Bis-Tris pH 6.8, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 4:	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 5:	150 mM NaCl, 20 mM Bis-Tris pH 6.8, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 6:	150 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.1M arginine, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 7:	150 mM NaCl, 20 mM Bis-Tris pH 6.8, 0.1M glycine, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 8:	150 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.1M glycine, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 9:	150 mM NaCl, 20 mM Bis-Tris pH 6.8, 0.1 % NP-40, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 10:	150 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.1 % Triton X-100, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 11:	150 mM NaCl, 20 mM Bis-Tris pH 6.8, 0.1 % Triton X-100, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 12:	150 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.1 % NP-40, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 13:	150 mM NaCl, 20 mM Bis-Tris pH 6.8, 0.1 % β-OG, 20 mM imidazole pH 8.0, 3 mM BME
Lysis Buffer 14:	150 mM NaCl, 20 mM Tris-HCl pH 8.0, 0.1 % β-OG, 20 mM imidazole pH 8.0, 3 mM BME

A.12 Full Length PatE - Intein

Lysis Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0, 8 M Urea
Elution Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0, 8 M Urea
Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP
Cleavage Buffer:	150 mM NaCl, 10 mM HEPES pH 7.4, 10 mM TCEP, 200 mM MESNA/MPAA

A.13 PatE – Intein (Small, Large, GLEA)

Lysis Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 20 mM imidazole pH 8.0,
Elution Buffer	500 mM NaCl, 20 mM Tris-HCl pH 8.0, 250 mM imidazole pH 8.0,
Gel Filtration Buffer	150 mM NaCl, 10 mM HEPES pH 7.4, 1 mM TCEP
Cleavage Buffer:	150 mM NaCl, 10 mM HEPES pH 7.4, 10 mM TCEP, 200 mM MESNA/MPAA

Appendix B. Sequence Alignment of Members of the PatF Family

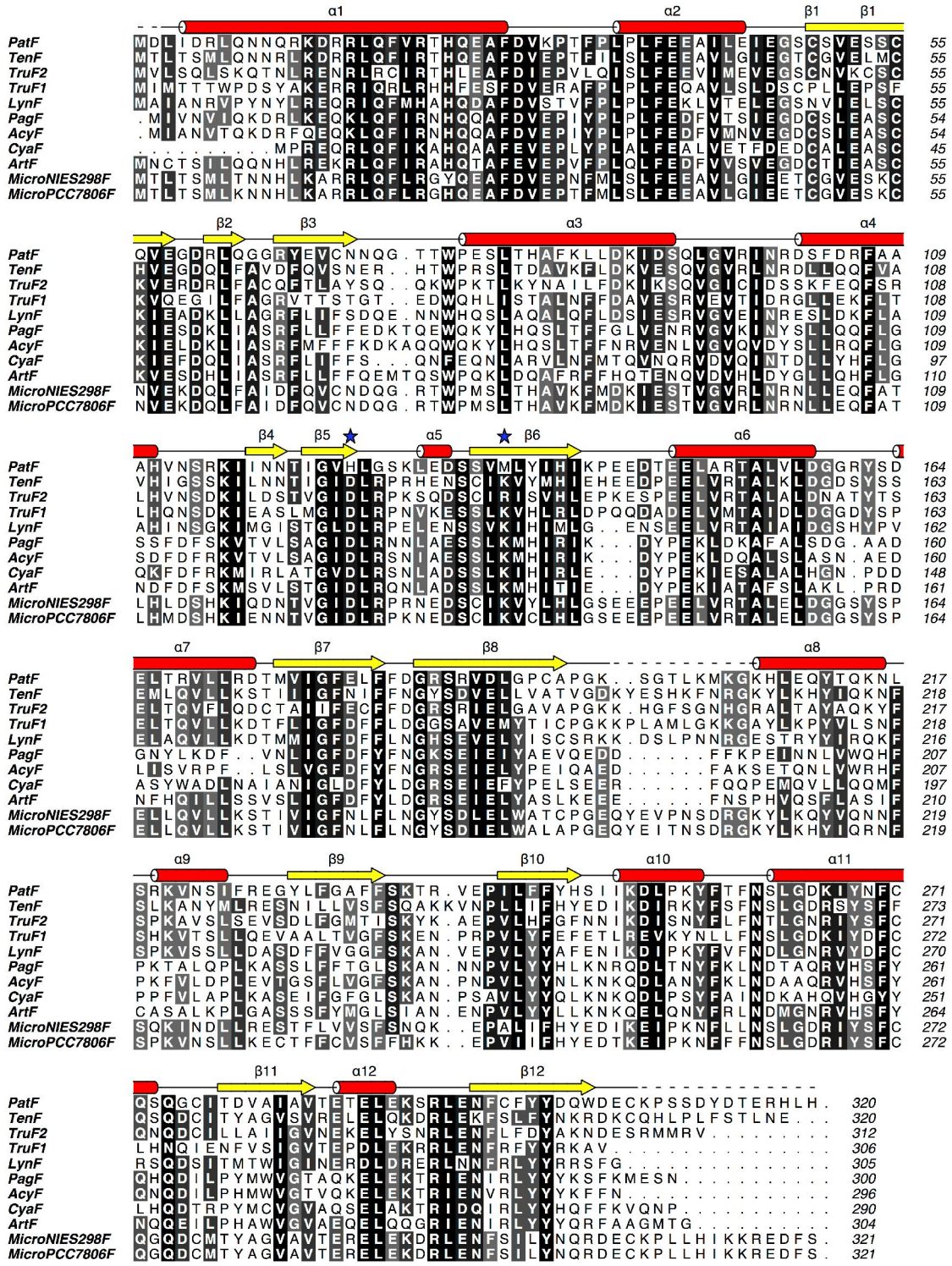


Figure B: Sequence Alignment of Members of the PatF Family. Absolutely conserved residues are shown in black. Secondary structure elements for PatF highlighted above (alpha helices – red, beta strands – yellow). Potential DMAPP binding residues are indicated by stars.

Appendix C. Sequence Alignment of Members of the PatG Macrocyclase Family

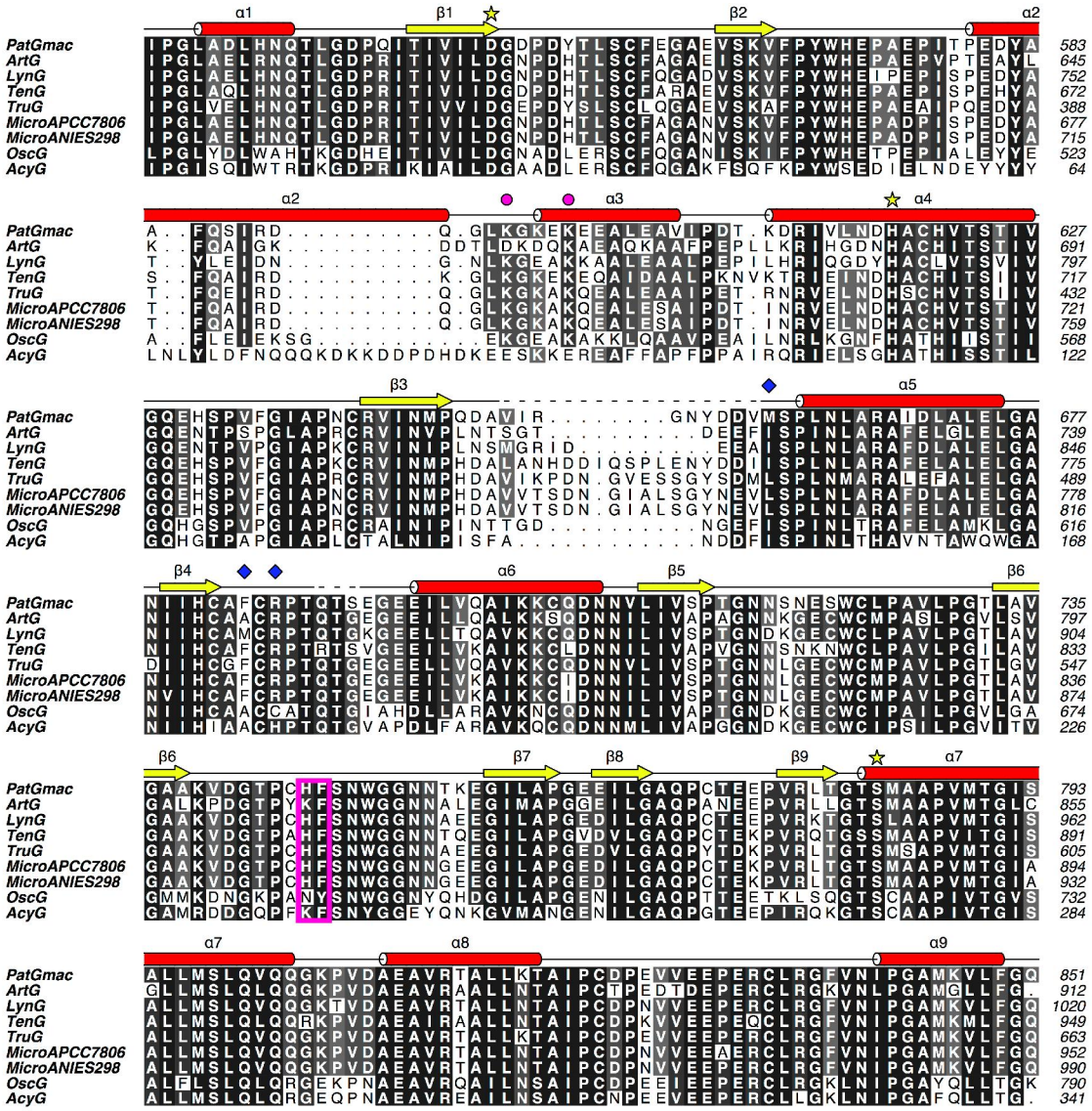


Figure C: Sequence Alignment of Members of the PatG Macrocyclase Family. Secondary structure elements are shown in red (alpha helices) and yellow (beta strands). Active site residues are indicated by yellow stars, residues blocking the S3 and S4 sites as blue diamonds, lysines forming salt-bridges with the substrate as purple circles and His and Phe residues involved in substrate binding are marked by a magenta box.

Appendix D. Mass spectrometry data for *in vitro* derived cyclic compounds.

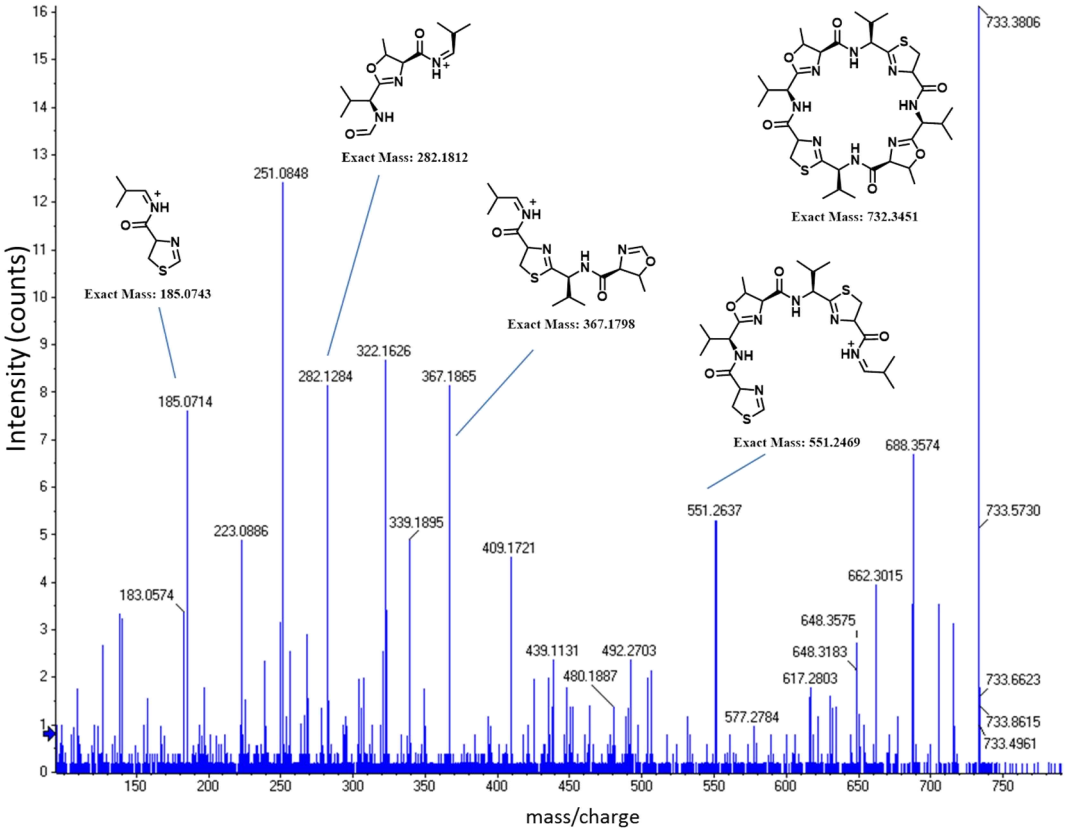


Figure D.1: Mass Fragmentation of cyclo[VT^{Oxn}VC^{Thn}VT^{Oxn}VC^{Thn}].

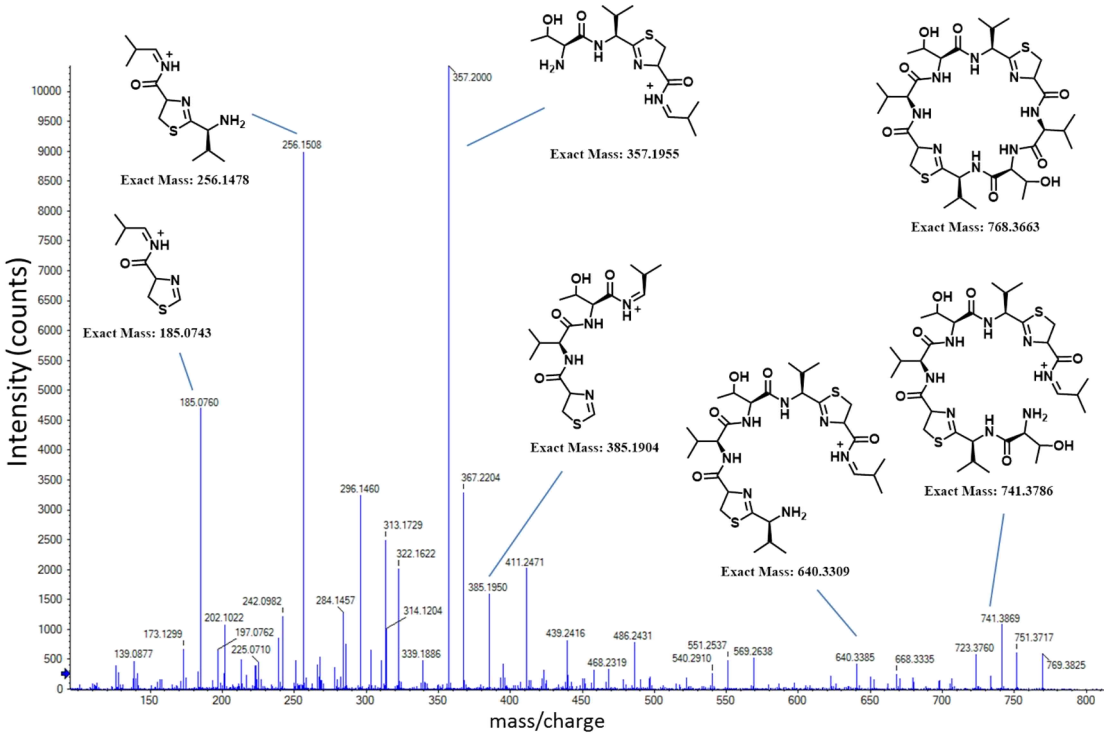


Figure D.2: Mass Fragmentation of cyclo[VTVC^{Thn}VTVC^{Thn}].

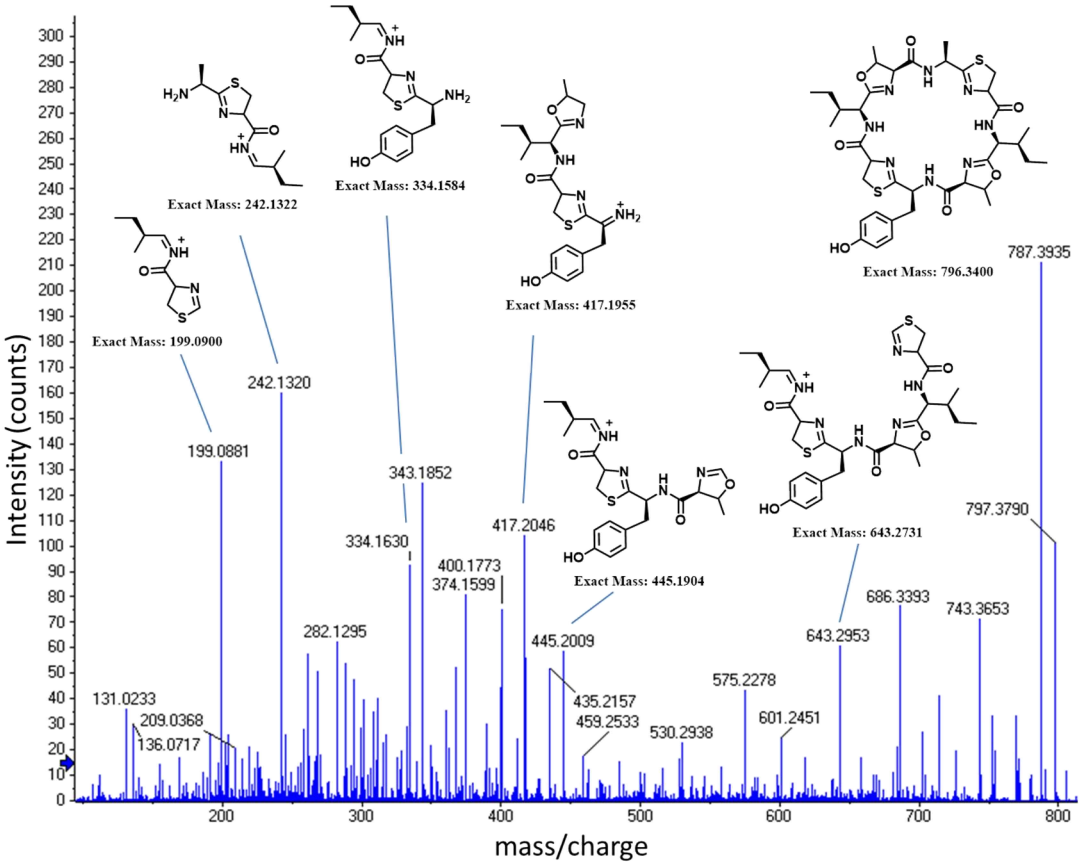


Figure D.3: Mass Fragmentation of cyclo[IT^{Oxn}AC^{Thn}IT^{Oxn}YC^{Thn}].

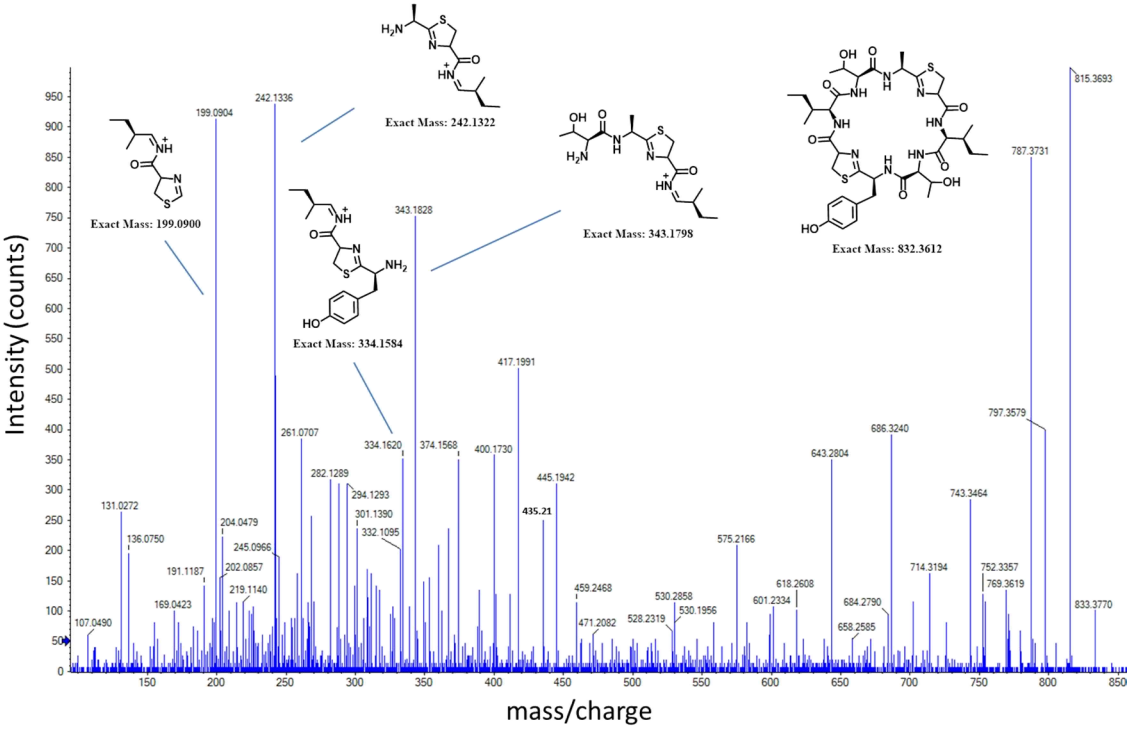


Figure D.4: Mass Fragmentation of cyclo[ITAC^{Thn}ITYC^{Thn}].

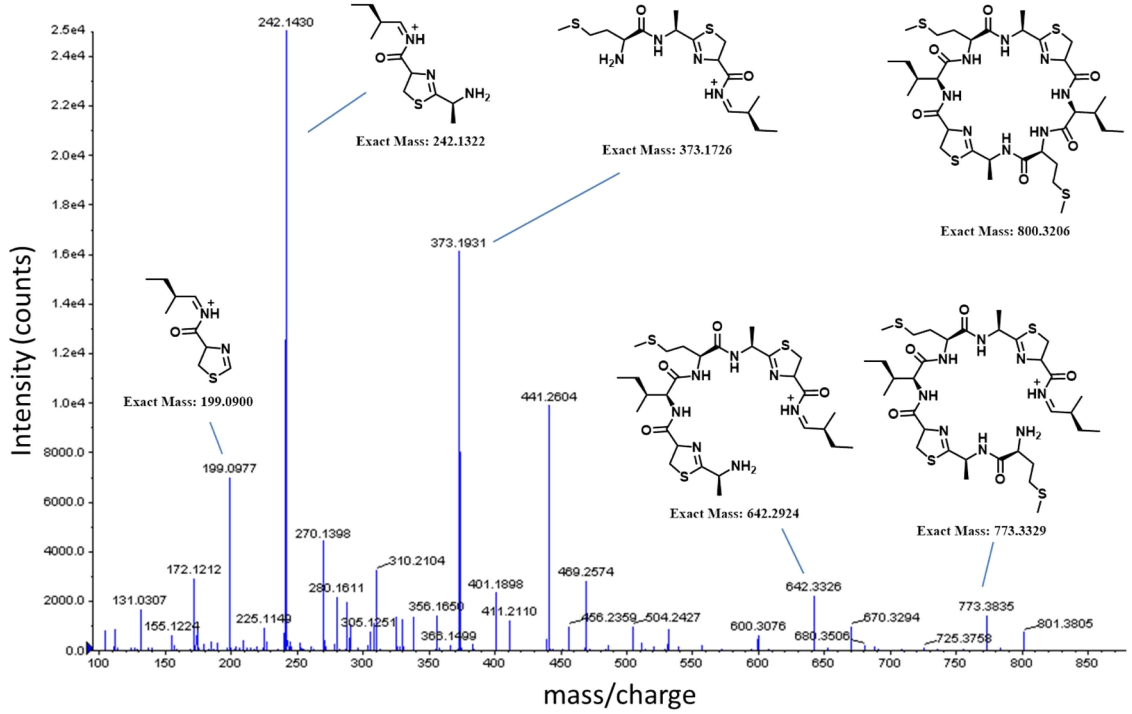


Figure D.5: Mass Fragmentation of cyclo[IMAC^{Thn}IMAC^{Thn}].

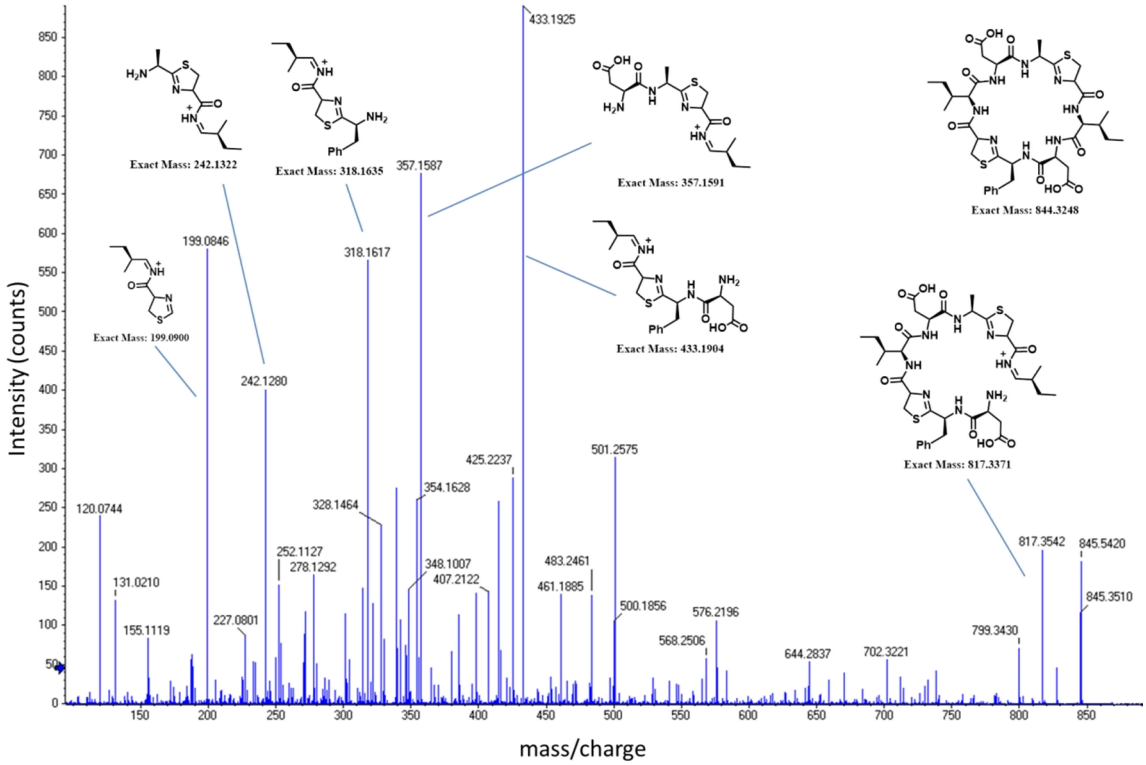


Figure D.6: Mass Fragmentation of cyclo[IDAC^{Thn}IDFC^{Thn}].

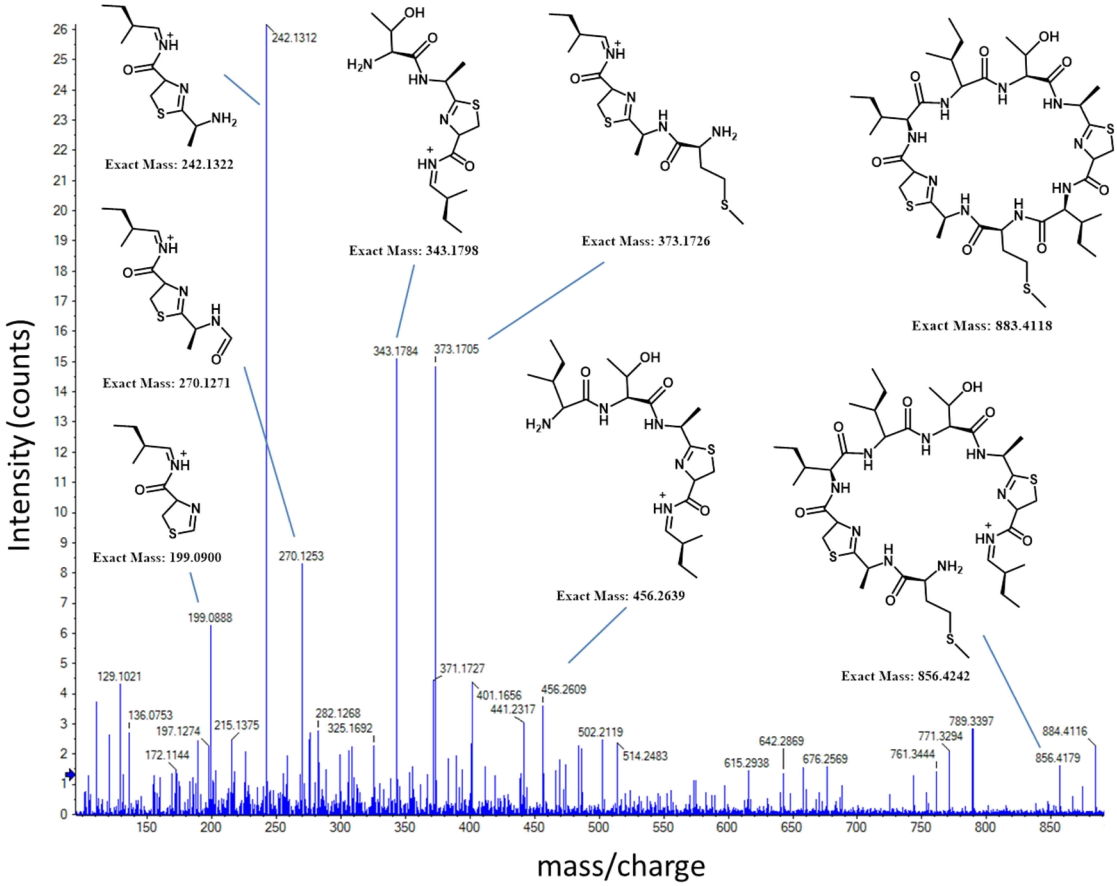


Figure D.7: Mass Fragmentation of cyclo[IITAC^{Thn}IMAC^{Thn}].

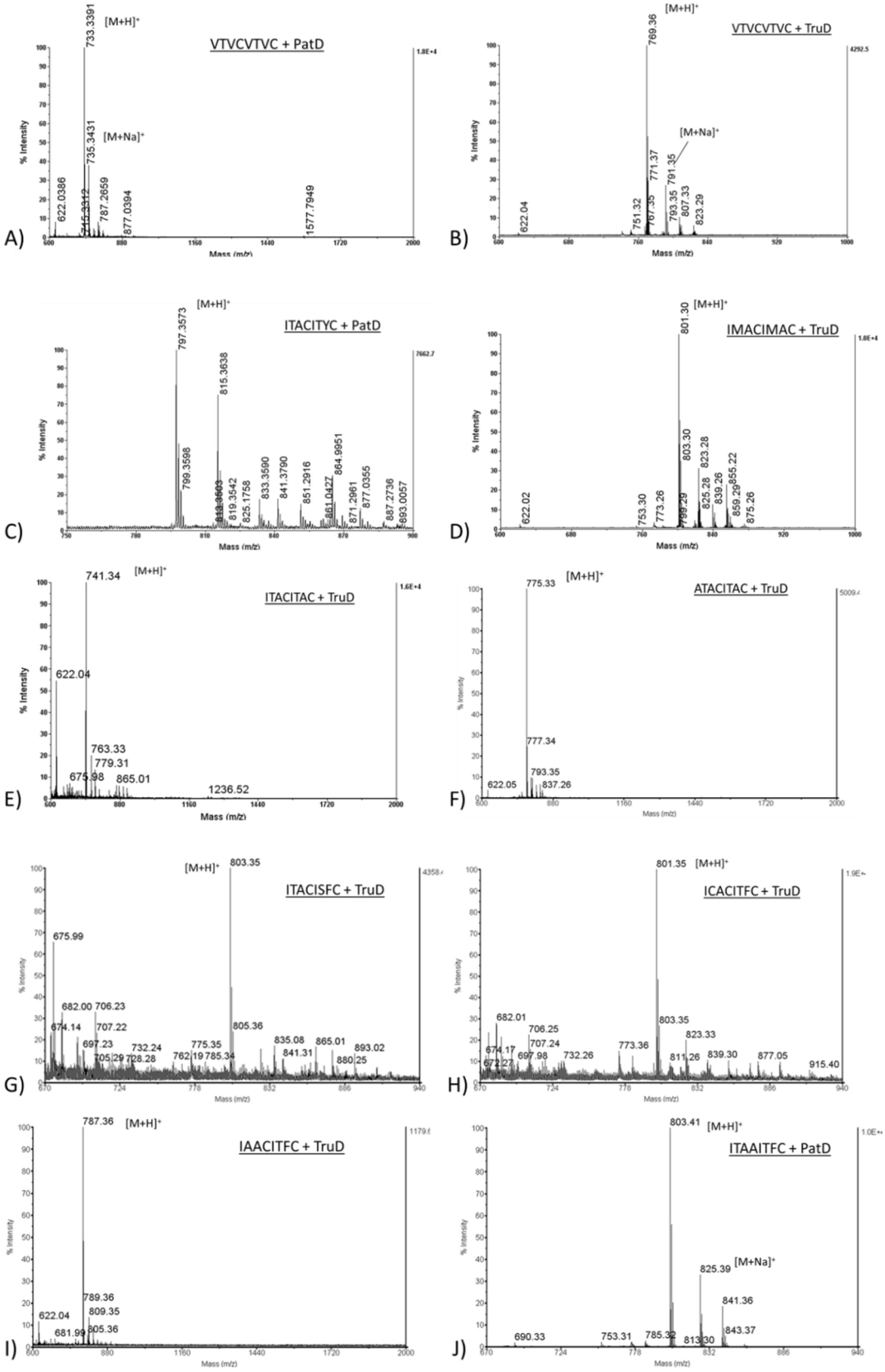


Figure D.8: MALDI-MS of *in vitro* Derived Cyclic Peptides. A) cyclo[VT^{Oxn}VC^{Thn}VT^{Oxn}VC^{Thn}], B) cyclo[VTVC^{Thn}VTVC^{Thn}], C) cyclo[IT^{Oxn}AC^{Thn}IT^{Oxn}YC^{Thn}], D) cyclo[IMAC^{Thn}IMAC^{Thn}]., E) cyclo[ITAC^{Thn}ITAC^{Thn}], F) cyclo[ATAC^{Thn}ITAC^{Thn}], G) cyclo[ITAC^{Thn}ISFC^{Thn}], H) cyclo[IC^{Thn}AC^{Thn}ITFC^{Thn}], I) cyclo[IAAC^{Thn}ITFC^{Thn}] and J) cyclo[ITAAITFC^{Thn}].

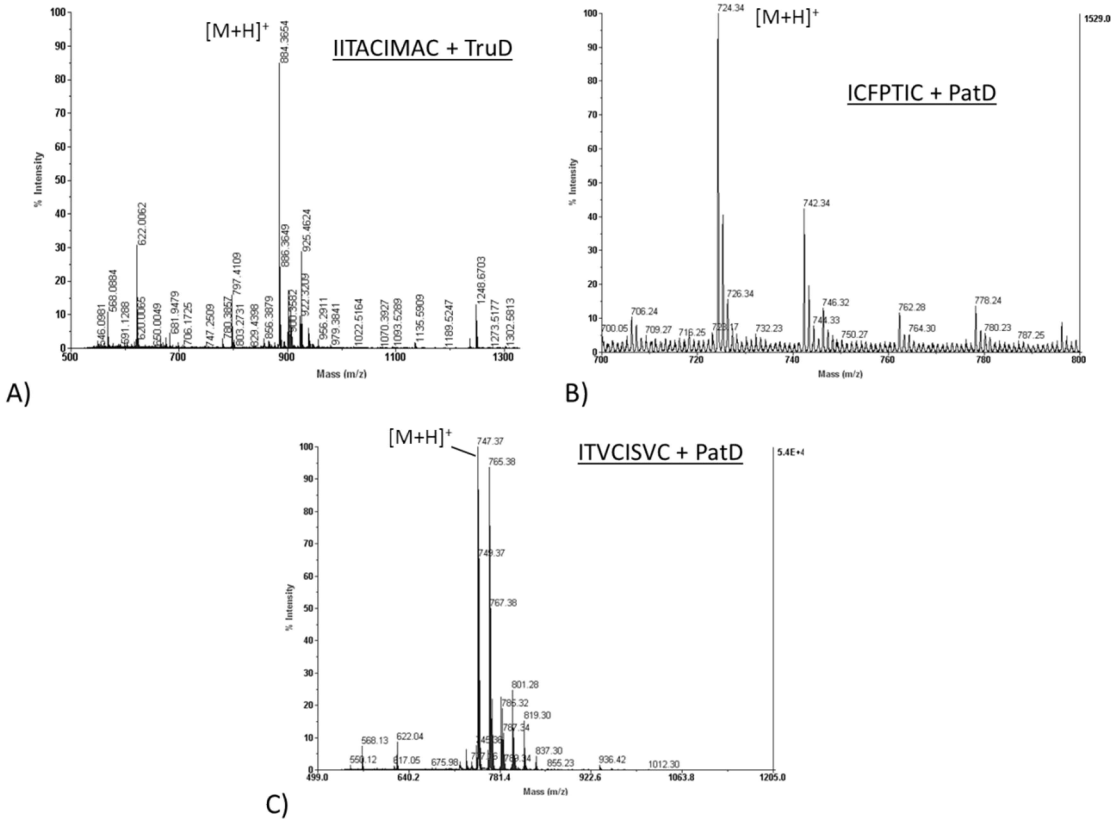


Figure D.9: MALDI-MS of *in vitro* Derived Cyclic Peptides. A) cyclo[IITAC^{Thn}IMAC^{Thn}], B) cyclo[IC^{Thn}FPT^{Oxn}IC^{Thn}] and C) cyclo[IT^{Oxn}VC^{Thn}IS^{Oxn}VC^{Thn}]

Appendix E. Stochastic Crystallisation Screen Composition Tables.

E.1 Stochastic Screen 18
- generated in-house

No.	Condition
1	27.8 % PEG8000, 0.03 M Zinc chloride
2	0.1 M HEPES pH 7.5, 26.4 % PEG6000, 0.09 M Calcium acetate, 1.91 % Ethylene glycol
3	0.1 M Tris pH 8.0, 20.0 % PEGMME5000, 0.06 M Sodium chloride
4	0.1 M Sodium citrate pH 5.0, 1.0 M Sodium citrate
5	0.1 M Sodium cacodylate pH 7.0, 16.3 % PEG10000, 0.05 M Sodium potassium phosphate
6	38.9 % PEG1500, 0.08 M Zinc acetate
7	0.1 M HEPES pH 7.5, 52.8 % PEG400, 0.09 M Sodium bromide, 0.4 % Methanol
8	0.1 M MES pH 6.0, 22.5 % PEGMME5000, 0.08 M Zinc acetate, 0.9 % Methanol
9	25.5 % PEG3350, 0.03 M Zinc chloride
10	0.1 M Tris pH 8.5, 31.8 % PEGMME2000, 0.09 M Calcium acetate
11	0.1 M Sodium citrate pH 5.5, 1.8 M Sodium formate, 0.28 M Magnesium formate
12	0.1 M Sodium acetate pH 5.0, 2.3 M Sodium chloride, 0.11 M Potassium thiocyanate, 3.5 % PEGMME350
13	0.1 M CHES pH 9.5, 27.2 % PEG6000, 0.29 M Sodium chloride
14	0.1 M Tris pH 8.0, 2.8 M Sodium malonate
15	0.1 M Bicine pH 9.0, 51.6 % MPD
16	24.5 % PEGMME2000, 0.09 M Sodium potassium tartrate
17	0.1 M MOPS pH 6.5, 2.8 M Sodium acetate
18	0.1 M Sodium citrate pH 5.0, 20.0 % Ethanol, 0.16 M Lithium chloride, 1.6 % PEG400
19	0.1 M Bicine pH 8.5, 14.0 % PEG4000, 2.5 % PEGDME250
20	0.1 M MES pH 6.5, 2.3 M Ammonium phosphate, 1.7 % Butanediol
21	0.1 M Tris pH 8.5, 12.8 % Isopropanol, 0.11 M Calcium chloride
22	0.1 M Tris pH 7.5, 2.0 M Sodium malate
23	0.1 M Sodium acetate pH 5.5, 0.90 M Sodium citrate
24	0.1 M Bicine pH 9.0, 44.0 % PEGMME550, 0.13 M Ammonium citrate
25	0.1 M Bicine pH 9.5, 11.6 % PEG10000, 0.04 M Ammonium tartrate
26	1.8 M Sodium potassium phosphate, 0.2 M Ammonium phosphate
27	0.1 M Sodium acetate pH 5.0, 1.4 M Magnesium sulfate, 0.09 M Sodium citrate
28	0.1 M MOPS pH 6.5, 2.1 M Sodium malate, 2.4 % PEG400
29	0.1 M MOPS pH 7.0, 28.2 % PEG3350, 0.07 M Sodium citrate
30	1.1 M Sodium tartrate, 0.13 M Ammonium citrate, 3.7 mM BME
31	0.1 M Sodium acetate pH 4.5, 18.1 % PEGMME5000, 0.1 M Calcium chloride, 0.06 % LDAO
32	0.1 M MES pH 6.0, 34.8 % PEG1500, 0.25 M Lithium sulfate
33	17.1 % PEG4000, 0.09 M Zinc chloride
34	0.1 M Bicine pH 9.0, 1.1 M Magnesium sulfate
35	0.1 M CHES pH 9.0, 39.6 % PEG400, 0.29 M Lithium sulfate, 2.7 % PEG400
36	0.1 M Bicine pH 9.5, 2.4 M Sodium malonate
37	0.1 M Sodium acetate pH 5.5, 29.0 % PEG4000, 0.05 M Sodium potassium phosphate
38	0.1 M Sodium acetate pH 5.5, 34.8 % PEGMME550, 0.03 M Calcium acetate
39	0.1 M Sodium citrate pH 5.0, 2.1 M Sodium malonate, 0.22 M Potassium thiocyanate
40	0.1 M MOPS pH 7.0, 2.2 M Sodium malate
41	0.1 M Sodium cacodylate pH 6.5, 1.9 M Sodium potassium phosphate
42	0.1 M Bicine pH 8.5, 1.9 M Ammonium sulfate
43	0.1 M Sodium citrate pH 4.5, 2.4 M Sodium malate, 0.19 M Potassium nitrate
44	0.1 M Tris pH 7.5, 1.7 M Ammonium phosphate, 0.29 M Potassium chloride
45	0.1 M Sodium citrate pH 4.5, 30.8 % PEG1500, 0.08 % CHAPS
46	0.1 M Tris pH 8.0, 16.2 % PEG10000, 0.05 M Zinc acetate

47	0.1 M MOPS pH 6.5, 34.5 % PEG3350, 0.09 M Sodium potassium tartrate
48	53.9 % MPD, 0.15 M Ammonium sulfate, 3.0 % DMSO
49	0.1 M Tris pH 8.0, 20.0 % Ethanol, 0.18 M Magnesium acetate, 0.08 % βOG
50	0.1 M MES pH 6.0, 16.6 % PEG10000, 1.3 % Dioxane
51	0.1 M BisTris pH 6.0, 27.0 % PEG3350, 0.08 M Zinc acetate, 0.05 % CHAPS
52	0.1 M Bicine pH 9.0, 2.5 M Ammonium sulfate, 0.23 M Potassium nitrate, 0.9 % PEGMME350
53	0.1 M CHES pH 9.0, 39.5 % MPD
54	0.1 M Sodium cacodylate pH 7.0, 2.4 M Sodium malonate, 0.13 M Ammonium tartrate
55	0.1 M Tris pH 8.5, 26.5 % PEG6000, 0.1 M Ammonium citrate
56	0.1 M MES pH 6.0, 20.2 % PEGMME5000, 0.10 M Magnesium formate
57	0.1 M MES pH 6.5, 1.1 M Sodium citrate, 0.29 M Magnesium sulfate, 2.2 % Hexanediol
58	0.1 M MOPS pH 7.0, 20.5 % PEGMME5000
59	0.1 M Bicine pH 8.5, 51.3 % PEG400, 0.11 M Sodium chloride
60	0.1 M Sodium acetate pH 5.5, 22.0 % PEGMME2000
61	0.1 M Tris pH 8.5, 2.7 M Sodium malonate, 0.10 M Potassium chloride
62	24.8 % PEGMME2000, 0.17 M Sodium bromide
63	0.1 M Sodium acetate pH 5.0, 19.0 % PEG6000, 0.11 M Lithium chloride
64	0.1 M CHES pH 9.5, 2.1 M Sodium malate, 0.17 M Ammonium sulfate
65	0.1 M Sodium citrate pH 5.0, 27.1 % PEG4000, 0.01 % LDAO
66	0.1 M Tris pH 8.0, 24.3 % PEG8000, 0.05 M Ammonium phosphate
67	0.1 M Sodium citrate pH 5.0, 1.6 M Ammonium sulfate, 0.14 M Ammonium acetate, 2.9 % Glycerol
68	0.1 M BisTris pH 6.0, 2.4 M Sodium chloride, 0.08 M Calcium acetate
69	0.1 M Sodium acetate pH 4.5, 53.0 % MPD, 0. 28 M Potassium thiocyanate, 1.3 % PEGDME250
70	0.1 M Sodium citrate pH 5.5, 31.8 % PEGMME2000, 0.10 M Ammonium acetate, 2.7 % PEGMME350
71	0.1 M Sodium acetate pH 4.5, 25.5 % PEG400, 0.24 M Potassium nitrate
72	0.1 M Sodium citrate pH 4.5, 2.0 M Ammonium phosphate, 0.07 M Potassium chloride, 3.2 % Ethylene glycol
73	0.1 M Sodium citrate pH 5.5, 38.0 % PEGMME550, 0.26 M Lithium sulfate, 0.1 % LDAO
74	0.1 M MOPS pH 7.0, 37.8 % PEG1500, 0.19 M Potassium chloride, 6.3 mM EDTA
75	0.1 M HEPES pH 8.0, 21.6 % PEGMME550, 0.07 M Magnesium chloride
76	0.1 M Sodium cacodylate pH 7.0, 2.2 M Ammonium phosphate, 0.1 M Ammonium acetate
77	0.1 M Tris pH 7.5, 2.4 M Sodium acetate, 0.11 M Potassium thiocyanate
78	0.1 M CHES pH 9.0, 18.7 % PEGMME5000, 0.27 M Ammonium phosphate
79	0.1 M Sodium cacodylate pH 7.0, 2.5 M Sodium formate
80	0.1 M Bicine pH 9.5, 19.7 % PEG8000
81	0.1 M BisTris pH 6.5, 2.0 M Sodium acetate
82	0.1 M Sodium cacodylate pH 7.0, 2.0 M Sodium malate, 0.09 M Ammonium phosphate
83	0.1 M Tris pH 8.6, 25.3 % PEG400, 0.17 M Magnesium acetate
84	0.1 M Sodium citrate pH 5.5, 1.7 M ammonium phosphate, 0.10 M Sodium acetate
85	0.1 M Tris pH 7.5, 1.7 M Magnesium sulfate, 0.16 M Magnesium chloride
86	0.1 M Tris pH 7.5, 23.5 % PEG6000
87	0.1 M BisTris pH 6.5, 22.8 PEG8000
88	0.1 M Bicine pH 9.5, 11.2 % PEG10000, 0.11 M Sodium bromide
89	0.1 M MOPS pH 6.5, 2.4 M Sodium chloride, 0.07 M Calcium chloride
90	0.1 M Tris pH 8.0, 15.5 % PEGMME5000, 0.13 M Ammonium phosphate
91	0.1 M HEPES pH 8.0, 1.8 M Sodium potassium phosphate
92	0.1 M Sodium citrate pH 4.5, 30.3 % PEG1500, 0.20 M Ammonium sulfate
93	0.1 M CHES pH 9.0, 0.70 M Sodium citrate, 0.23 M Sodium chloride
94	0.1 M HEPES pH 8.0, 2.7 M Sodium chloride, 0.08 M Calcium chloride
95	0.1 M HEPES pH 7.5, 2.2 M Sodium malonate, 0.30 M Potassium chloride
96	0.1 M Sodium acetate pH 5.5, 29.8 % PEG1500

E.2 Stochastic Screen 20

- generated in-house

No.	Condition
1	2.4 M Sodium acetate, 0.11 M Zinc acetate, 0.06 % LDAO
2	0.1 M Sodium acetate pH 5.5, 14.7 % Isopropanol, 0.29 M Magnesium acetate
3	0.1 M MES pH 6.0, 28.8 % PEG3350, 0.04 M Calcium chloride
4	0.1 M Bicine pH 9.0, 23.4 % PEG4000, 0.15 M Ammonium acetate, 1.0 % MPD
5	0.1 M MES pH 6.0, 22.8 % PEGMME2000, 0.13 M Lithium sulfate
6	0.1 M Sodium acetate pH 5.0, 2.5 M Ammonium sulfate, 0.13 M Lithium sulfate
7	0.1 M Sodium acetate pH 5.5, 1.7 M Ammonium phosphate, 0.08 M Ammonium citrate
8	0.1 M CHES pH 9.5, 2.5 M Ammonium phosphate
9	0.1 M Bicine pH 9.0, 2.3 M Sodium malate, 0.24 M Sodium chloride
10	0.1 M MES pH 6.5, 2.2 M Sodium acetate, 0.11 M Calcium acetate
11	0.1 M Sodium acetate pH 5.0, 18.0 % PEG3350, 0.05 M Zinc chloride
12	0.1 M MES pH 60, 2.7 M Sodium chloride, 0.13 M calcium chloride
13	0.1 M Bicine pH 9.0, 2.3 M Sodium malate, 0.24 M Sodium chloride
14	0.1 M MES pH 6.5, 2.2 M Sodium acetate, 0.11 M Calcium acetate
15	0.1 M Sodium acetate pH 5.0, 18.0 % PEG3350, 0.05 M Zinc chloride
16	0.1 M MES pH 6.0, 2.7 M Sodium chloride, 0.13 M Calcium chloride
17	0.1 M Sodium citrate pH 5.5, 0.65 M Sodium citrate, 0.06 M Sodium bromide
18	0.1 M BisTris pH 6.0, 15.8 % PEG4000, 0.07 M Magnesium sulfate, 2.0 % PEG400
19	0.1 M Sodium acetate pH 5.0, 1.3 M Magnesium sulfate, 0.05 % CHAPS
20	0.1 M BisTris pH 6.5, 24.8 % PEG8000, 0.07 M Magnesium formate
21	0.1 M Bicine pH 8.5, 12.7 % PEG6000, 0.05 M Sodium acetate
22	2.3 M Sodium acetate, 0.16 M Lithium chloride
23	0.1 M CHES pH 9.0, 28.0 % PEG4000
24	0.1 M Sodium citrate pH 5.5, 36.7 % PEGMME2000, 0.03 M Sodium citrate
25	0.1 M CHES pH 9.0, 18.3 % PEG10000, 0.07 M Sodium citrate
26	0.1 M HEPES pH 7.5, 2.7 M Sodium potassium phosphate, 0.05 % CHAPS
27	0.1 M BisTris pH 6.0, 25.6 % PEGMME5000, 1.3 % Butanediol
28	0.1 M Sodium acetate pH 5.0, 25.1 % PEGMME2000, 0.17 M Potassium chloride, 0.9 % DMSO
29	0.1 M Sodium cacodylate pH 6.5, 25.9 % PEG6000, 0.08 M Lithium sulfate
30	0.1 M Sodium acetate pH 4.5, 22.6 % PEGMME2000, 0.25 M Lithium sulfate
31	0.1 M Bicine pH 9.5, 2.5 M Sodium malonate, 0.07 M Sodium potassium tartrate
32	0.1 M Sodium acetate pH 5.5, 14.8 % PEGMME5000, 0.07 M Lithium chloride
33	0.1 M Tris pH 7.5, 3.2 M Sodium chloride, 0.14 M Ammonium citrate
34	0.1 M MOPS pH 7.0, 24.5 % PEGMME550, 0.24 M Ammonium phosphate
35	0.1 M HEPES pH 8.0, 1.7 M Magnesium sulfate, 0.18 M Potassium nitrate
36	0.1 M Bicine pH 9.5, 19.4 % Ethanol, 0.29 M Magnesium chloride
37	0.1 M Tris pH 7.5, 34.2 % PEGMME2000
38	0.1 M Sodium acetate pH 5.0, 21.7 % PEG3350, 0.11 M Zinc acetate
39	0.1 M Sodium citrate pH 4.5, 1.6 M Sodium formate
40	0.1 M BisTris pH 6.0, 2.7 M Ammonium phosphate, 0.08 M Ammonium sulfate
41	0.1 M MOPS pH 7.0, 32.0 % PEGMME550, 0.27 M Sodium potassium phosphate, 2.8 % Glycerol
42	0.1 M MOPS pH 7.0, 53.3 % MPD, 0.8 % Methanol
43	0.1 M Sodium citrate pH 4.5, 18.4 % PEGMME5000, 2.4 % PEGMME350
44	0.1 M Sodium cacodylate pH 7.0, 2.3 M Sodium malate
45	0.1 M MOPS pH 6.5, 16.9 % PEGMME5000, 0.08 M Zinc acetate
46	0.1 M MES pH 6.0, 2.1 M Sodium formate, 0.05 M Sodium bromide
47	0.1 M CHES pH 9.5, 2.1 M Sodium malate
48	0.1 M CHES pH 9.5, 32.3 % PEG1500, 0.08 M Sodium bromide, 0.4 % Ethylene glycol

49	12.8 % PEG8000, 0.1 M Calcium acetate
50	0.1 M Bicine pH 8.5, 18.6 % PEG8000, 0.04 M Ammonium tartrate
51	0.1 M Tris pH 8.5, 31.8 % PEG400
52	0.1 M Tris pH 8.5, 41.0 % PEGMME550, 0.14 M Magnesium acetate
53	0.1 M HEPES pH 8.0, 14.8 % PEG4000, 0.04 M Ammonium citrate, 0.01 % LDAO
54	0.1 M BisTris pH 6.5, 2.2 M Ammonium sulfate
55	0.1 M BisTris pH 6.5, 32.2 % PEG3350, 0.09 M Sodium chloride, 7.8 mM BME
56	0.1 M Sodium acetate pH 5.0, 12.0 % Isopropanol
57	0.1 M HEPES pH 7.5, 0.72 M Sodium citrate, 0.04 M Magnesium formate, 2.9 % Ethylene glycol
58	0.1 M HEPES pH 7.5, 1.9 M Sodium potassium phosphate, 0.24 M Ammonium phosphate, 4.0 % PEGDME250
59	0.1 M Sodium acetate pH 4.5, 1.3 M Sodium tartrate
60	16.3 % Isopropanol
61	0.1 M Sodium cacodylate pH 6.5, 14.2 % PEG4000, 0.04 M Zinc acetate
62	0.1 M Sodium acetate pH 4.5, 21.8 % PEG4000, 0.08 M Lithium chloride
63	0.1 M HEPES pH 7.5, 1.3 M Sodium citrate
64	0.1 M MOPS pH 7.0, 2.0 M Sodium potassium phosphate, 0.22 M Sodium bromide
65	0.1 M BisTris pH 6.0, 1.8 M Sodium formate, 0.13 M Ammonium tartrate
66	0.1 M CHES pH 9.0, 1.3 M Magnesium sulfate, 0.2 M Magnesium acetate
67	0.1 M Bicine pH 9.5, 20.5 % PEG3350, 0.14 M Potassium thiocyanate
68	0.1 M MES pH 6.0, 19.6 % PEG6000, 0.11 M Ammonium acetate
69	1.0 M Sodium citrate, 0.11 M Sodium acetate
70	11.3 % PEG10000
71	0.1 M Sodium cacodylate pH 6.5, 2.3 % Ethanol, 0.13 M Magnesium chloride
72	0.1 M Bicine pH 9.0, 23.2 % PEG1500, 0.05 M Magnesium formate, 4.6 mM BME
73	0.1 M Sodium citrate pH 5.5, 2.7 M Sodium potassium phosphate, 0.06 M Potassium nitrate
74	0.1 M Sodium cacodylate pH 7.0, 14.7 % PEGMME5000, 0.07 M Calcium acetate
75	0.1 M Tris pH 8.5, 4.1 M Sodium chloride
76	0.1 M Sodium citrate pH 5.5, 47.3 % PEGMME550, 3.7 % Hexanediol
77	17.4 % Isopropanol, 0.1 M Sodium potassium tartrate
78	0.1 M Bicine pH 8.5, 9.3 % PEG10000, 0.26 M Lithium sulfate
79	0.1 M Sodium cacodylate pH 7.0, 25.1 % PEGMME5000, 0.09 M Ammonium tartrate
80	0.1 M Tris pH 8.0, 36.0 % PEGMME550, 0.18 M Magnesium acetate
81	0.1 M CHES pH 9.0, 38.3 % PEG400, 0.13 M Sodium bromide
82	0.1 M Sodium acetate pH 5.5, 1.9 M Sodium malonate
83	0.1 M Bicine pH 8.5, 0.90 M Sodium tartrate, 0.05 M Sodium potassium phosphate
84	0.1 M Tris pH 8.5, 2.8 M Sodium chloride, 0.1 M Zinc acetate
85	0.1 M Tris pH 8.5, 0.90 M Sodium citrate, 0.27 M Magnesium chloride
86	0.1 M BisTris pH 6.0, 31.4 % PEG1500, 0.06 M Calcium chloride
87	0.1 M HEPES pH 7.5, 9.3 % Ethanol, 0.13 M Ammonium sulfate, 0.03 % βOG
88	0.1 M Tris pH 8.0, 20.8 % PEG4000, 0.14 M Potassium chloride
89	0.1 M Sodium acetate pH 4.5, 2.2 M Sodium malate, 4.4 mM EDTA
90	0.1 M HEPES pH 8.0, 19.0 % PEG10000
91	32.0 % PEGMME2000, 0.06 M Zinc chloride, 3.5 % Hexanediol
92	0.1 M Sodium acetate pH 5.5, 20.7 % PEG1500, 0.29 M Sodium chloride
93	0.1 M Sodium citrate pH 5.0, 0.89 M Sodium citrate, 0.12 M Ammonium acetate
94	0.1 M CHES pH 9.5, 29.8 % PEG400
95	0.1 M HEPES pH 8.0, 20.4 % PEG3350, 0.26 M Sodium acetate
96	0.1 M Sodium cacodylate pH 7.0, 1.1 M Sodium tartrate, 0.13 M Ammonium citrate, 3.4 % DMSO

E.3 Stochastic Screen Pegs 4
- generated in-house

No.	Condition
1	0.1 M HEPES pH 8.0, 25.2 % PEG400
2	0.1 M Sodium acetate pH 5.0, 31.5 % PEGMME550, 0.13 M Zinc acetate
3	34.6 % PEGMME2000, 1.0 mM EDTA
4	0.1 M Sodium citrate pH 4.5, 19.3 % PEG10000, 0.07 M Magnesium formate
5	0.1 M Sodium cacodylate pH 6.5, 35.6 % PEGMME2000, 0.14 M Magnesium formate
6	0.1 M Sodium acetate pH 5.0, 48.4 % PEGMME550, 0.08 M Ammonium acetate
7	0.1 M Sodium cacodylate pH 7.0, 24.2 % PEGMME5000, 0.29 M Sodium chloride, 3.2 % PEGDME250
8	12.1 % PEG10000
9	23.1 % PEG3350, 0.05 M Calcium acetate
10	0.1 M Sodium acetate pH 5.5, 21.4 % PEG4000, 0.20 M Magnesium chloride
11	0.1 M MES pH 6.5, 32.2 % PEG1500, 0.15 M Calcium chloride
12	23.9 % PEG8000, 0.11 M Sodium potassium phosphate
13	0.1 M Tris pH 7.5, 30.2 % PEGMME550
14	0.1 M Sodium cacodylate pH 6.5, 27.2 % PEG3350
15	0.1 M Tris pH 8.0, 19.5 % PEG3350, 0.10 M Potassium nitrate
16	0.1 M HEPES pH 7.5, 21.8 % PEG1500, 0.09 M Zinc chloride
17	0.1 M Bicine pH 8.5, 17.9 % PEG10000, 0.26 M Lithium chloride, 3.9 % Ethylene glycol
18	0.1 M Sodium citrate pH 5.5, 31.8 % PEGMME550, 0.16 M Ammonium sulfate
19	0.1 M Sodium acetate pH 5.5, 24.3 % PEG400, 0.11 M Sodium chloride
20	0.1 M Sodium citrate pH 5.5, 14.6 % PEG6000, 0.08 M Lithium chloride, 1.6 % Dioxane
21	24.2 % PEG3350, 0.04 M Sodium potassium tartrate
22	0.1 M CHES pH 9.0, 14.7 % PEG6000, 0.07 M Sodium acetate
23	0.1 M BisTris pH 6.5, 19.7 % PEG1500, 0.17 M Ammonium phosphate, 4.0 % PEGMME350
24	0.1 M BisTris pH 6.0, 37.8 % PEG400, 0.08 M Potassium chloride, 0.1 % LDAO
25	0.1 M Sodium acetate pH 5.5, 24.1 % PEG8000, 0.14 M Sodium citrate, 0.01 % βOG
26	0.1 M Bicine pH 9.5, 27.8 % PEGMME2000, 0.22 M Magnesium sulfate, 2.1 % Ethylene glycol
27	0.1 M HEPES pH 7.5, 27.3 % PEG6000, 0.22 M Ammonium phosphate, 3.1 % DMSO
28	0.1 M HEPES pH 7.5, 13.1 % PEGMME5000
29	0.1 M HEPES pH 7.5, 41.8 % PEG400, 0.10 M Calcium acetate
30	0.1 M CHES pH 9.0, 21.3 % PEG8000, 0.22 M Sodium potassium phosphate
31	0.1 M CHES pH 9.5, 9.8 % PEG10000, 0.14 M Ammonium acetate
32	26.3 % PEG400, 0.15 M Sodium chloride
33	0.1 M Sodium acetate pH 4.5, 22.8 % PEGMME5000, 0.05 M Zinc chloride, 3.6 % PEG400
34	0.1 M CHES pH 9.5, 27.1 % PEG3350, 0.10 M Potassium chloride
35	0.1 M Sodium citrate pH 5.0, 30.6 % PEG3350, 0.10 M Sodium potassium phosphate, 0.8 % MPD
36	0.1 M MES pH 6.0, 30.4 % PEGMME2000, 0.06 M Sodium acetate
37	0.1 M Sodium cacodylate pH 7.0, 21.2 % PEG6000, 0.09 M Calcium chloride
38	0.1 MOPS pH 7.0, 25.2 % PEG6000, 0.15 M Potassium chloride, 1.8 % PEG400
39	0.1 M Tris pH 8.0, 29.0 % PEG3350, 0.03 M Ammonium citrate
40	0.1 M MES pH 6.0, 42.7 % PEGMME550
41	0.1 M MOPS pH 7.0, 26.7 % PEGMME2000
42	0.1 M Tris pH 8.0, 13.9 % PEG4000
43	44.0 % PEG400, 0.05 M Sodium potassium tartrate
44	0.1 M Sodium acetate pH 4.5, 47.6 % PEGMME550, 0.18 M Magnesium chloride
45	0.1 M BisTris pH 6.0, 15.6 % PEGMME5000, 0.22 M Magnesium acetate
46	0.1 M Sodium acetate pH 5.0, 36.1 % PEG1500, 0.09 M Sodium chloride

47	0.1 M MES pH 6.0, 47.0 % PEGMME550, 0.26 M Potassium thiocyanate
48	0.1 M Bicine pH 9.0, 18.5 % PEGMME5000, 1.6 % Methanol
49	0.1 M Sodium citrate pH 5.0, 22.8 % PEG4000, 0.05 M Magnesium formate
50	0.1 M Sodium cacodylate pH 7.0, 15.8 % PEG6000, 0.07 M Zinc acetate, 3.0 % Methanol
51	0.1 M Sodium citrate pH 4.5, 32.4 % PEG3350, 0.06 M Magnesium chloride
52	0.1 M Sodium acetate pH 5.0, 25.4 % PEGMME2000, 0.12 M Zinc acetate
53	0.1 M Sodium acetate pH 5.0, 18.5 % PEG4000, 0.13 M Sodium chloride
54	0.1 M Tris pH 8.0, 23.8 % PEG3350
55	0.1 M Tris pH 7.5, 25.3 % PEG8000, 0.04 M Ammonium tartrate
56	0.1 M HEPES pH 8.0, 15.9 % PEG6000, 0.03 M Calcium chloride
57	0.1 M BisTris pH 6.0, 21.3 % PEG1500, 0.17 M Ammonium sulfate, 0.08 % CHAPS
58	0.1 M Bicine pH 9.0, 16.5 % PEGMME5000
59	0.1 M Sodium cacodylate pH 7.0, 24.8 % PEG4000, 0.03 M Zinc chloride
60	0.1 M MOPS pH 7.0, 15.1 % PEG10000, 0.09 M Potassium chloride
61	0.1 M Bicine pH 9.5, 15.7 % PEG8000, 0.12 M Magnesium sulfate
62	34.8 % PEG1500, 0.25 M Lithium sulfate
63	0.1 Bicine pH 9.5, 19.3 % PEG10000
64	0.1 M Tris pH 8.5, 15.1 % PEGMME5000, 0.10 M Ammonium citrate
65	0.1 M BisTris pH 6.0, 20.5 % PEG1500, 0.28 M Lithium sulfate
66	0.1 M Sodium citrate pH 4.5, 42.8 % PEGMME550
67	0.1 M Bicine pH 8.5, 33.2 % PEGMME2000, 0.07 M Calcium chloride
68	0.1 M Sodium citrate pH 5.0, 24.6 % PEG4000, 0.05 M Ammonium sulfate
69	0.1 M HEPES pH 8.0, 14.5 % PEG10000, 0.01 % CHAPS
70	0.1 M MOPS pH 6.0, 24.3 % PEG8000, 0.07 M Zinc chloride, 0.1 % CHAPS
71	0.1 M Tris pH 8.5, 34.3 % PEGMME550, 1.4 % PEGDME250
72	0.1 M MOPS pH 7.0, 24.0 % PEG6000, 3.0 % PEGMME350
73	24.1 % PEGMME550, 0.06 M Sodium bromide, 0.5 % Hexanediol
74	0.1 M MES pH 6.0, 18.4 % PEG1500, 6.1 mM BME
75	0.1 M Sodium citrate pH 4.5, 12.9 % PEGMME5000, 0.09 M Lithium sulfate
76	0.1 M Sodium acetate pH 4.5, 55.4 % PEG400
77	0.1 M Bicine pH 8.5, 25.8 % PEGMME2000, 0.11 M Calcium acetate
78	0.1 M MES pH 6.5, 30.1 % PEG4000
79	0.1 M BisTris pH 6.0, 26.7 % PEGMME5000, 0.14 M Sodium acetate
80	23.3 % PEGMME2000, 0.12 M Sodium potassium tartrate
81	0.1 M MOPS pH 7.0, 29.6 % PEG3350, 0.08 M Ammonium phosphate
82	0.1 M Bicine pH 9.5, 29.6 % PEG10000
83	0.1 M CHES pH 9.0, 23.3 % PEG6000, 0.11 M Sodium potassium tartrate
84	0.1 M MES pH 6.5, 9.5 % PEG10000, 0.07 M Potassium thiocyanate
85	0.1 M Sodium acetate pH 5.5, 24.5 % PEG400, 0.07 M Zinc chloride, 3.4 % PEGDME250
86	0.1 M HEPES pH 7.5, 32.1 % PEG1500, 0.21 M Magnesium acetate
87	0.1 M Sodium acetate pH 5.5, 15.0 % PEG10000, 0.25 M Magnesium chloride
88	0.1 M Sodium citrate pH 5.5, 23.3 % PEG6000, 0.06 M Ammonium tartrate
89	0.1 M Bicine pH 9.5, 11.7 % PEG10000, 0.06 M Sodium citrate
90	0.1 M Sodium cacodylate pH 7.0, 22.6 % PEG3350, 0.27 M Magnesium sulfate
91	0.1 M Tris pH 8.5, 31.2 % PEG4000, 0.19 M Sodium potassium phosphate, 3.9 % Glycerol
92	0.1 M Bicine pH 9.0, 50.6 % PEG400
93	0.1 M CHES PH 9.0, 17.6 % PEG4000, 0.8 % Glycerol
94	0.1 M MES pH 6.0, 39.0 % PEGMME2000, 0.14 M Potassium nitrate
95	0.1 M CHES pH 9.5, 20.7 % PEGMME5000, 1.9 % PEGMME350
96	0.1 M Sodium acetate pH 5.5, 26.8 % PEG8000, 0.22 M Sodium bromide

E.4 Commercial Screens

JCSG+ Screen, PEGS I Suite and PEGS II Suite were purchased from Qiagen

Compositions can be found at:
http://www.qiagen.com/products/protein/crystallization/compositiontables/default.aspx

Wizard Classics Screen was purchased from Rigaku Reagents

Compositions can be found at:
http://www.rigakureagents.com/p-1-wizard-classic-crystallization-screen-series.aspx