Expanding the Biocatalytic Toolkit: Utilising Biofilms as an Adaptable Biocatalytic Platform and the Development of a General, Regioselective Pyrrole Halogenase

Michael Winn

Thesis submitted to the University of East Anglia for the degree of Doctor of Philosophy

School of Chemistry

University of East Anglia

April 2012

Abstract

Biocatalysts can offer advantages over traditional chemical methodologies in the synthesis of fine chemicals but their usage in synthesis remains rare; this can be attributed to a lack of available useful biocatalysts, limited substrate scope and poor enzyme stability under harsh chemical conditions. There is scope to develop new robust biocatalysts with novel, useful reactivities and also to improve the stability of existing enzymes.

In Nature microorganisms protect themselves from harsh conditions by forming robust multi-cellular communities known as biofilms, there is potential to use these structures as robust biocatalysts.

This thesis details the development of a novel method of engineered biofilm formation by spin-coating cultures of the biofilm producing *E. coli* strain PHL644 onto glass microscope slides. This technique proved to be fast and reproducible and compared favourably with traditional methods and also enabled detailed analysis of the maturation of the biofilm using various microscopy techniques.

Spin coating a strain of *E. coli* PHL644 that overexpressed the enzyme tryptophan synthase from *Salmonella typhimurium* enabled the assessment of the biocatalytic potential of the spin-coated biofilm to perform the PLP-dependent biotransformation reaction between haloindole and L-serine to form enantiomerically pure L-halo-tryptophan. The biofilm demonstrated improved yields and increased catalytic longevity compared to an equivalent amount of planktonic cells, cell free extract or immobilised pure enzyme.

Additionally, research was also carried out into the expression and over-production of the flavin-dependent halogenase PrnC. The ability of PrnC to regio-specifically incorporate a chlorine atom during Pyrrolnitrin biosynthesis is a feat not easily reproduced by traditional chemical techniques; therefore this enzyme represents a potentially powerful new tool for biocatalysis. The enzyme proved to be very difficult to produce in *E. coli* but numerous attempts were made to optimise production and purification, providing sufficient amounts of semi-pure protein to test enzyme activity with two substrate analogues.

Table of Contents

List of Figures

Chapter 1: Background: Utilising Biological Catalysts

Chapter 2: Engineered Biofilm Formation and Physical Analysis

Chapter 3: Performance of the Spin Coated Engineered Biofilm as an Immobilised Catalyst and Comparison to the Activity of Planktonic Bacteria and Traditionally Immobilised Enzyme

Chapter 4: Background: Natural Products and Halogenation

Chapter 5: Overexpression and Purification of PrnC

Chapter 6: Testing the Activity of Purified PrnC

List of Schemes

Chapter 1: Background: Utilising Biological Catalysts

Chapter 6: Testing the Activity of Purified PrnC

List of Tables

Chapter 1: Background: Utilising Biological Catalysts

Chapter 2: Engineered Biofilm Formation and Physical Analysis

Chapter 3: Performance of the Spin Coated Engineered Biofilm as an Immobilised Catalyst and Comparison to the Activity of Planktonic Bacteria and Traditionally Immobilised Enzyme

Chapter 5: Overexpression and Purification of PrnC

Abbreviations

$(MeO)_3B$	Methyl borate
(v/v)	Percentage volume to volume
(w/v)	Percentage weight to volume
(w/w)	Percentage weight to weight
A	Adenine
A	Absorbance
A. xylinum	*Acetobacter xylinum*
ACN	Acetonitrile
AFM	Atomic force microscopy
AHL	*N*-acyl-L-homoserine lactone
Amp	Ampicillin
ATCC	American type culture collection
B. subtilis	*Bacillus subtilis*
Bap	Biofilm Associated Protein
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
c	Concentration of sample
cAMP	Cyclic adenosine monophosphate
CAP	Catabolite activator protein
Clm	Chloramphenicol
CSLM	Confocal scanning laser microscopy
DAAO	D-amino acid oxidase
DCM	Dichloromethane
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
ε	Molar absorption coefficient (extinction coefficient)
E. coli	*Escherichia coli*
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EPS	Exopolysaccharides or Extracellular polymeric substances
ESEM	Environmental scanning electron microscopy
EtBr	Ethidium bromide
EtOH	Ethanol
FA	Formic acid
FAD	Flavin adenine dinucleotide
FORM	Formamide
FPLC	Fast protein liquid chromatography
G	Guanosine
G3P	Glyceraldehyde-3-phosphate
GFP	Green fluorescent protein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOCl	Hypochlorous acid
HPLC	High performance liquid chromatography
HPO	Haloperoxidase
IDO	Indoleamine 2,3-dioxygenase
IGP	Indole-3-glycerol phosphate

IPA	Isopropyl alcohol/Propan-2-ol
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kan	Kanamycin
kb	Kilobase
l	Path length
L. monocytogenes	*Listeria monocytogenes*
LCMS	Liquid chromatography mass spectrometry
LIC	Ligation independent cloning
MALDI	Matrix-assisted laser desorption/ionisation
MCS	Multiple cloning site
MeOH	Methanol
MMO	Methane monooxygenase
MW	Microwave
N	Newtons (SI unit of force)
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NBS	*N*-Bromosuccinimide
NEB	New England Biolabs
NRPS	Non-ribosomal peptide synthetase
NTA	Nitrilotriacetic acid
OD	Optical density
OD_{600}	Optical density at 600 nm
P. aeruginosa	*Pseudomonas aeruginosa*
PCE	Tetrachloroethylene
PCR	Polymerase chain reaction
PDA	Photo-diode array
PIA	Polysaccharide Intercellular Adhesins
PKS	Polyketide synthase
PLL	Poly-lysine
PLP	Pyridoxal phosphate
Prn	Pyrrolnitrin
r.t	Room temperature
RNA	Ribonucleic acid
Rpm	Revolutions per minute
S. aureus	*Staphylococcus aureus*
S. coeruleorubidus	*Streptomyces coeruleorubidus*
SAM	*S*-adenosyl-L-methionine
SCB	Spin coated biofilm
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
T	Thymine
TBAF	Tetra-*n*-butylammonium fluoride
t-BuLi	*tert*-Butyllithium
THF	Tetrahydrofuran
TIPS	Triisopropylsilyl
Tris	Tris(hydroxymethyl)aminomethane
U	Enzyme activity unit
V. fischeri	*Vibrio fischeri*
Z. mobilis	*Zymomonas mobilis*
α-KG	α-Ketoglutarate
β-1,6-GlcNAc	β-1,6-*N*-acetylglucosamine

Acknowledgements

Firstly I would like to thank Dr Rebecca Goss for giving me the opportunity to work towards a PhD in her lab and for providing me with her support, supervision and guidance throughout my studies. The last three and a half years have been an experience that I have immensely enjoyed, despite what I may have said if you had asked me during my writing up period. I would particularly like to thank the original members of the lab, mostly since moved on, who helped me settle into my first year of research and also all of the members past and present who have made working in the lab a real pleasure. Eternal gratitude goes to those who provided good cheer (and beer) during my writing up, they know who they are.

The biofilm research would not have been possible without the assistance of Dr Andreas Tsoligkas from the University of Birmingham who was instrumental in the development, implementation and characterisation of the spin-coating technology and Dr James Bowen who performed all the beautiful AFM analysis presented in this work. I would also like to thank Dr Mark Simmons and Dr Tim Overton, also from the University of Birmingham, who provided the chemical and biochemical engineering insight that helped to drive this project forward.

For the PrnC project I have to thank Dr Sabine Grüschow whose work on a parallel project provided helpful insight as well as some of the enzyme constructs used in this study. I would also particularly like to thank Dr Abhijeet Deb Roy and Tony Abou Fayyad who very kindly took up some of their busy time synthesising substrate analogues for my enzyme activity testing. Also thanks to Dr Lionel Hill for the LCMS analysis at the John Innes Centre (JIC).

I would like to thank my parents and family for being supportive of me during this process and forgiving me for not calling home as often as I should.

And finally thanks to you, the reader, for venturing inside this tome. I hope you enjoy reading it as much as I enjoyed researching it.

Declaration

The research contained in this thesis was carried out in the School of Chemical Sciences at the University of East Anglia between October 2008 and February 2012. The experimental work contained is original, and was carried out independently, unless otherwise stated.

Chapter 1:

Background: Utilising Biological Catalysts

1.1 The role of biological catalysts in chemical synthesis

Organisms may be thought of as biological factories, capable of synthesising all manner of complicated products from the simplest of starting materials. The idea of harnessing this intrinsic manufacturing ability of Nature is not a new one; microbial cells and enzymes have been used in the production of foods since ancient times. However it has only been in the last few decades that the potential of biology has been explored in the arena of chemical synthesis.

Over 3000 different enzymes have been identified to date[1] and many of them have already been exploited for organic synthesis. Most common examples include the use of acylases and lipases in the production of optically pure amino acids and alcohols from racemic mixtures and the production of bulk chemicals such as the nitrile hydratase catalysed formation of acrylamide from acrylonitrile.[1] The discovery that some enzymes such as lipases can be used with organic solvents and in solid phase has further encouraged the integration of such biological catalysts into the synthetic chemists tool kit.[2]

Biocatalysts can offer several advantages over traditional chemical methodologies. Evolution has optimised enzymes to perform a specific task and therefore they can act as very efficient catalysts with associated rate enhancements (as much as 10^{12}).[3] There is also an increasing trend in the chemical industry towards greener, more environmentally benign processes and enzymes provide the facility to perform complex reactions under mild, often aqueous conditions generating waste products that require no special treatment; this contrasts with traditional chemical synthesis involving high temperatures, organic solvents and often hazardous reagents and catalysts.

The main reason behind current utilisation of biocatalysts however is the high degree of regio-, stereo- and enantio-selectivity that enzyme reactions provide. This selectivity is due to the way in which enzymes control the environment in which the reactions take place; often limiting the number of reactive options by preventing access of reagents to peripheral functional groups.[3] With traditional chemistry, selectivity relies on differing reactivity rates of functional groups and the use of tedious protecting and de-protecting steps to block reagent access to sites prone to reaction.[4]

The disadvantages of enzymes include their instability when isolated as activity very often is reliant on maintaining a specific folded structure. Enzymes are highly optimised and although many exhibit

broad substrate specificity it is still hard to predict how even a well characterised enzyme will catalyse the conversion of an unnatural substrate. Enzymes such as oxidoreductases also require expensive cofactors that need to be recycled and the overproduction of enzymes and fermentation processes can involve skills not present in chemistry laboratories. Some of these problems can be overcome by using whole cells rather than isolated enzymes and by utilising directed evolution techniques to enhance activity, broaden substrate specificity and improve stability.[5, 6]

Despite widespread research the number of biocatalysts used for synthesis remains small; this can be attributed to a lack of available useful biocatalysts, limited substrate scope and operational stability.[6] It is therefore important to identify and develop robust enzymes with potentially useful applications and to improve stability of existing enzymes. Enzyme stability can be improved by increasing the structural rigidity by directed evolution and/or multipoint immobilisation (see Mateo 2007[7] and Hernandez 2011[8] for recent reviews). Problems can arise, however, when these techniques used to improve stability lead to poorer catalytic performance. Enzymes are fluid in nature and often go through rapid and precise changes in structure during catalysis. Too much conformational stability restricts this protein movement and therefore activity.[9] Enzymes tend to have optimised activity for the physiological temperature of the producing organism, balancing the conformational rigidity needed for the environment whilst allowing as much flexibility for catalysis as possible.[10] Rather than pursue directed evolution strategies to obtain stability at the cost of performance, it may perhaps be advantageous to examine robust enzyme casings instead that protect the enzyme from less ideal environments. In this arena nature has provided a method of protecting biological systems from harsh environments: the biofilm.

1.2 Biofilms

1.2.1 Biofilm characteristics

The archetypal image of bacteria is one of individual, single celled microorganisms existing in free floating, planktonic form. However in nature bacteria are predominantly found in surface-attached multicellular aggregates known as biofilms.[11] In the process of biofilm formation, microorganisms make a shift from the classic unicellular paradigm to a state that is analogous to a more complex organism where single cells make up part of a greater whole. While biofilm formation was once thought to only be in the repertoire of a few specific species, it is now considered to be a near universal trait, often with many different species co-existing within a single community. Despite the prevalence of these biofilms the actual structure and construction vary enormously with factors such

as participating species and the environment influencing formation. In spite of this diversity two characteristics have been assigned that define a biofilm:

The primary characteristic is that the participating cells are surrounded by an extracellular matrix that is produced by the cells themselves. This ECM is composed of numerous compounds which are known collectively as the EPS. The composition of this matrix varies significantly from species to species. Approximately 97% of the ECM is composed of water with the rest made up from a mix of secreted polymers, absorbed nutrients, metabolites and products from cells that have lysed.[12] Of the secreted polymers the most ubiquitous are high molecular weight sugars known as exopolysaccharides. It is from these sugars that the EPS derives its name. In addition to these polysaccharides, proteins and nucleic acids also form a part of the EPS; because of this the EPS acronym can also stand for extracellular polymeric substances. The EPS plays a significant role in protecting the enclosed cells from external threats including host defences and also acts as a diffusion barrier for toxic molecules. The high percentage of water also potentially offers a hydrating shell that protects the bacteria from desiccation.[13]

The secondary characteristic is that biofilm development is triggered by extracellular signals. These signals include environmental factors and self-produced signalling molecules called autoinducers (quorum sensing). This second biofilm attribute is just as diverse as the first with multiple pathways in different species, although similar adhesion molecules that promote biofilm formation are found in many microorganisms.[14] This process is discussed in more detail in section 1.2.3.

Biofilms in nature often exhibit vast population differences within them and not just because they are generally found as a mix of different species. Even single species biofilms that have arisen from a single cell often reveal distinct sub-populations. These differences are due to the bacterial cells adapting to growth in their individual environment and variations in the extracellular conditions can trigger differential gene expression.[15] Much of the heterogeneity within biofilms can be explained by differences in the chemical environment that exists inside them, particularly involving concentration gradients of substrates, products, oxygen or electron acceptors.[16] For example in *Pseudomonas aeruginosa* biofilms the activity of oxygen-dependent alkaline phosphatase has been used to show that oxygen only penetrates the outer regions of the biofilm.[17] This lack of internal available oxygen is not just caused by poor diffusion. The matrix is mostly made of water and it has been shown that oxygen can diffuse through this at a rate that is 60% of the rate in pure water.[18] The lack of oxygen in the lower levels is due to the ratio between the amount of oxygen diffusing through the biofilm and the rate that it is being respired by cells along the way. Oxygen never reaches the base because it is effectively being used up quicker than it can be replaced.[15]

In contrast to this, metabolites accumulate inside the biofilm with lower concentrations on the outside. A simple model of a facultative anaerobe where oxygen and nutrients are able to penetrate the whole biofilm predicts at least three different states that member cells can live in. The surface residing cells inhabit a niche that is replete with both oxygen and substrates and are therefore able to grow aerobically. Several layers down the oxygen starts to become depleted but substrates are available so the organism grows anaerobically. At the base both oxygen and substrates are limited and metabolites are accumulated. These cells are unable to exist normally and would become dormant or begin to die.[15] Planktonically grown aerobic cultures are constantly provided with oxygen; in this case it is the substrate that becomes limited during stationary phase. This high oxygen, low substrate state is not present inside the biofilm and has been suggested to explain why planktonic cells exhibit a different range of phenotypes than biofilm entrapped cells (also in a pseudo stationary phase).[15] In a similar fashion, observing DNA and protein synthesis in *Staphylococcus aureus* biofilms grown on agar led to the discovery that in this system metabolic activity was occurring only at the surface and the base of the biofilm where the cells were exposed to either oxygen or nutrients. The study indicated that roughly two thirds of the biofilm was metabolically inactive.[19] It is these micro-environments inside the biofilm (which are likely to be much more complicated in practise) that give rise to the observed high levels of biofilm heterogeneity.

1.2.2 General models of biofilm formation

To facilitate the general modelling of biofilm formation, microorganisms can be grouped into two classes based on their planktonic state: motile and non-motile. These categories differentiate between bacteria that are capable of self propulsion (through propeller like structures such as flagella) and those that rely on passive processes like Brownian motion. Separate models exist that generalise how these two classes of bacteria form biofilms.

Biofilms form at the interface of surfaces and initial bacterial attachment to this surface is the essential first step in forming a biofilm. In the case of non-motile bacteria movement speed and direction is controlled by external factors, therefore when a bacteria comes across a surface it needs to act fast if it wants to stick. Hence when conditions are right for biofilm formation, individual bacteria increase the expression of adhesion molecules, such as fimbriae, on the cell surface. These hair-like structures are short pili fibres that can promote adhesion to either a surface or to other bacterial cells already immobilised. By pre-expressing these surface features the bacteria cell stands an increased chance of adhesion upon contact with a surface or other surface bound organisms (see figure 1.1). An example of this can be found in some staphylococcal species where a large

extracellular protein known as Bap (biofilm associated protein) promotes cell-cell and cell-surface adhesion.[20] This protein is composed of repeating recombinogenic domains which can result in the production of variable length protein chains within a population. Homologues of Bap have been found in many species of bacteria. It has been suggested that due to the presence of dimerisation domains, Bap proteins from different bacteria can dimerize together and therefore promote cell-cell attachment.[21] Presumably the variation in chain length enables a cell to form as short as possible (therefore strongest) connections with close neighbours and those further away. Once bound together non-motile bacteria produce the EPS components that surround the cells. In staphylococcal species, polysaccharides known as PIA (Polysaccharide interceullar adhesins composed of β-1,6-linked *N*-acetylglucosamines) make up a large percentage of the matrix and mutations in the *ica* operon that encodes for these sugars leads to cells that are biofilm deficient.[21]

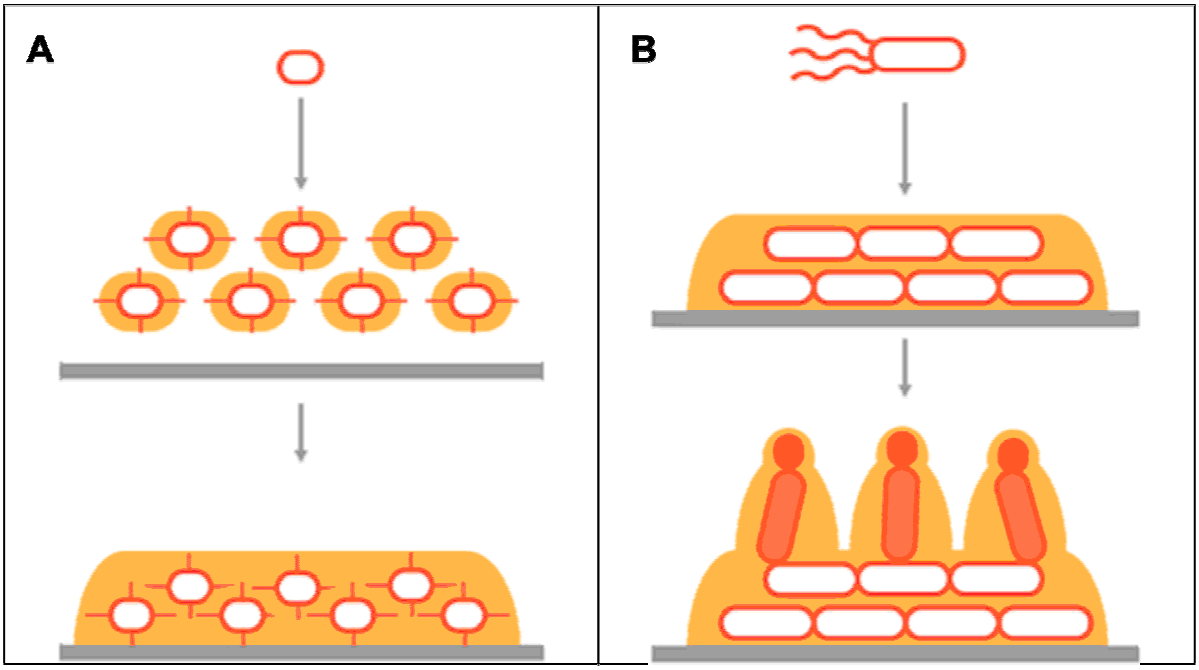


Figure 1.1: Cartoons illustrating the two different models of biofilm formation for non-motile (A) and motile (B) bacteria. (A) When conditions are right for biofilm formation, non-motile cells increase the expression of adhesion molecules on the cell surface and begin to produce EPS components. Upon contact with a surface the adhesins facilitate attachment to either the surface or other surface attached cells. EPS components surround the community in an ECM. **(B)** Motile bacteria actively travel to a surface. Once attached some motility is lost and adhesins and EPS components are made. Continued twitching motility (a tethering and dragging movement driven by type IV pili) can give rise to mushroom shaped colonies and water channels. Changes in local conditions also lead to phenotypic diversification. Figure adapted from Lemon *et al.* 2008.[22]

Motile bacteria do not have to rely on chance guiding them to a surface. Locomotive structures such as flagella allow them to move directly to a surface. Once localised adjacent to a surface, motility is lost and ECM components are synthesised. The initial surface attachment is recognised as the first of five distinct stages of motile bacterial biofilm formation.[22] At this stage the process is still reversible and the cell can either return to a planktonic state or switch to a biofilm mode. Once sufficient bacteria have attached to a surface a monolayer is formed, this is described as the second stage. The third stage involves migration of colonies within the monolayer to form multi-layered microcolonies. The fourth stage is marked by the production of ECM components. Finally a mature biofilm forms with characteristic three dimensional structures including the distinctive mushroom shaped colonies that are a hallmark of biofilms. This is a generalised view of biofilm formation and it is possible for these steps to occur in a different order or simultaneously. For example ECM production usually occurs throughout all the stages following the initial attachment.

The locomotion that flagella provide is significant to biofilm formation. Flagella negative mutants of *Listeria monocytogenes* are defective in biofilm formation.[23] The action of the flagella is thought to enable bacteria to overcome repulsive electrostatic and hydrodynamic forces found around surfaces in order to make that initial surface connection. Initial surface attachment of these mutants can be restored by gently pushing the cells toward the surface *via* centrifugation but full biofilm maturation is not seen, suggesting that motility not only influences initial surface attachment but also maturation.[23] Computer models have also shown that motility plays an important role in the formation of the three dimensional morphology within a biofilm.[24] The models revealed patterns also observed in real life *Pseudomonas aeruginosa* biofilms and illustrate that non-motile bacteria tend to form round colonies whilst motile cells form flat, spread out biofilms. The formation of the distinctive mushroom shaped structures are thought to offer motile cells an advantage by becoming less affected by mass transfer limitations. The emergence of channels within the surface of a biofilm allows the penetration of resources (including water and oxygen) down into the lower echelons of the film. The formation of mushroom colonies are driven by substrate concentrations with substrate limited bacteria detaching from the biofilm and migrating to an area with higher resources before reattaching. The driving force of this movement seems to be twitching motility, a process driven by type IV pili fibres, where cells use the small hair-like appendages on the cell surface to drag themselves forward, although the flagella may also play a role.[24] Studies with *E. coli* have provided additional evidence for the importance of motility for biofilm maturation. Indole is a common interspecies signalling molecule and has been shown to repress cell motility in *E. coli*. Biofilms formed in the presence of indole show a shift from the scattered towers seen in its absence to much flatter colonies with none of the three dimensional architecture.[25]

1.2.3 Role of signalling in biofilm formation

The decision to change from a planktonic lifestyle to a biofilm involves the activation of multiple metabolic pathways and therefore is not a trivial one. The decision can be influenced by environmental factors but also by signals produced by other bacteria. These signals are called autoinducers and at high concentrations can trigger signal transduction cascades that control a variety of developmental processes including biofilm formation.[26] This process of cell-cell signalling is known as quorum sensing and in some ways confuses the distinction between prokaryotes and eukaryotes as it enables bacteria to make decisions as a multi-cellular organism on a population wide scale.[27]

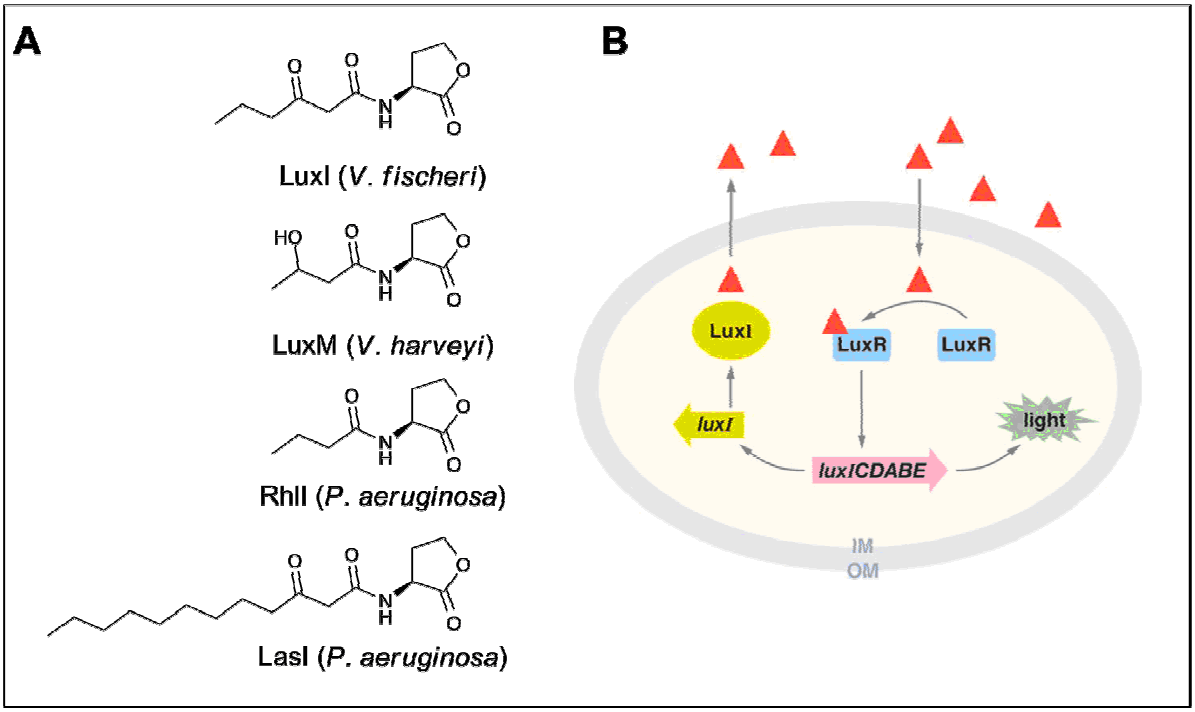


Figure 1.2: Examples of *N*-acyl-L-homoserine lactone signalling molecules from different species of gram negative bacteria and diagram illustrating an example pathway from *Vibrio fischeri*. (A) Structures of different AHLs vary by the length of the side chain. **(B)** Example of the LuxI mediated quorum sensing circuit from *V. fischeri.* The AHL produced by LuxI diffuses out of the cell. AHL (red triangles) external concentration increases with increased cell density. At critical concentration AHL binds to LuxR and transcriptional activation of the operon proceeds. LuxR-AHL also induces expression of *luxI* for positive feedback of the quorum sensing signal. Diagram taken from Waters (2005).[27]

In many gram-negative organisms, such as *P. aeruginosa,* quorum sensing is mediated by a class of compounds known as *N*-acyl-L-homoserine lactones (AHLs). Many species produce compounds in this family but variations in the chain length provide signals that are species specific. Some example compounds can be seen in figure 1.2A.

Quorum Sensing was first observed with the bioluminescent marine bacterium *Vibrio fischeri* and the mechanism of this system provides the classic example of quorum sensing which can be applied to most gram-negative bacteria (figure 1.2).[28] This species of bacteria colonise the light organ of the Hawaiian Bobtailed squid. In this organ they grow to high cell densities and induce the expression of genes required for bioluminescence. In a case of symbiosis the squid uses the counter-illumination to mask its shadow, while the nutrient rich organ allows the bacteria to proliferate in numbers unachievable in sea water. This bacterium contains the operon for the luminescent protein luciferase (*lux*ICDABE). Expression of this operon is controlled by two proteins: LuxI (autoinducer synthase) and LuxR (cytoplasmic autoinducer receptor/DNA binding transcriptional activator). The AHL produced by LuxI diffuses out of the cell and AHL extracellular concentration increases with increased cell density. At a specific critical concentration AHL binds to LuxR and transcriptional activation of the operon proceeds, leading to bioluminescence. The LuxR-AHL complex also induces expression of *luxI* in a positive feedback mechanism which produces more signal.[27]

Crystal structures obtained of LuxR type proteins suggest that they possess a specific acyl-binding pocket that only permits each LuxR to bind and be activated by its cognate signal.[29] Similarly the AHL synthases possess an acyl binding pocket that fits a particular side chain. The specificity of each LuxI and each LuxR means that in a mixed species environment, with multiple AHL signals, each species can distinguish and respond to the correct signals.[30]

P. aeruginosa possesses two quorum sensing systems known as *las* and *rhl*. Both these systems have their own AHL synthase (LasI and RhlI) and transcriptional regulator (LasR and RhlR) and are both involved in cell adhesion, biofilm formation and virulence factor expression. Expression of the RhlI synthase is activated by the Las pathway, meaning that this must occur before Rhl pathway can be activated. This produces a temporally ordered sequence of gene expression that may be critical for ordering of early and late events required for successful infection (see figure 1.3). Analysis of gene expression demonstrates that some genes respond to one of the inducers while some require both.[31] Both of these quorum sensing systems have been shown to be important to biofilm matrix production and structure. Deletion of the las pathway leads to flat, unstructured biofilm[32] while deletion of the rhl pathway makes *P. aeruginosa* biofilms less resistant to antibiotic treatments.[33]

Quorum sensing through the AHL pathway has been shown to be involved in the biofilm forming ability of many species of gram negative bacteria. A summary of different species and the effect of deletions in quorum sensing pathways can be seen in table 1.1. As is the case with *P. aeruginosa* most AHL quorum sensing pathways in gram negative bacteria have been found to encourage biofilm formation.

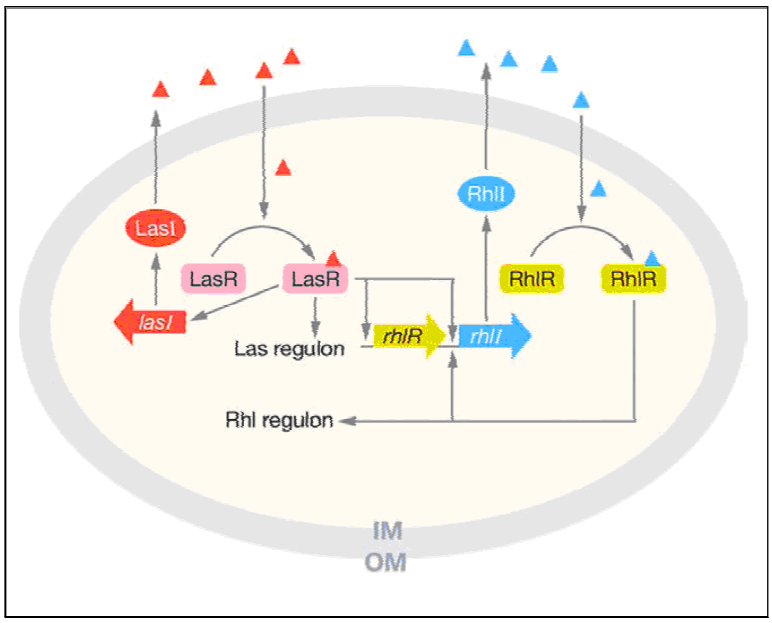


Figure 1.3: The Las and Rhl mediated quorum sensing circuit in *P. aeruginosa*. The lasI synthase produces lasI which diffuses outside of the cell (red triangles). When the extracellular concentration is great enough the AHL can bind to the transcriptional regulator LasR. As well as switching on transcription of the Las regulon, LasR also positively regulates *lasI* and induces the transcription of the *rhl* (blue triangles) pathway which functions in a similar fashion. Diagram reproduced from Waters *et al.* (2005).[27]

There are some examples however where the deletion of quorum sensing proteins leads to more robust biofilms or reduced biofilm dissociation, e.g. *Pseudomonas putida* or *Serratia marcescens* (see table 1.1). These cases illustrate that sometimes it is advantageous for cells within a biofilm to be able to control biofilm dissociation when the size of a population reaches critical mass. An example of this can be found within the quorum sensing pathway of gram positive *S. aureus.* Gram positive bacteria do not use the AHL system of signalling. Instead cyclic oligopeptides, with variable chain lengths, are used as signalling molecules (see figure 1.4). The specificity of signals and receptors functions in the same way as the LuxIR systems. Precursor peptides are synthesised inside the cell before being exported out of the cell by transporter proteins. As part of the transport process the larger precursor is cleaved to reveal the active oligopeptide which is subsequently modified to

contain lactone and thiolactone rings, lanthionine and isoprenyl groups. These peptides are not able to diffuse across the membrane like AHL signals, so the system utilises a two-component, membrane bound histidine kinase receptor to detect the quorum signal. Signal transduction is transferred via a phosphorylation cascade that influences the activity of DNA-binding transcriptional regulators which control the downstream regulon (for review see Waters 2005).[27]

Organism	Disabled AHL Pathway	Effect on Biofilm	Reference
Aeromonas hydrophila	*ahy*	Defective maturation of biofilm	Lynch et al. 2002[34]
Burkholderia cenocepacia	*cep*	More susceptible to ciprofloxacin	Huber et al. 2001[35]
	cci	More susceptible to SDS*	Tomlin et al. 2005[36]
Pseudomonas aeruginosa	*las*	Flat, unstructured biofilm. More sensitive to SDS*	Davies et al. 1998[32]
	rhl	More susceptible to tobramycin and H_2O_2	Bjarnsholt et al. 2005[33]
Pseudomonas putida	*ppu*	Formation of a more structured biofilm with distinct microcolonies and water channels.	Steidle et al. 2002[37]
Serratia liquefaciens	*swr*	Thinner biofilm, lacking cell aggregates	Labbate et al. 2004[38]
Serratia marcescens	*swr*	No biofilm dispersal	Rice et al. 2005[39]

* SDS susceptibility indicates less developed EPS structure

Table 1.1: Effect of disabled AHL quorum sensing pathways on biofilm formation. Table adapted from Irie (2008)[40]

The *S. aureus* quorum sensing system is regulated by the *Agr* operon. A diagram of the whole quorum sensing pathway can be seen in figure 1.4. The signalling peptide AIP is encoded by the *agrD* gene. There are four variations of AIP peptide (see figure 1.4) and four corresponding strains of *S. aureus* depending on which AIP they produce. The peptide product is exported and modified by AgrB and detected by the histidine kinase AgrC. This triggers the phosphorylation of AgrA which induces the production of further AgrD in a positive feedback loop similar to the AHL pathways. Phosphorylated AgrA also activates regulatory gene RNAIII. This leads to downstream effects including the repression of cell adhesion factors.[41] Through this pathway *S. aureus* biofilm formation is favoured at low cell density. Cells inside a biofilm are benign and not pathogenic. Once the concentration of bacteria has reached critical numbers for the quorum sensing pathway to activate,

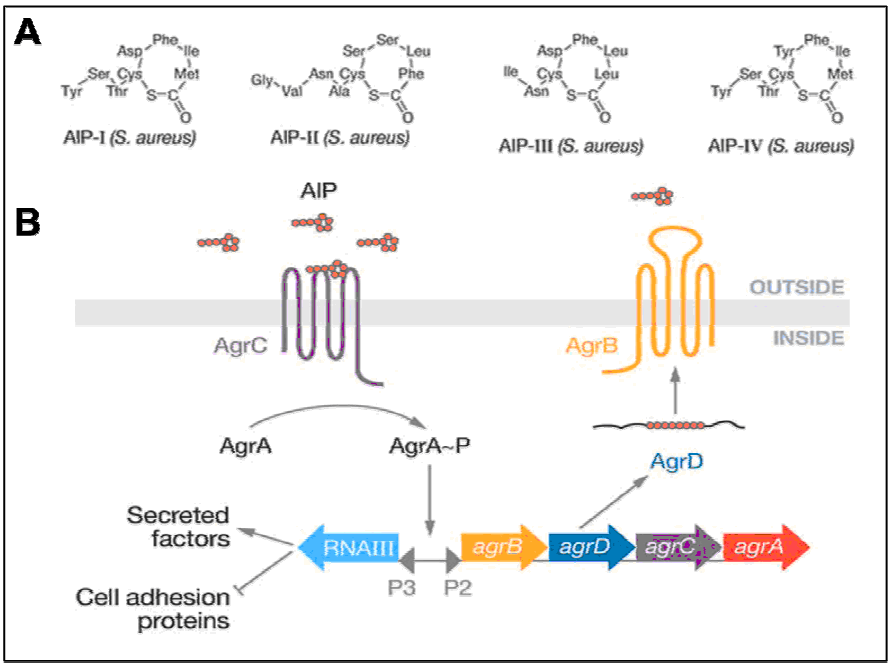


Figure 1.4. The two-component histidine kinase quorum sensing regulatory system of *S. aureus.* **(A)** Structures of the four different AIP peptides that different strains of *S. aureus* produce. **(B)** Diagram showing the sequence of events leading to activation of RNAIII. The *agrD* gene product (AIP) is transported out of the cell and modified by the AgrB exporter. Once the extracellular concentration of AIP has reached a threshold concentration a phosphorylation cascade is trigged upon AgrC binding that leads to upregulation of the Agr genes and RNAIII. Images taken from Waters 2005.[27]

the genes encoding cell adhesion proteins are switched off and cells dissociate from the biofilm. At the same time the secretion of toxins and proteases required for dissemination are switched on. The end result are vast numbers of deadly pathogens that can overwhelm a host immune system. In this case biofilm development is a way of assembling the troops for a large scale attack and quorum sensing is used as the method to control dissemination at the optimum time.

In addition to quorum sensing networks, various small molecules such as the antibiotic tobramycin have been seen to have an effect on biofilm formation (tobramycin can induce the formation of biofilms of *P. aeruginosa and E. coli*). These are very species dependent however and the pathways diverse (for review see Lopez 2010).[16]

1.2.4 *Escherichia coli* biofilms

1.2.4.1 Biofilm formation in *Escherichia coli*

Escherichia coli are gram negative rod shaped bacteria that are very commonly found in the intestinal tract. Importantly it is a well studied model organism and cultivated strains like *E. coli* K-12 are routinely used in all aspects of biology and biotechnology. *E. coli* was an original test bed for the first work on recombinant DNA and is very commonly used for the heterologous expression of proteins. Wild type strains are generally harmless and exist in the gut as components of mixed species biofilms. Attenuated lab strains however have lost the ability to form biofilms.

Wild type *E. coli* is a motile bacterium and therefore initial adhesion is dependent on flagellum. This dependency can however be circumvented. In a non-motile strain of *E. coli* K-12 (PHL644) that over-expresses curli surface adhesin fibres, it has been shown that flagella are dispensable for initial attachment and biofilm maturation.[42] Curli fimbriae are heteropolymeric proteinaceous fibres that are known to play a role in cell adhesion. These fibres aggregate at the cell surface to form structures with diameters in the range of 6-12 nm. They can reach up to 1 µm in length. As well as being demonstrated to play a role in attaching ECM components such as fibronectin, laminin and plasminogen (which enables *E. coli* to adhere to human cells)[43], they have also been shown to promote cell-cell and cell-surface interactions.[42]

The genes encoding production of curli are present in two operons. The *csgBA* operon encodes the structural components such as curlin (major element of curli) while the *csgDEFG* operon codes for the transcriptional regulator csgD and the apparatus to export the fibres to the exterior of the cell (csgE-G).[44] Expression of these genes is controlled by the OmpR/EnvZ two-component regulatory system which is responsible for sensing the osmolarity of the environment. Vidal and co-workers observed that, although K-12 *E. coli* does not generally form biofilms, when grown in continuous culture mutants eventually emerge that visibly adhere to the wall of the culture vessel. It was from one of these strains that the PHL644 mutant was developed.[42] It was established that this curli over-producing mutant contained a single point mutation (L43R) in the OmpR gene (OmpR234). This mutation improves the binding of the regulator to the *csgD* promoter region and up regulates the transcription of csgD. The OmpR234 mutation switches on curli production at much higher levels of osmolarity than with the normal laboratory strains of *E. coli.* With the normal strains the osmolarity of the environment has to be exceedingly low before the curli genes are switched on and as a consequence the fibres are never seen under lab conditions. As mentioned above the appearance of these adhesive fibres is sufficient for mutant K-12 *E. coli* strains to form thick biofilms in the absence

of flagella. Indeed even in strains with active flagella, disruption in the ability to produce these fibres results in a 50% decrease in biofilm formation.[45] Interestingly the csgD gene that codes for the curli structural genes is also involved in the production of cellulose, which forms part of the *E. coli* EPS. Like staphylococcal bacteria another common constituent of the *E. coli* EPS is β-1,6-*N*-acetylglucosamine (β-1,6-GlcNAc) which acts as an adhesin that stabilises the biofilms of many species. Treatment of *E. coli* biofilms with enzymes that degrade this polymer results in almost total dispersion, indicating the essential nature of the ECM.

Other studies have shown that *E. coli* strains bearing the conjugative plasmid IncF (increases the frequency of sex pili on the cell surface) are able to form biofilms despite being deficient in flagella, fimbriae and curli.[46] These findings seem to suggest that it is possible for motile bacteria to form biofilms in the absence of flagella as long as there are sufficient adhesins present on the cell exterior to overcome repulsive surface forces. In this case the initial attachment closely resembles the techniques employed by wild type non-motile bacteria.

1.2.4.2 Role of signalling in *Escherichia coli* biofilm formation

E. coli is not known to produce AHL signalling molecules and has no AHL synthase-like proteins in its genome.[13] It does however contain a gene, *SdiA*, which encodes a protein in the LuxR family. Knockout mutants of this gene produce three times less biofilm than the wild type equivalent.[47] The exact environmental conditions that lead to activation of SdiA are not yet known but the likelihood is that although *E. coli* cannot produce its own AHL signals it is able to detect quorum sensing molecules released by other species, perhaps allowing it to respond to signals from other bacteria in mixed species biofilms.

E. coli does however use other signalling molecules to mediated biofilm forming ability independent of quorum sensing pathways. Perhaps the most important of these is indole. This is produced inside cells in two ways. The first is derived from indole-3-glycerol phosphate during the biosynthesis of tryptophan in the α-subunit of the enzyme tryptophan synthase. This indole is consumed by the β-subunit of the same enzyme and is consequently never free in solution.[48] The second pathway produces indole during the degradation of tryptophane by the enzyme tryptophanase.[49] Indole has been shown to play a role in numerous signalling pathways[50] but specifically it also plays a somewhat ambiguous role in biofilm formation.[25] Studies using global transcriptome analyses, confocal microscopy, isogenic mutants and dual-species biofilms have shown that indole is a non-toxic signal that represses motility in *E. coli via* the autoinducer-I quorum sensing pathway controlled by the SdiA gene.[25] Motility is very important both for initial biofilm formation and for the migration of

colonies inside biofilms along concentration gradients (the causal event behind mushroom colony formation). *E. coli* biofilms grown in the lab supplemented with 500 μM indole show reduced biofilm formation and flat colonies compared to the scattered towers found without addition of indole. Conversely however it has also been demonstrated that bacterial species that possess the tryptophanase gene (breaks down tryptophan into indole) show enhanced biofilm forming ability. Di Martino and co-workers found that a strain carrying a mutation in tnaA, the gene that encodes tryptophanase, presented a decreased tendency to form biofilms, a feature that was attenuated with the addition of physiological concentrations of indole.[51] Transcriptome analysis of this mutant found repression of seven different genes associated with motility. So in this case indole seems to increase cell motility when fed to this mutant strain. Therefore there seems to be evidence that indole can both repress and enhance biofilm formation depending on concentration but in both instances it is decreased cell motility that actually leads to reduced biofilm formation.

Interestingly indole has been seen to stimulate biofilm formation in *P. aeruginosa* in mixed species biofilms with *E. coli*. Mirroring the presence of AHL receptors in *E. coli*, this species of *Pseudomonas* does not synthesise indole itself but is able to respond to the production of it in others. The fact that it seems to have the opposite reaction (enhancing biofilm formation rather than reducing it) opens up the possibility that different species can use these signals to control the population of different bacteria.[25] Signalling therefore is extremely important to biofilm formation and regulation. Understanding it holds the key to creating or combating biofilms at will.

1.2.5 Mechanisms of biofilm resistance

There has been significant research into the formation and regulation of biofilms but this has historically been motivated by the significant problems they represent to industry and medicine. Their robust nature makes them difficult to remove from surfaces and they have been implicated in the corrosion of water pipes and industrial reactor fouling.[52] Perhaps more importantly is their involvement in pathogenesis and infection which is exacerbated by their observed enhanced resistance to antibiotics. They are recognised to be involved in a plethora of medical conditions such as periodontitis, cystic fibrosis pneumonia, recurrent tonsillitis and chronic wound infections.[53] Cells within biofilms have been seen to exhibit up to a 1000 fold increase in antibiotic resistance compared to the planktonic equivalent.[54] The general mechanisms that afford biofilms with this increased resistance can be divided into two categories: innate and induced resistance.

Structural features of the biofilm itself, such as the EPS matrix, can act as an effective barrier against hostile environments and antimicrobial compounds. These resistance mechanisms are therefore

thought of as an innate attribute arising from the conversion of cells from a planktonic to a surface attached lifestyle. Although the increased robustness they bring was probably a large driving force in influencing how biofilm structure evolved in the first instance. A summary of the innate resistance mechanisms can be seen in figure 1.5.

We have already seen that oxygen and nutrient diffusion within a biofilm is limited. In the same way harmful chemicals must traverse through the EPS matrix in order to reach target cells. The decrease seen in antibiotic effectiveness with biofilms may be because less of the compound is actually reaching the cells.[54] Some studies with *P. aeruginosa* have demonstrated that some anionic compounds within the EPS (most significantly alginate) can trap cationic antimicrobial peptides, effectively delaying their diffusion.[55] The antibiotic tobramycin is trapped in this manner, although the antibiotic still penetrates the biofilm eventually.[56] These anionic EPS components can also protect biofilm cells by binding toxic heavy metals, allowing bacteria to grow in environments hostile to planktonic cells.[57] In some cases the production of EPS components like alginate have been seen to be up regulated in the presence of sub inhibitory concentrations of antibiotics which further supports the view that the EPS plays a vital role in biofilm resistance.[58] However there have been many studies that show free diffusion of antibiotics through biofilms, ciprofloxacin for example penetrates throughout *P. aeruginosa* biofilms[56] but the affect of the drug is only seen on the edge of the biofilm, therefore additional resistance mechanisms must exist. The low oxygen and nutrient concentrations within a biofilm result in internal cells with reduced metabolic activity and growth rates. Many antibiotic compounds act by disrupting cell division. If internal biofilm cells are not growing or dividing then these classes of antibiotic will have limited effectiveness.[59] This helps to explain why the affect of some antibiotics can only be observed on the surface and not lower down despite uninhibited diffusion.

There are cells within the biofilm that take slow metabolism to the extreme degree; these cells are called persisters. These are a small subpopulation of bacteria that differentiate into a dormant spore-like state that can survive even extreme antibiotic treatment. These types of cell can be found in planktonic and biofilm cultures.[54] The triggers that direct cells to become persisters usually involve a stressful condition such as starvation. Bacteria grown under optimal growth conditions have not been seen to develop persisters.[60] After receiving the signal that leads to the persister state it can take time for the cell to switch back, even if the offending signal is removed. In addition to starvation, an increase in the number of persister cells in a culture can be influenced by the amino acid starvation stringent response. This is a well known stress response in bacteria that diverts resources away from growth towards amino acid synthesis to ensure cell survival. Other factors that

encourage the formation of persisters are biofilm formation and quorum sensing pathways although the exact mechanisms of these are unclear.[60] Finally antibiotic treatment itself has been seen to trigger growth arrest and persistence. Some *E. coli* cells exposed to ciprofloxacin enter the dormant state due to the SOS response (cell cycle arrested and DNA repair boosted) that this antibiotic induces.[61]

Not all cells exposed to the same conditions will turn into persisters however. It is suggested that the decision to become persistent is a bi-stable system; essentially persister formation is either set to 'on' or 'off'. Analogous to a coin toss there are two possible outcomes and small differences in initial condition are enough to result in different outcomes at different times.[60] Research is still ongoing to identify the specific pathways leading to persister formation.

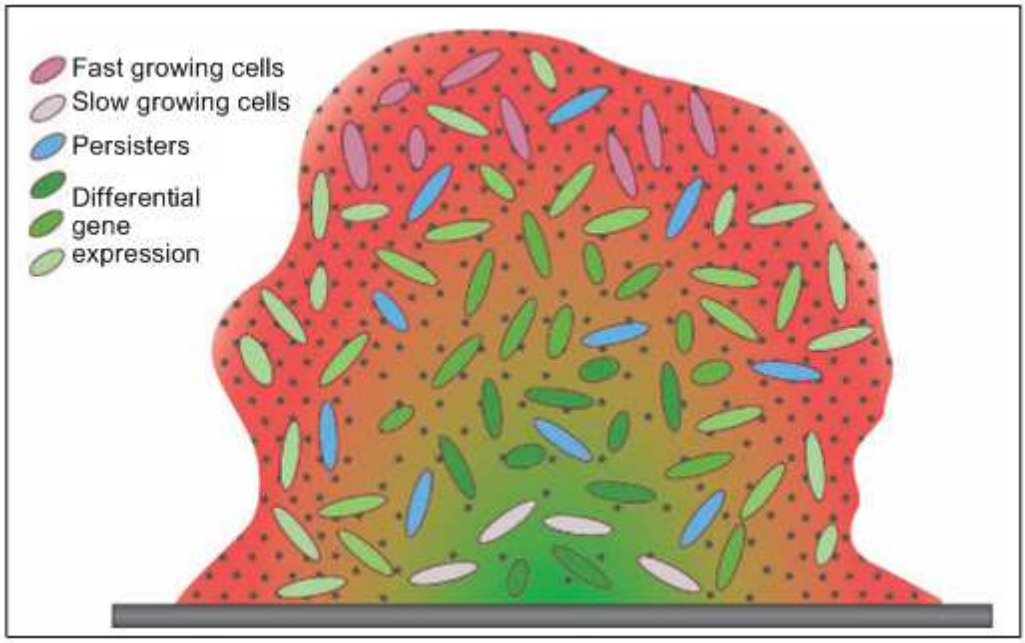


Figure 1.5: Innate biofilm resistance mechanisms. Cells within a biofilm are surrounded by an extracellular matrix (shown as coloured background). The concentration of oxygen and nutrients decreases closer to the centre (shown by colour gradient) which slows the metabolic rate of the cells. Faster growing cells are shown at the outside (dark pink) while slowly growing cells are present at the base of the biofilm (light pink). Non growing persister cells are dotted around the biofilm (blue). Different cells within the biofilm respond to the variations in conditions by displaying a range of phenotypes, including antibiotic resistance genes. Antimicrobial compounds are shown as dark spots, the concentration of these decreases as they penetrate deeper into the biofilm. Image reproduced from Anderson *et al.* (2008).[54]

As well as the innate mechanisms there are also induced resistance mechanisms that help to protect bacteria from stresses. In induced mechanisms, external factors are responsible for triggering resistance pathways. We have already seen an example of this in *P. aeruginosa* where treatment of the biofilm with imipenem (a β-lactam antibiotic) led to increased production of the EPS component alginate.[58] In another study with *P. aeruginosa*, tobramycin treatment of biofilms resulted in the upregulation of two putitive antibiotic efflux systems. Efflux pumps are one of the key mechanisms that bacteria utilise to overcome antibiotic challenges as they can remove harmful compounds from the cell. Several studies in *E. coli* have shown that efflux pumps are generally upregulated in biofilms and play such an important role that mutations in the relevant efflux pump genes show a decrease in the biofilm forming ability of those strains.[62, 63] The rapid export of harmful compounds from the inside of cells possibly also plays a role in the robustness of biofilms.

In addition to antimicrobials and toxic metal ions, the robustness of biofilms also protects constituent cells from toxic organic solvents. A biofilm composed of the gram negative bacteria *Zymomonas mobilis* has been shown to be more resistant to benzaldehyde than planktonic cells.[64] Biofilm cells demonstrated six times higher metabolic activity than planktonic equivalents after three hours of exposure to 30 mM benzaldehyde. Exposure to 50 mM benzaldehyde resulted in the complete inactivation of planktonic cells while biofilm cells still exhibited 45% residual metabolism in the same time period. A more recent study with a *Pseudomonas* species demonstrated biofilm resistance to styrene that was significantly greater than comparable planktonic cells.[65] These results indicate that while biofilm resistance is a problem in the medical arena, it may be advantageous in the field of biotechnology.

1.2.6 Biofilms as catalysts

Biocatalysis is a very effective and environmentally low impact tool for the production of industrially relevant chemicals. The family of biocatalysts includes bacteria, fungi and the assortment of enzymes that they produce. The advantages of using a biocatalyst include the potential for high regio and stereo specificity that the reactions can show and the fact that they can be performed under mild conditions. The catalysts themselves are biodegradable and waste streams are generally less toxic than traditional chemical equivalents. Sequential reactions can be performed within the same cell and therefore circumvent the need to purify intermediates and the associated financial cost that accompanies this.

Biocatalysts can be whole cells or purified enzymes. In industry whole cell catalysts predominate as they have a number of advantages over purified enzymes for large scale processes. Compared to

isolated enzymes, whole cells do not require lengthy protein purification procedures and the enzymes are generally better protected against denaturation caused by non-optimal conditions. More complicated, multi-step syntheses also require multiple enzymes and cofactors which can all be contained and recycled in single cells.[66]

Because of the obvious advantages that biocatalysts afford, there are many examples where biocatalysts have been incorporated into chemical synthesis. According to a recent review there are 125 biotransformation processes currently in use industrially and over half of these are catalysed by whole cell systems.[67]

One of the main challenges of using enzymes and whole cell systems is the toxicity that can arise due to choice of substrate and conditions that are present in many chemical reactions. Limited aqueous solubility of potential organic substrates and the toxicity of high concentrations of organic solvents to whole cells and isolated enzymes, limits the number of useful biocatalysts available to the industry. In some cases biphasic biotransformation systems have been used to great effect to solubilise substrates and the development of water immiscible ionic liquids as a less toxic alternative to organic solvents for biphasic reactions have been employed,[68, 69] but the added expensive of these solutions has limited their uptake for large scale industrial processes.

The robustness of biofilms (demonstrated with increased resistance to antibiotics) is seen as a problem in medicine but this increased resilience can be beneficial for biotechnological applications, helping the cells to survive in otherwise toxic conditions. The enhanced solvent protection that biofilms provide cells has already been discussed. In addition to their increased robustness, a biofilm also presents a method of whole cell immobilisation. Of the 66 currently utilised industrial biotransformation reactions most are run in batch mode, only seven are continuous flow processes.[67] Continuous reactor operation is significantly more cost effective than individual batch reactions. One of the currently preferred types of bioreactor for whole cells is a membrane reactor featuring a permeable membrane that is used to entrap cells but allows the passage of substrates and products. The problem with this type of reactor is the cost and high incidence of reactor fouling.[70] Alternatively the cells themselves can be entrapped within a polymer such as calcium alginate cross-linked beads but this requires intricate preparation steps that add to the cost.[71] Due to the cross links, the transfer of substrates and products through the polymer may be limited. Immobilising cells within a biofilm offers an alternative.

One area where biofilms have found a successful catalytic niche is in the bioremediation of waste water, gases and soils. Using biofilms in this manner can be traced back to the early 1880s.[70] These

days, bioremediation usually involves taking an environmental sample from a hostile environment and growing up cultures of bacteria that have adapted to living under those conditions using the logic that the organisms would have developed methods to remove or utilise the toxic products. Biofilms produced in this way are inevitably composed of mixed species. Using this approach processes have been developed to reduce the chemical oxygen demand of waste water streams or for nitrification/denitrification processes.[70] Continuing developments in bioremediation have produced systems that can process harmful industrial waste products. Tetrachloroethylene (also known as PCE) is a potentially carcinogenic solvent used as an industrial degreasing agent and in dry cleaning fluids. It is one of 14 compounds on the US Environmental Protection Agency's priority pollutant list and very strict release limits are imposed on it. Carter and co-workers demonstrated an anaerobic biofilm reactor that converted PCE into lesser chlorinated compounds including vinyl chloride at 98% efficiency to levels below that recommended in the Safe Drinking Water Act.[72] Other examples include the sequestration of selenium and arsenic from waste water using biofilms cultured from coal mining effluent[73] and the transformation of *o*-xylene into *o*-methyl-benzoic acid using a biofilm cultured from sewage sludge.[74]

In addition to the membrane bound reactor, many bioremediation processes, especially waste water treatment, use a variation of a type of reactor system called a rotating biological contactor.[70] In this process a rotating drum is covered with a thin film of biofilm and waste water passed over it. The rate of the rotation is set to a speed that allows optimum contact with the waste stream inside the particular system. The transformation of *o*-xylene and sequestration of selenium and arsenic in the previous examples both use this type of system (200 and 160 rpm respectively). The stream is recycled over the drum to ensure efficient catalysis.

The most commonly used industrial reactor system is the bed reactor (figure 1.6). In this type of system the catalyst, which could be a biofilm, is coated onto individual supports, usually spherical particles. For biofilm formation these particles are packed into the cylindrical reactor before being seeded with a bacterial culture. The flow through the reactor is set to low levels (or removed altogether) to allow the cells to attach to the surface of the supports. Depending on the type of reactor and the species used, a biofilm can take as little as a few days or as long as a couple of weeks to form.[70]

There are three main types of bed reactor that are generally used. The first type is a packed bed reactor. In this system the biofilm coated supports are packed into the reactor and the feed solution either added from the top (trickle packed bed reactor) or pumped through the supports from the base (submerged packed bed reactor). The liquid is passed through the supports at low pressure and

travels through the gaps between the spherical supports. This method has the advantage of very low sheer forces so weakly bound biofilms can be retained throughout the reaction.[67] The trickle bed reactor does present potential problems with cleaved cells fouling the output; the reverse flow of the packed bed reactor overcomes this issue.

The second type of bed reactor is the expanded bed reactor. The fundamental process is the same as a packed bed reactor but the fluid velocity passing through the supports is greater. This has the effect of forcing the particles apart, effectively increasing the active surface area available for catalysis.[67] This form of bioreactor has also seen use in several bioremediation processes such as the PCE biotransformation mentioned above.

The third type of bed reactor is referred to as a fluidised bed reactor. In this system the flow rate is increased even further which causes the supports to bounce around the reactor. This increases the catalytic surface area even more than the expanded bed reactor and therefore has the potential to reach higher productivity than the other types of bed reactor.[67] This type of the reactor is

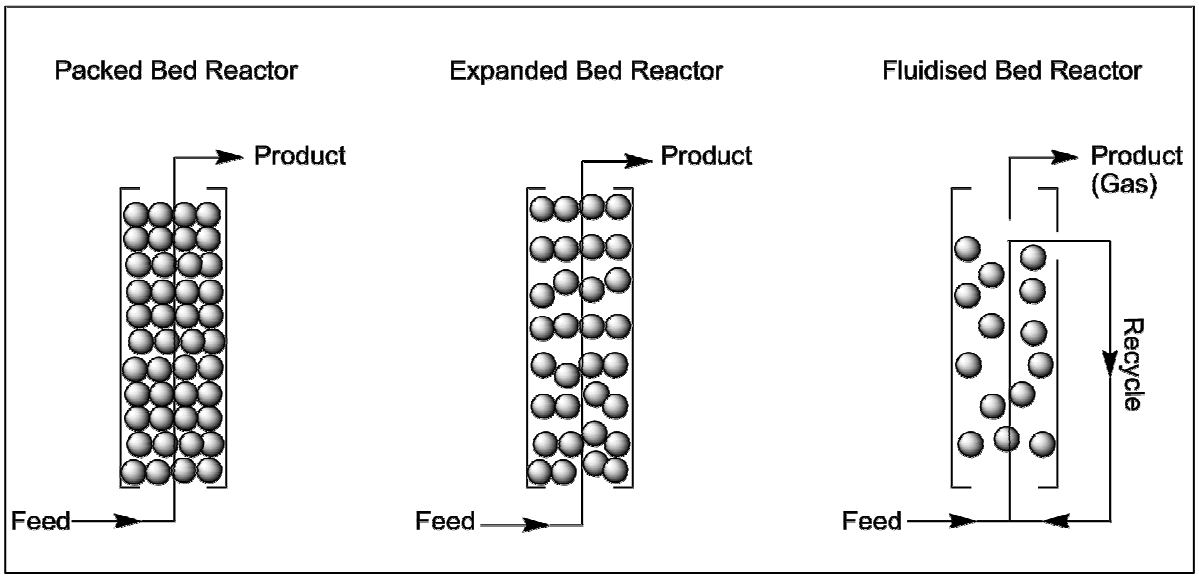


Figure 1.6: Diagrams of various types of bed bioreactor. The packed bed bioreactor contains catalyst coated spheres through which a feed stream is pumped in the upward direction. Product is then collected from the top. Expanded bed bioreactors work with a similar principle but with higher stream velocities supporting the catalysts weight and forcing them upwards. Fluidised bed reactors use a high velocity feed stream that moves the particles around in the reactor creating a greater surface area, promoting mass transfer but introducing higher sheer forces. This type of reactor is especially suited to biphasic systems where a gas is evolved as the product which can be removed from the top of the reactor. The liquid can be recycled.

particularly useful when dealing with biphasic reactions which involve liberating a gas from a liquid. The gas can be removed from the top of the reactor while it is straightforward to recycle the feed stream, as seen in figure 1.6.

As the fluid velocity increases in the expanded bed and fluidised bed, improved mass transfer of the system can be seen as the movement of the beads is analogous to stirring. However the sheer forces acting on the biofilm increase with increasing flow rate and can therefore make the formed biofilms unstable.

Apart from a few examples, for instance the production of single species *Neurospora crassa* fungal biofilms for the bioremediation of phenol,[75] bioremediation processes utilise biofilms that contain a mix of different bacterial species. Controlling biofilm formation and limiting the bacteria to a single species opens up the possibility of using immobilised whole cells in the production of targeted industrial chemicals. Many of these industrial processes take advantage of pathways already present in the biofilm organism, such as the ability of the bakers/brewer's yeast *Saccharomyces cerevisiae* to produce ethanol. For many ethanol production processes *Zymomonas mobilis* biofilms formed inside reactors has surpassed traditional yeast as the ethanol producer of choice.[76] In one example the bacteria was adsorbed onto vermiculite which was used as the support inside a biofilm attached expanded bed bioreactor. Using this method Bland *et al*. fermented glucose into ethanol and achieved a productivity of 105 g L^{-1} h^{-1} which showed enhanced operating parameters over traditional free cell reactors (< 4 g L^{-1} h^{-1}).[70, 77] Using similar methods other important biofuels have been produced including butanol from biofilms of *Clostridium acetobutylicum* and 2,3-butanediol from *Klebsiella pneumoniae* biofilms (for extensive list see Qureshi 2005).[70]

The applications of immobilised biofilms have been stretched further than simple fermentations and involve the formation of other chemicals and the involvement of more complex biotransformations. As seen earlier, *Z. mobilis* biofilms demonstrate a significantly higher tolerance to benzaldehyde than planktonic cells. This tolerance was explored using biofilms that were formed inside flow cells. Reactor technology was then exploited to produce benzyl alcohol in a continuous flow process. The cells were immobilised onto glass beads which were packed into a modified test tube. Benzaldehyde was pumped through the beads from the base, making a rudimentary packed bed bioreactor.[64] The resulting reactor was shown to produce product at a rate of 8.11 g of dry product per day with a 90% molar yield after 45 hours. Another interesting example harnesses the methane monooxygenase enzyme (MMO) from methanotrophic bacteria. This enzyme is the first step in the methane utilisation pathway in these bacteria and has a broad substrate specificity that can oxygenate a range of *n*-alkanes and *n*-alkenes. Xin and colleges demonstrated the ability of a mixed methotrophic

species biofilm to monooxygenate propene to epoxypropane at a rate of 110-150 µmol per day. This was run in an attached film fluidised bed bioreactor with the biofilm formed on diatomite particles. The reactor was operated continuously for 53 days without any loss in activity.[78] The production of these optically pure epoxides is important as they are valuable synthons in organic synthesis.[79]

All of the examples mentioned so far utilise native pathways already present in the biofilm forming species to produce a chemical or perform a biotransformation. The real potential for this technology is to engineer genetic pathways or to introduce recombinant DNA to enable the formation of biofilms capable of performing a specific programmed task. To achieve this, biofilms need to be constructed from genetically well characterised organisms. Li and co-workers screened 68 common industrial and laboratory strains of bacteria, spread over 40 genera and 5 phyla. Each was evaluated and the strains biofilm forming ability was scored. By testing various culturing conditions, 97% of the strains tested were able to form biofilms and over half were classified as having strong biofilm forming tendencies.[80] Most of these strains are genetically well characterised which shows there is huge scope for producing engineered bacterial biofilms that are capable of performing biotransformations using recombinant enzymes. Currently however there are very few examples in the literature of biotransformations performed by biofilms composed of engineered bacteria strains.

One of the bacteria screened in the study by Li *et al*. was a strain of *Pseudomonas* (VLB120) that was known for its ability to mineralize styrene through an (*S*)-styrene intermediate using the styrene monooxygenase StyAB.[80] A tubular membrane bound biofilm reactor was demonstrated by Gross *et al*. for this conversion of styrene to (*S*)-styrene oxide. The strain was engineered to include a mutation in the isomerase enzyme that is the second step in the pathway. This led to the process of styrene degradation stalling at the formation of (*S*)-styrene oxide which was the desired product. The product is formed in high enantiomeric excess (99.9%). The biofilm reactor was stable for at least 55 days at a maximum rate of 16 g of product per litre per day and a yield of 9 mol%.[81] Continuous flow processing allowed the removal of product to prevent product inhibition of the enzyme which improved the overall life of the reactor compared to batch processes.

Currently there are only two examples in the literature of totally recombinant strains of bacteria being used for biofilm mediated biotransformations. One of these utilises recombinant *E. coli* B strain KO11 which contains chromosomally integrated genes for ethanol biosynthesis (from *Z. mobilis*).[82] Planktonic cells of this strain showed lower levels of ethanol tolerance than traditional yeast however biofilm immobilised cells grown on glass micro spheres inside an expanded bed bioreactor showed improved stability. Non-immobilised cells saw a decline in ethanol yield to 60% after only 8-9 days of continuous fermentation. By contrast the biofilm immobilised *E. coli* saw a

stable conversion of >85% for at least ten days before reaction rate slowed to a still respectable 70% for at least another 40 days of fermentation.

The only other example of recombinant biofilm strains being utilised as immobilised biocatalysts was reported by Setyawati *et al*.[83] A recombinant strain of *Acetobacter xylinum* was transformed with a plasmid containing the D-amino acid oxidase (DAAO) enzyme from *Rhodosporidium toruloides*. DAAO catalyses the formation of α-keto acids from corresponding D-amino acids and is useful in the resolution of racemic amino acids, the production of amino cephalosporanic acid (a key intermediate for the production of new cephalosporin antibiotics) and for the production of the α-keto acids themselves. The resulting recombinant strain was allowed to form a biofilm-like structure. *A. xylinum* is a gram negative obligate anaerobe that can produce a thick white cellulose pellicle at air-liquid interfaces in static culture. During growth the bacterial cells cover themselves in cellulose fibres that form an inter-locking weave that surrounds the cells. Although different from traditional biofilm structure, the overall effect is the same. These formed pellicles protect the contained cells from harsh environments, retain moisture and hold the cells in an aerobic environment.

Cultures of *A. xylinum* were shown to form pellicles in static and shaking cultures. Initial evidence of cellulose pellicle formation occurred within 3 days. DAAO production was induced by addition of IPTG. 4 days after protein induction the immobilised beads were tested for DAAO activity. 100 mg of wet beads were used to convert D-alanine into the corresponding α-keto acid in a simple shaking flask system. In parallel the same biotransformation was carried out using the same weight of non-immobilised (non-pellicle) cells and the cell free extract. The activity of the immobilised cells was shown to be reduced in comparison to the cell free extract equivalent (less than 10% relative activity). The planktonic cells fared better and showed 50% activity relative to the cell extract.[83]

The pellicles were revealed to be composed of 75% cellulose EPS and this property was blamed for the comparatively bad catalytic performance. This does highlight a possible problem with biofilm mediated biocatalysis, that excessive matrix production may limit substrate access to the cells themselves.[67, 83] This is a conundrum as the matrix production is what makes biofilms so attractive to biocatalysis. The advantage is clearly seen when hostile environments or toxic substrates are used, but the benefits are less clear when conducting biotransformations under mild conditions.

Despite the advantages that biofilm biocatalysts can offer to chemical synthesis and the many demonstrated applications shown as laboratory concepts, there has not been wide uptake of this technology in industry.[67] The only real example of a successful application so far is a trickle bed reactor for the production of vinegar using biofilms composed of acetic acid bacteria. The bacteria

are immobilised onto beechwood shavings and large fermenters up to 60 000 litres have been used.[70] In 2009 it was estimated that approximately 300-400 of these reactors were in operation.[67] This leaves huge scope for the development of an effective and efficient biofilm attached bioreactor that can be shown to be of worthwhile interest to industry.

Within industry, however, it is not just the problem of enzyme stability that limits biocatalyst uptake; the number of biocatalysts capable of performing useful chemistry is still limited. The development of new and interesting biocatalysts is essential if this technology is to experience wide application in the chemical industry. One such area where biocatalysts are being developed to offer advantages over traditional chemical methodologies is biological halogenation.

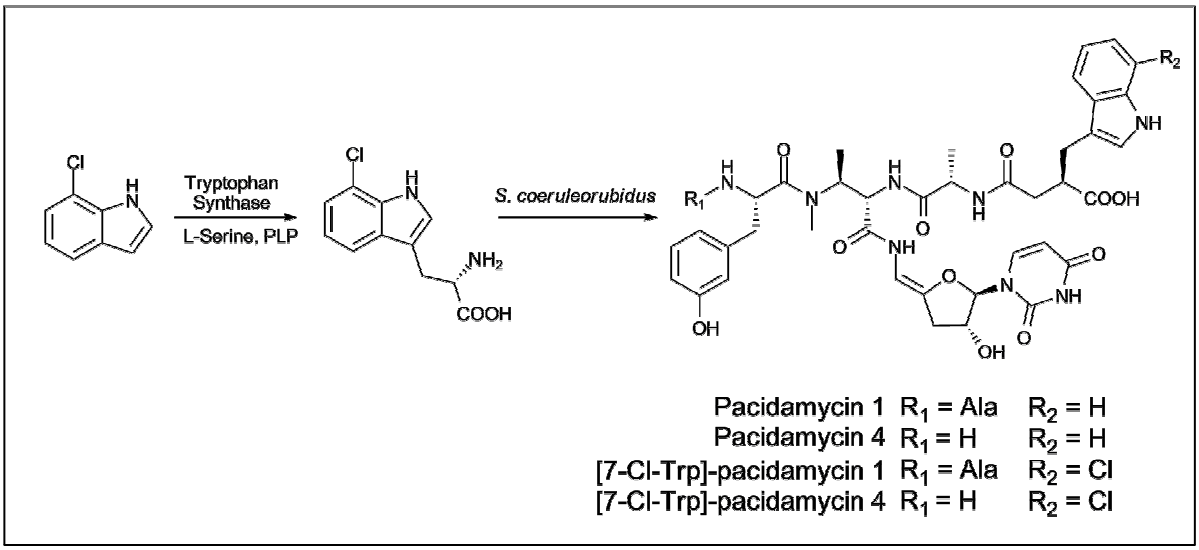
1.3 Halogenation as a route to new compounds of medicinal interest

Nature's inventory of secondary metabolites has long been mined as a source of medicinal lead compounds. In 2005 it was estimated that around half of the drugs in clinical use at the time were either natural products themselves, natural product derived or semi-synthetic natural products.[84, 85] However the discovery of new types of natural products with unique modes of action has slowed and as a result pharmaceutical companies have increasingly turned to combinatorial chemistry and diversity-orientated synthesis to generate large compound libraries.[86] These libraries are capable of generating vast numbers of potential lead compounds with significant structural diversity but this approach results in very low hit rates and the potential for increased side effects due to often less specific binding characteristics.[85] With the advent of increasingly cheap and rapid genome scanning techniques and tools for mining novel bioactive molecules, natural products are returning to favour as they are themselves the result of a natural form of high-throughput screen and often therefore possess high levels of specificity and potency.

A key problem with natural products however is that they tend to demonstrate poor physiochemical properties or fail to show levels of activity desirable for therapeutic drugs. This opens up the possibility that natural products may be modified to fine-tune activity and create a better lead compound. Instead of viewing combinatorial synthesis as an entirely separate entity, strategies are now being employed that combine both approaches to introduce semi-synthetic modifications to target compounds.[87] Within this there is large scope for incorporating novel biocatalysts.

The introduction of a halogen is a very simple change that can result in large differences to the physicochemical properties of a compound. For example, Grüschow *et al.* have shown that incorporation of a halogen into the antimicrobial compound pacidamycin increases its lipophilicity which can affect activity.[88] Pacidamycins are a family of nucleoside antibiotics that inhibit the

function of translocase I (an enzyme involved in bacterial cell wall assembly) using a clinically unexploited mode of action.[89] Members of this group of antibiotics differ in the N and C-terminal residues; some contain a tryptophan residue at the N-terminus while others have a 3-hydroxyphenyl ring (meta-tyrosine). The C-terminus either contains alanine, glycine or hydrogen (see scheme 1.1). Grüschow and co-workers were able to show that the incorporation of a single halogen at the 7 position of the N-terminal tryptophan residue contained within pacidamycin 1 and 4 (the most abundant tryptophan containing pacidamycin analogues) resulted in altered cytotoxic properties.[88] The incorporation was made possible by using precursor directed biosynthesis and biotransformation techniques. Tryptophan chlorinated at the 7-position was produced from commercially available 7-chloroindole using a biotransformation catalysed by the enzyme tryptophan synthase (see scheme 1.1). The chlorinated tryptophan was then fed into the growth media of the producing organism (*Streptomyces coeruleorubidus)* at resting phase where it showed good incorporation into pacidamycin 1 and 4 (scheme 1.1). The resulting 7-chloro-tryptophan analogue of pacidamycin 1 proved to have an MIC of 32 μg ml^{-1} which was four times lower than the parent compound. Conversely chlorinated pacidamycin 4 was twofold less active than the parent compound. This illustrates the profound effect that a single chlorine atom can have on the activity of a natural product, even if that effect is not consistent or predictable. This also demonstrates how the enzyme tryptophan synthase can be utilised to produce useful tryptophan analogues that may be harder to produce using traditional chemistry.



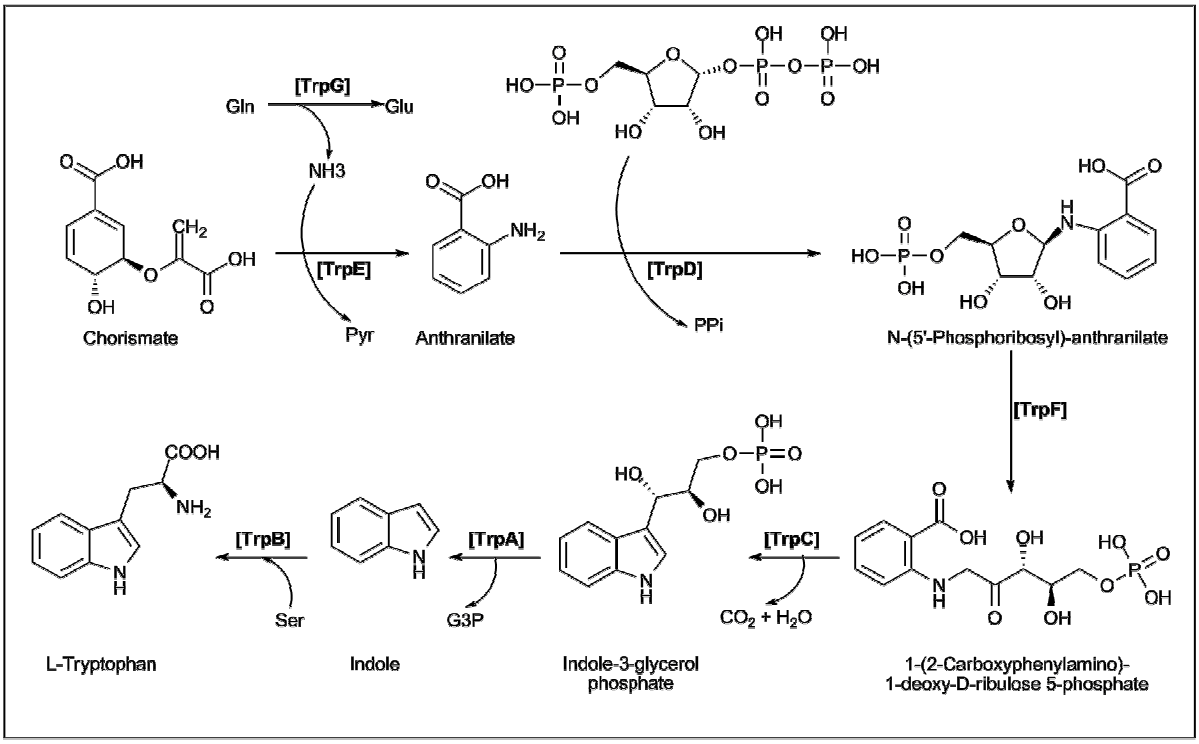
Scheme 1.1: Precursor-directed biosynthesis of chlorinated pacidamycin derivatives. 7-chloro-tryptophan produced from 7-chloroindole by a tryptophan synthase mediated biotransformation. Feeding to pacidamycin producer *S. coeruleorubidus* results in some chlorinated pacidamycin 1 and 4. [88]

1.4 Tryptophan synthase

1.4.1 Tryptophan synthesis

Synthesising tryptophan analogues using traditional organic synthetic methods is far from straightforward. The first synthesis of a chlorinated tryptophan analogue (7-chlorotryptophan) was published in 1955[90] and even in the years since the standard synthetic methods of producing modified tryptophan analogues often involve multi-step processes.[91] Obtaining tryptophan that is enantiomerically pure is also difficult to achieve by separating out the individual diastereoisomers,[92] although more recently a biocatalytic approach to this problem has been demonstrated by an enzymatic optical resolution using an enantiospecific acylase enzyme.[93] A far simpler procedure involves harnessing the activity of the enzyme tryptophan synthase, the enzyme responsible for the final stages of L-tryptophan biosynthesis.

The facility for tryptophan biosynthesis is missing in mammals but is available to prokaryotes, eukaryotic microorganisms and higher plants.[94] Tryptophan biosynthesis is controlled by 7 catalytic modules (scheme 1.2). In *Bacillus subtilis* these seven protein domains are encoded by separate genes (*trpA*-*G*),[94] while in other organisms, such as *E. coli,* several of these domains are encoded within a single gene (*E. coli trpD* for instance encodes the equivalent of the separate *trpG* and *trpD* genes from *B. Subtilis)*. The role of these domains can be seen in scheme 1.2. The starting point for the biosynthesis is chorismate which is converted into anthranilate by the TrpE anthranilate synthase. The reaction proceeds with the addition of ammonia (released from glutamaic acid by TrpG aminase) and subsequent loss of pyruvate. The gene product of *trpD* (anthranilate phosporibosyl transferase) catalyses the formation of *N*-(5-Phosphoribosyl)-anthranilate from anthranilate and phosphoribosyl pyrophosphate which is then isomerised into the ribulose form (1-(2-carboxyphenylamino)-1-deoxy-D-ribulose 5-phospate) by TrpF. Indoleglycerol phosphate synthase (TrpC) then mediates the indole ring formation with loss of water and CO_2 that results in indole-3-glycerol phosphate. This acts as the substrate for the gene product of *trpA* which forms indole with the loss of glyceraldehyde-3-phosphate. Indole is then combined with serine to form tryptophan, catalysed by *trpB*. Tryptophan synthase consists of the *trpA* and the *trpB* catalytic modules.



Scheme 1.2: Biochemical pathway of tryptophan biosynthesis from *Bacillus subtilis*. Figure modified from Xie *et al.*[94]

The tryptophan synthase enzyme is made up of two separate subunits; the alpha subunit (encoded by *trpA*) and the beta subunit (encoded by *trpB*).The best characterised and mostly widely used tryptophan synthase variant comes from *Salmonella typhimurium* and is found as a tetramer of two alpha and two beta subunits ($\alpha_2\beta_2$) arranged in an elongated αββα structure with the two beta subunits forming an interface between the two functional αβ-dimers with the alpha subunits on the outside.[95]

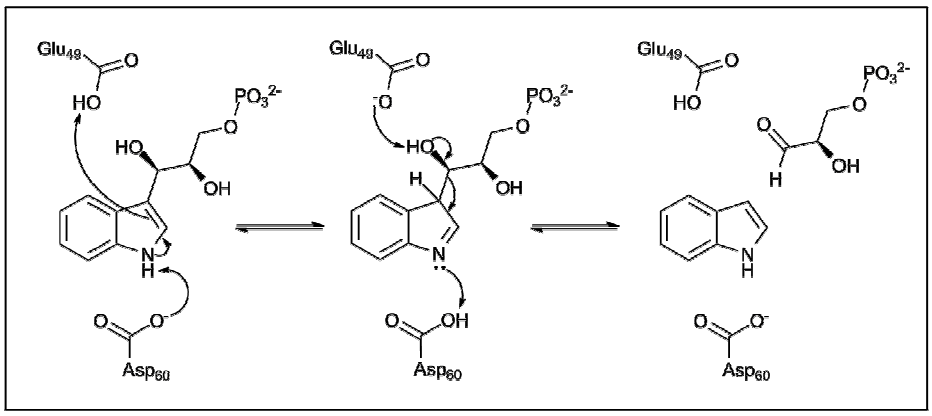
Tryptophan production is highly regulated using a negative repressive feedback mechanism. The gene *trpR* encodes for a repressor protein which is continuously expressed at low levels. When bound to tryptophan, this repressor protein forms a conformation that allows it to bind to the tryptophan operator, effectively blocking the association of the RNA polymerase and therefore blocking tryptophan production when tryptophan is in excess.[96] To ensure full repression an attenuation method is also present mediated by a leader transcript sequence at the beginning of the trp operon which contains two tryptophan residues. In the presence of tryptophan this leader sequence is complete and a transcriptional terminator hairpin loop forms which promotes the dissocation of the RNA polymerase and therefore prevents transcription of the trp operon. If the concentration of tryptophan in the cell is low then the ribosome stalls while trying to translate the

sequence. This promotes the formation of an anti-termination hairloop which allows the RNA polymerase to transcribe the tryptophan operon in full.[97]

1.4.2 The α-subunit reaction

The α-subunit is responsible for preparing the indole substrate for the β-subunit. It catalyses the reversible retro-aldol cleavage of indole-3-glycerol phosphate (IGP) to yield indole and glyceraldehyde-3-phosphate (G3P) (scheme 1.3). The active site of the α-subunit was located by binding the competitive inhibitor indole-3-propanol phosphate to the crystalline enzyme.[98] Site directed mutagenesis studies on the active site of the α-subunit identified two key catalytic residues that were important for activity; aspartic acid 60 and glutamic acid 49.[98] Crystal structures resulting from the same study also showed these residues were located near the scissile bond and in a position suitable for catalytic groups. Replacement of aspartic acid 60 and glutamic acid 49 with alanine led to loss of activity. Replacement of aspartic acid 60 with a glutamic acid residue showed some restoration of activity, therefore a general acid-base catalysed retro-aldol cleavage mechanism, as can be seen in scheme 1.3, was suggested.[98]

The cleavage of the C3'-C3 bond in IGP is driven by the tautomerisation of the indole ring to the indolenine. This process is performed by the two catalytic residues; Glu49 protonates the indole ring while Asp60 abstracts the proton from N1 of the indole. The cleavage is completed by deprotonation of the C3' hydroxyl group of the glycerol phosphate by Glu49 to release G3P and indole.[98, 99]



Scheme 1.3: Mechanism of the α-reaction of tryptophan synthase. General acid-base catalysis is powered by two important residues in the active site αGlu49 and αAsp60. Adapted from Phillips 2004.

The released G3P is held in the active site to prevent indole from escaping into solution. Instead the indole is transferred intramolecularly down a 25Å tunnel into the β-subunit active site for the next part of the process.

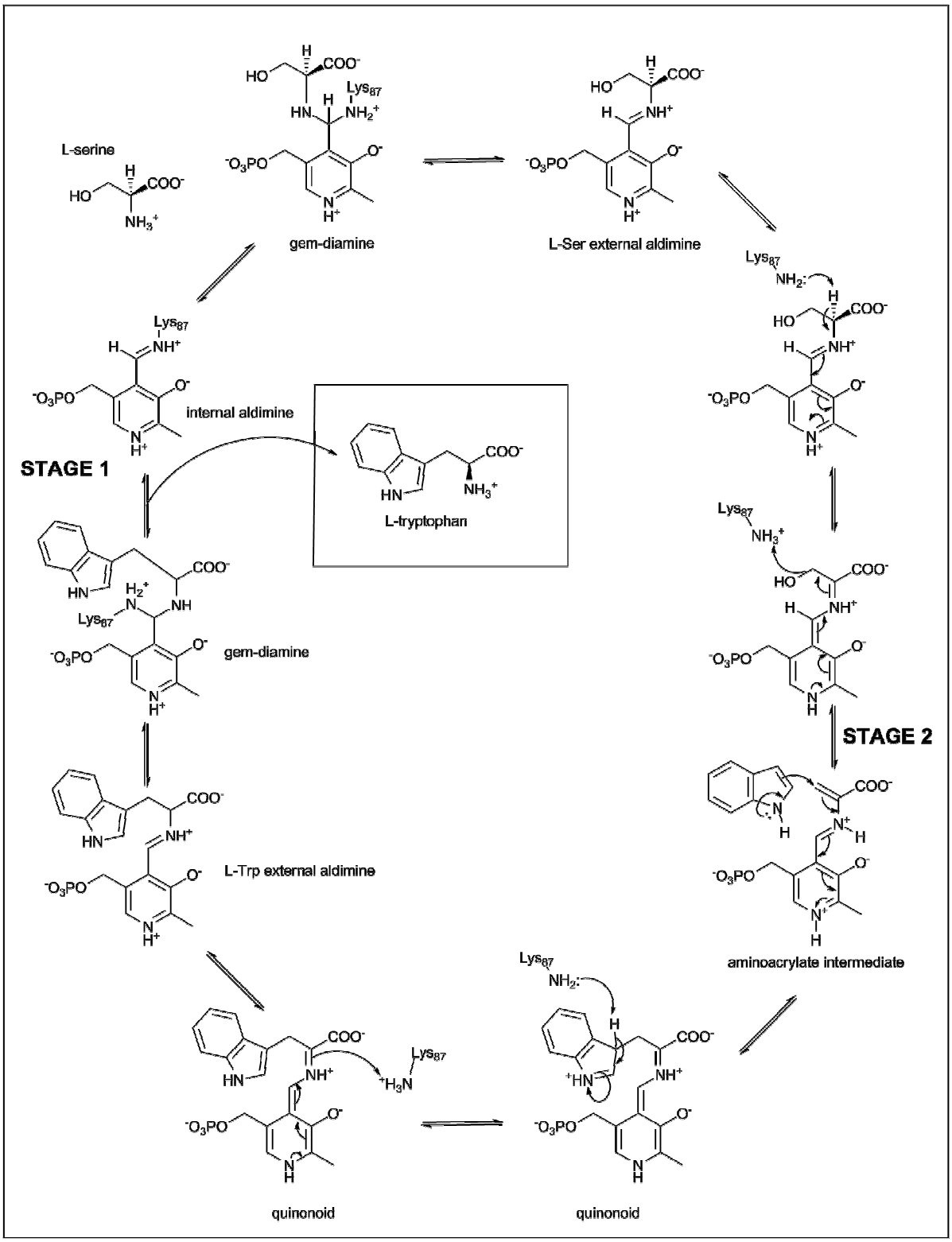
1.4.3 The β-subunit reaction

The reaction performed in the beta subunit is a PLP-catalysed β-replacement reaction between indole and L-serine to form L-tryptophan. The reaction is summarised in scheme 1.4. The reaction in this subunit occurs in two stages with the first stage occurring before the release of indole from the α-subunit.

At the start of the first stage the PLP cofactor is bound to an active site lysine residue (βLys87) forming an internal aldimine between the PLP and the ε-amino group of the lysine.[99] The taking up of L-serine into the active site results in a reversible reaction with the internal aldimine to form a *gem*-diamine intermediate which results in the loss of lysine and the formation of an external aldimine between PLP and serine. The same lysine residue is then suggested to act as a base to abstract the α-proton of serine to yield a quinonoid intermediate which is stabilised by the delocalisation of electrons around the PLP ring. Elimination of water from the serine is driven by the cofactor, PLP, returning to its aromatic state and results in the formation of a stable α-aminoacrylate species. This concludes the first stage of the β-reaction. The subunit waits for indole to be provided by the α-subunit before the second stage of catalysis begins. At this point the α-aminoacrylate species exists in equilibrium with the external aldimine. This equilibrium is affected by pH, temperature, hydrostatic pressure and α-subunit ligands. The external aldimine is favoured under low temperature,[100] high pressure[101] and high pH.[100] The binding of α-site ligands favours the formation of the aminoacrylate which suggests a temporal order to the catalysis – where events in one substrate dictate events in the other.[100]

Evidence of this temporal ordering of the two substrates was provided by binding studies with α-site ligands such as indoleacetyl glycine or indoleacetyl aspartate. Binding of these substrates in the active site of the alpha subunit was seen to disrupt the equilibrium between the two subunits and stabilise the α-aminoacrylate Schiff base in the β-subunit, showing formation of intermediate occurs independent of the release of indole and that changes in the alpha subunit are capable of adjusting behaviour in the beta subunit.[102]

The arrival of indole from the alpha subunit triggers the start of the second stage of catalysis which is proposed to proceed via a Michael reaction between the indole and the previously formed aminoacrylate intermediate. This results in the formation of a resonance-stabilised indoleninium



Scheme 1.4: Mechanism of the β-replacement reaction of tryptophan synthase. Stage 1: the first stage of the reaction occurs with the formation of the internally bound PLP-lysine aldimine and concludes with the formation of the stable α-aminoacrylate intermediate. Stage 2: Michael reaction between indole and the aminoacrylate generated in stage 1 leads to formation of tryptophan external aldimine and then tryptophan release to regenerate the internal aldimine required for stage 1. Adapted from Phillips 2004.[103]

quinonoid carbanion (scheme 1.4). Deprotonation of this intermediate at C3 of the indole forms the quinonoid intermediate of L-tryptophan (in scheme 1.4 this is shown to be catalysed by Lys87 although the involvement of glu109 has also been implicated)[104]. Protonation of this next intermediate generates the external aldimine of L-tryptophan. A transamination reaction with Lys87 results in the *gem*-diamine and subsequent release of L-tryptophan and the regeneration of the enzyme bound internal aldimine.

This β-replacement reaction occurs with retention of configuration at the β-carbon due to the fact that both the α-proton and the β-hydroxyl leave from the same face when generating the aminoacrylate intermediate and that the later Michael reaction occurs on the *si*-face of the indole to produce the *(S)*-indolenine.[105, 106]

1.4.4 Allosteric coupling and kinetics of tryptophan synthase

As discussed above, it has been shown that substrate binding to the α-subunit stabilises the aminoacrylate intermediate required for catalysis and demonstrates allosteric communication between the two substrates. This is not surprising as the two separate active sites are linked by a substrate channel which suggests that communication between the two subunits is needed to ensure cooperation. Due to the unusual nature of the relationship between the two subunits there has been significant research into how they communicate.[99, 104]

The first evidence of communication comes when the α and β sites are separated. Subunit association increases the affinity for the substrates and increases the rates of reaction.[107] In an isolated state the β-subunit prefers β-chloro-L-alanine, whereas L-serine is the substrate of choice for the tetramer, suggested to be because the chloride is an excellent leaving group and so does not require optimised catalysis.[107] Crystallographic data has suggested that this communication is largely mediated by β-helix6 interacting with α-loop2 (contains the active site of the α-subunit). Recent studies have suggested that the decrease in activity seen with the isolated β subunit is due to substrate gate residues Tyr279 and Phe280 blocking entry to the active site, although access can still be granted due to a cleft that forms between two domains of the enzyme in the absence of the α subunit.[108] Both subunits in an isolated state have also been shown to be more stable and rigid than in the tetramer and are therefore unable to perform the conformational changes that are apparently important for activity.[109]

Further evidence of complex allosteric regulation can be found in the ordering of catalytic events. Binding of substrate (or analogue) to the α-subunit triggers the binding of serine to the β-subunit and stabilises the aminoacrylate intermediate,[110] the formation of which leads to a 30 fold increase

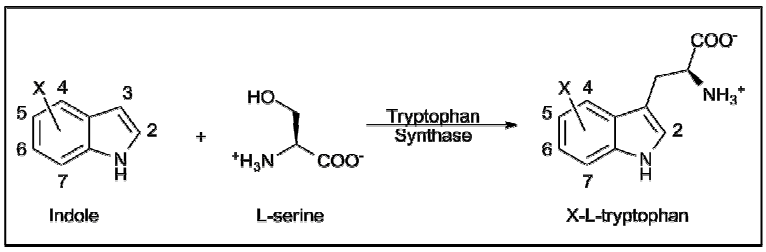
in activity of the α-subunit to produce the indole.[111] Finally the conversion of the tryptophan quinonoid species to the external aldimine of tryptophan triggers the alpha subunit to finally release bound G3P, returning it to starting conditions.[111] This form of communication is mediated by interactions between αGly181 (on α-loop6) and βSer178 (β-helix6). In the absence of α-subunit substrate the α-loop6 is disordered,[95] on substrate binding this loop is stabilised in a conformation that covers the α-active site, forcing it into a closed state.[112] Resulting conformational changes also move the substrate towards the αGlu49 catalytic residue, favouring the retro-aldol catalysis.[104] This information has led to the concept that the subunits of tryptophan synthase can be described to exist in one of two forms; open or closed. In the absence of ligands both substrates are in the open state. As we have seen, binding of IGP to the α-subunit stabilises the loop over the active site and takes the subunit to the closed form. This triggers the stabilisation of the aminoacrylate and increases the preference of the β-subunit to bind serine, which forces the β-subunit also into the closed form and triggers cleavage of indole. This event is mediated by the β-helix6 that interacts with the α-subunit. This all favours the Michael reaction with the indole that results in the tryptophan external aldimine. The β-subunit enters the open conformation again which destabilises the closed α-subunit, increasing the disorder of the closing loop and triggering the release of the G3P from the α-subunit to solution. Tryptophan is released from the enzyme and the catalytic cycle can begin again.

Using a spectrophotometric assay, Kawasaki *et al.* were able to identify the steady state kinetic parameters of the tryptophan synthase dimer as well as the kinetics of the individual α and β-subunit reactions.[113] The activity of the α-subunit was determined by varying the concentration of indole 3-glycerol phosphate (IGP, 0.02 to 0.2 mM). The production of glyceraldehyde-3-phosphate (G3P) within the α/β dimer was followed in the absence of L-serine, using an assay coupled with the NADH producing glyceraldehyde-3-phosphate dehydrogenase. The k_{cat} and the K_m values of this reaction were determined to be 3.6 s^{-1} and 0.1 mM respectively with a specificity constant (k_{cat}/K_m) of 1.4 mM^{-1} s^{-1}. The k_{cat} and K_m values for the β-subunit were determined by varying the concentration of indole (0.013 to 0.2 mM) in the presence of 40 mM L-serine. The conversion of indole to tryptophan was monitored by the change in absorbance at 290 nm. The k_{cat} value for this reaction was estimated to be 3.6 s^{-1} and the K_m to be 0.015 mM. Interestingly in the presence of 10 mM L-serine the K_m and k_{cat} of the α-subunit reaction with G3P improved dramatically to 0.02 mM and 3.0 s^{-1} and the turnover increased to 150 mM^{-1} s^{-1}. This further illustrates the presence of allosteric coupling between the subunits and that the activity of the α-subunit is heavily dependent on the β-subunit.

Anderson *et al.* used chemical quench-flow and stopped-flow methodologies to study the individual rates of each step of the two subunits of tryptophan synthase.[114] Within the α-subunit active site, indole formation from IGP was shown to proceed at a rate of 24 s^{-1}. As seen in the earlier steady state kinetic experiments, in the absence of serine the rate of IGP formation was shown to be significantly reduced to 0.16 s^{-1}. When IGP and serine were added simultaneously a lag phase was observed before G3P was produced. The kinetics of this lag phase (45 s^{-1}) was found to be identical to the rate of aminoacrylate formation within the β-subunit, suggesting that the cleavage of indole within the α-subunit is dependent on the dehydration of serine in the β-subunit. In single turnover experiments with radiolabeled IGP, only trace amounts of indole could be detected, this implied that the rate of tryptophan formation was rapid and that the channelling of the indole from the α-subunit and the subsequent formation of tryptophan must occur at a rate ≥ 1000 s^{-1}. The overall turnover of the enzyme however is limited by the slow rate of tryptophan release (8 s^{-1}) which is the rate limiting step for the whole process when in the presence of serine.

1.4.5 Biotechnological potential of tryptophan synthase

This natural biosynthetic machinery has been harnessed to produce enantiopure L-tryptophan derivatives including 4-, 5-, 6- and 7-chloro-L-tryptophan by providing the purified tryptophan synthase enzyme with a modified indole substrate.[115] A general scheme for this biotransformation can be seen in scheme 1.5. Following on from this concept Goss *et al*. demonstrated the production of a range of tryptophan derivatives including fluorinated, chlorinated, brominated and methylated tryptophan using a cell free lysate.[116] Amino-tryptophans have also been produced using this method.[117] It was discovered that the different halogens were better substrates in the range F > Cl > Br (5-fluoro-tryptophan was produced at 86% yield while 5-bromo-tryptophan was 26%), also indoles halogenated at the 5 or 6 position (see scheme 1.5) were better substrates than those



Scheme 1.5: Incubating indole analogues with the enzyme tryptophan synthase enables the generation of tryptophan analogues. The enzyme catalyses the β-replacement reaction between L-serine and the indole analogue to generate the corresponding enantiopure L-tryptophan analogue. The numbering system around the indole ring is shown.

halogenated at the 4 or 7 position (5-chloro-tryptophan yielded 61% compared to 9% for 7-chloro). This is due to the increasing size of the substituent from fluorine to bromine, the larger halogens finding it harder to squeeze down the narrow channel between the two subunits of tryptophan synthase.[116] Perhaps because of this size restriction, iodinated indoles show no conversion at all. Steric hinderance is also the reason for decreased activity with the 4- or 7-substitued indoles as substitution at this point increases the overall width of the molecule compared to position 5 or 6. The amino-tryptophan variants showed the opposite effect, the 5- and 6-substitued amino-tryptophans yielded 37 and 35% respectively, while 70 and 65% was recorded for the 4 and 7 positions.[117] It is not known what causes this discrepancy but the decreased size and the increased hydrogen bonding ability and solubility of these indole analogues compared to the halogenated counterparts may offer some explanation. It is possible that despite the problems squeezing down the active site tunnel the 4 and the 7 position are preferred inside the active site (perhaps due to steric factors) and therefore any factors that improve substrate access (such as smaller substituents or improved aqueous solubility) help to highlight this active site preference. However if this was the case we would expect to see a similar pattern with the fluorinated compounds as they also exhibit improved solubility, similar hydrogen bonding ability and decreased size, which is not the case. Comparison of the enzyme kinetics with these different analogues may help to identify if the transfer of indole from the α-subunit becomes rate limiting in the presence of sterically bulky groups or if another stage in the catalysis is causing the observed reduction in yield.

This method of harnessing tryptophan synthase shows considerable promise in the preparation of enantiopure L-tryptophan analogues. The fact that cell free extracts of *E. coli* (containing the tryptophan synthase gene from *Salmonella typhimurium*) have been used to catalyse these reactions makes this biotransformation very easy to use compared to having to use isolated enzymes, more recent developments showed that freeze-drying the cell lysate enabled the catalytic activity of tryptophan synthase to be sustained for a longer period (> 2 months) during storage at temperatures as high as 5°C.[117]

This method has already seen applications in biotechnology. As discussed earlier Goss and co-workers were able to produce chlorinated pacidamycin analogues using this tryptophan synthase biotransformation and a precursor-directed biosynthesis approach.[88] More recently the total synthesis of chloptosin utilised this biotransformation to produce the 7-chloro-tryptophan starting material.[118] For this synthesis by Oelke and co-workers the tryptophan synthase methodology was compared to more traditional forms of chemical chlorination and was found to be superior.

Chapter 2:

Engineered Biofilm Formation and Physical Analysis

2.1 Traditional methodologies for growing biofilms in the laboratory

The problems associated with the formation of biofilms, be that medical or industrial, means that significant study of the molecular and genetic basis of biofilm formation has been carried out, although many questions still remain. The mixed species biofilms that form in nature do not lend themselves very well to scientific analysis, therefore various artificial biofilm model systems have been developed that can allow biofilm formation under controlled conditions. There are four generally used methods that have been utilised in the laboratory (for review see Branda 2005).[11]

The most commonly used method involves a piece of equipment called a flow cell. These are used for studying submerged biofilms. Inside a flow cell a culture of the biofilm forming bacteria is constantly pumped through a circuit of tubing. The tubing is connected to a wider chamber inside which a flat substrate is placed (e.g. a glass slide). The pump introduces a slow laminar flow of bacteria that passes over the surface; this encourages the migration of bacteria to the interface of the substrate and allows initial attachment to occur. A laminar flow of nutrients is important to biofilm development as substrate uptake and metabolite excretion within the biofilm generally occurs by diffusion. Over time in a static system, with no flow over the surface, metabolites start to accumulate at the surface and limit the diffusion of fresh substrates. In a laminar system fresh substrates constantly flow over the surface of the biofilm and metabolites are carried away, keeping the concentration gradient within the biofilm as healthy as possible. The formation of biofilms on glass slides makes it possible to easily observe them, for example by using confocal scanning laser microscopy (CSLM). Formation of biofilms in this manner however can be time consuming and not particularly useful for high throughput mutant screening. Also not all biofilms grow on solid surfaces, some form floating biofilms at liquid-air interfaces. Organisms growing in this sort of biofilm require more complex structures to support themselves and in this case a strong laminar flow is destabilising.

A higher throughput method of studying submerged biofilms is in a no-flow system within microtiter plates. Although there is a lack of laminar flow the advantages of this method lie with the rapid production of large numbers of biofilms and this method has been used to study the genes involved in biofilm formation.

The final method is the most widely used in biology, although in most cases not intentionally for the study of biofilms. Bacterial colonies grown on the surface of agar plates are now recognised as a form of biofilm, although not all bacteria will form biofilms under these conditions. Like biofilms grown in microtiter plates these biofilms are amenable to high throughput analyses.[11]

All four methods have been used effectively to study biofilm systems but there can be differences in displayed phenotype if different methods are used. Ultimately none of the four methods can be described as the best, and are complementary to each other.[23]

The growing interest in harnessing biofilms for biocatalysis presents different challenges that suit the flow cell method less well. The ability to easily express a desired recombinant protein within the biofilm is a prerequisite of a versatile, customisable immobilised whole cell biocatalyst and many of the classically studied biofilm species are not appropriate for this task. Another essential requirement for a biocatalytic biofilm is the incorporation of a consistent and predictable quantity of biomass to ensure catalytic rate remains comparable between different batches. To this end, a more controlled and engineered deposition of a biofilm is attractive.

2.2 Results and Discussion

Spin-coating was explored as a method of biofilm formation. This is a method more commonly associated with the immobilisation of inorganic catalysts onto surfaces.[119] *E. coli* K-12 mutant strain PHL644 (MC4100 *malA-kan ompR*234) was chosen as the biofilm forming organism.[42] This strain contains a point mutation in the regulatory protein OmpR which results in the increased activation of *csgD* which leads to the emergence of curli fibres on the cell surface. Traditional K-12 variants are not capable of forming stable biofilms so this mutated strain enables the production of a stable biofilm comprised of an organism that is genetically well characterised, easy to transform and capable of expressing cloned recombinant proteins.

The spin-coating methodology was developed alongside chemical engineers at the University of Birmingham, specifically Dr Andreas Tsoligkas, and was achieved by centrifuging an overnight culture grown in LB/2 medium (OD_{600} > 2) in flasks which contained a standard size glass microscope slide placed at the base. The slide was supported on a bed of glass beads (4 mm in diameter) to reduce the likelihood of the slide cracking under the centrifugal forces. The slide was pre-treated with a poly-lysine (PLL) solution, which is generally used in microscopy to assist in the adherence of cells to glass substrates and has previously been shown to assist in the initial attachment of biofilms inside flow cells.[120] The result after centrifugation was a visible milky layer of microbial growth on the surface of the slide (see figure 2.1).

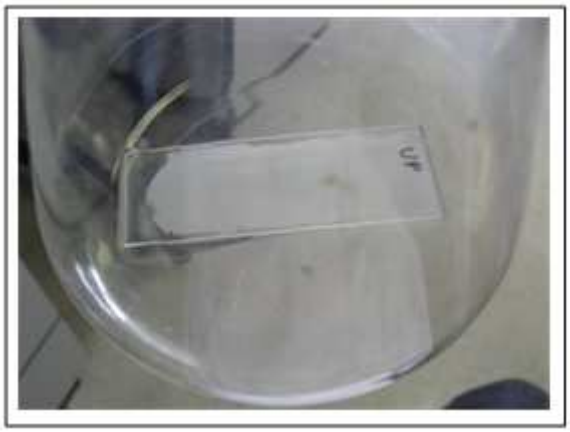


Figure 2.1: Example of a spin-coated *E. coli* PHL644 biofilm. After centrifugation of the *E. coli* PHL644 culture onto a standard microscope slide, a thick milky white layer of cells can be seen adhering to the surface of the glass.

The type of centrifuge rotor used proved to be very important in successfully spin-coating *E. coli* cells onto a surface. The initial spin-coating work was performed using a Jouan C4.22 centrifuge at The University of Birmingham. This centrifuge was fitted with a swinging bucket rotor capable of taking 750 ml centrifuge bottles. The swinging nature of this rotor angles the centrifugal force perpendicular to the glass slide, facilitating a top-down coating approach. A similar centrifuge was not available at the University of East Anglia. Spin-coating of an equivalent volume was attempted using a fixed angle centrifuge but this caused the culture to approach the slide at an incline, therefore a smooth and even coverage of cells was not achieved. The *E. coli* quickly sloughed off the slides once transferred into minimal media.

The only swinging bucket centrifuge available took centrifuge tubes of a maximum 50 ml volume. To compensate for this the microscope slides were cut into smaller squares (1.5 cm long) and into small rounds (2 cm diameter). The spin-coating was performed in the same way on both types (poly-lysine treatment of the reduced size slides and glass beads placed at the bottom of the tubes to prevent cracking). Following the coating, the films soon lost their stability in the minimal media (within 24 hours). This could be due to the increasing surface area to volume ratio as the biofilms become smaller. Since the surface is where the majority of sheering forces are found, it is likely that a smaller area of biofilm is subjected to relatively higher sheer forces than one with a larger area, promoting enhanced stability on the full sized microscope slides. In view of this all, biofilms conducted for this study were spin-coated and analysed using the facilities at the School of Chemical Engineering at the University of Birmingham.

The layer of cells that results after spin-coating cannot be defined as a biofilm, to achieve this title the individual cells should be networked by extra-cellular matrix components.[11] According to studies

in the literature, curli formation in *E. coli* is subject to complex regulation with fibres only being produced at temperatures below 30°C, at low osmolarity and in stationary growth phase.[44] Therefore under conditions of high osmolarity and temperature, curli and consequently biofilm formation is inhibited. For this reason the microbe coated slides were transferred from the relatively highly osmotic LB/2 medium into the minimal M63 medium and left to incubate at 30°C for several days with the hope that the selected conditions encouraged curli formation and subsequent biofilm maturation. A laminar flow is important to maintaining the concentration gradient on the surface of biofilms inside flow cells, so a pseudo laminar flow was introduced by gently agitating the slides in an orbital shaker incubator at 70 rpm. This low value was chosen to reduce the sheering forces on the surface of the biofilm that a higher rpm would introduce. The slides were monitored using environmental scanning electron microscopy (ESEM) for any signs that matrix components were being formed.

In a parallel control experiment, a poly-lysine coated glass slide was placed into a culture of *E. coli* PHL644, that had been grown overnight in LB/2 media to an OD_{600} of > 2 before being re-suspended into M63 media, instead of being spin-coated via centrifugation. This culture was incubated for the same period (> 7 days) whilst being very gently agitated. In this culture biofilm formation would rely on the traditional method of cells attaching to a substrate without any assistance.

ESEM images of the spin-coated biofilm (SCB) were captured by Dr Andreas Tsoligkas of the University of Birmingham. These images illustrated very different characteristics as the maturation period progressed (ESEM images shown in figure 2.2) and the SCB showed a marked increase in three dimensional structure compared to the cells that were allowed to form a biofilm by natural deposition. Comparison of a three day mature biofilm to a four day old specimen (figure 2.2 B, C) showed the formation of channels inside the biofilm and the appearance of deep pores. These pores and channels are often seen during biofilm development and are believed to allow nutrients to flow into and waste products to flow out of the biofilms lower strata.[121] By day five (figure 2.2, D) the spider's web-like evidence of EPS starting to form around the cells could be seen. On the sixth day (figure 2.2, E) the EPS production was very widespread covering all the cells in the white fibrous strands of the ECM and by the seventh day (figure 2.2, F) deep channels and pores were seen that are consistent with the theory of biofilm formation. The dramatic changes in biofilm surface morphology are best illustrated by comparing the third day to the sixth. The biofilm after six days maturation appeared to be much rougher with peaks of cells separated by valley-like channels and the volume of EPS had increased from a few connections between individual cells to a widespread network that encased the microbes into a polysaccharide matrix. The biofilm that was allowed to

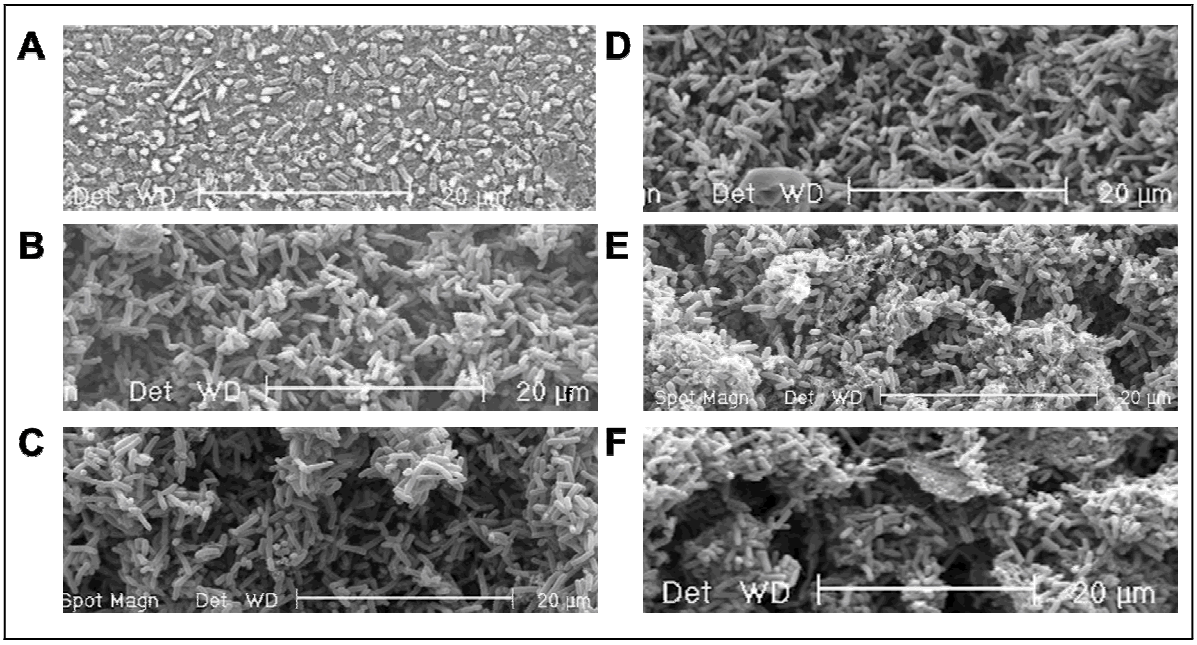


Figure 2.2: ESEM images of natural and spin-coated *E. coli* PHL644 biofilms at various maturation ages. (A) Image of the naturally formed biofilm after 7 days of maturation. Only a very sparse monolayer can be seen on the slide (3880x). **(B)** A 3 day mature spin-coated biofilm (SCB) shows the beginnings of channels within the biofilm but no significant quantities of EPS (3904x). **(C)** After 4 days of maturation of the SCB deeper pores and channels can be observed and slight EPS production has occurred (3913x). **(D)** A 5 day mature spin-coated biofilm still exhibits no significant EPS coverage (3914x). **(E)** A 6 day mature SCB shows considerable exo-polysaccharide production covering the bacterial cells (3906x). **(F)** A 7 day mature SCB continues to show EPS development and deep pores and channels indicating presence of a mature 3D structure (3894x). ESEM images captured by Dr Andreas Tsoligkas of the University of Birmingham.

form and mature naturally for 7 days showed only a sparse monolayer of microbes which contrasted strongly with the large volume of biomass and complex three dimensional structure present after spin-coating.

The increasing adhesive forces within the biofilm that are a consequence of the increase in EPS production were followed by atomic force microscopy (AFM) by Dr James Bowen of the University of Birmingham, using biofilms provided by myself or Dr Tsoligkas. This technique, also known as scanning force microscopy, has seen significant development over recent years in the field of biological imaging. The main advantage of AFM in this arena compared to alternative imaging methods (such as ESEM) is that it can image samples in a fluid environment, often without the need for any pre-treatment. This means that AFM can be used in the observation of biochemical or physiological processes in real time and at molecular resolutions. The theory behind the workings of

AFM has been described as being analogous to a record player.[122] A molecular needle-like tip is attached to a cantilever which has a spring constant that is much smaller than the intermolecular spring constant of the atoms being studied; because of this the cantilever is affected by the tiny forces acting on the tip from the surface it is in contact with. The changes in cantilever position can be followed with an optical deflection detection system. Essentially, a laser is directed onto the end of the cantilever and any shift in the position of the cantilever will cause the laser beam to reflect by a proportional amount, which can be monitored by photo-detectors. The ratio of the original movement of the tip and the deflection of the laser causes the overall change to be amplified by up to 1000 times, this means that even deflections as small as 1 nm or less can be recorded. In operation the micro-tip presses against the surface of the sample with a small loading force, the tip is then raster scanned in the *xy* plane (horizontally) over the sample, or by moving the sample past a static tip. The vertical position of the sample can also be measured in either constant force mode or constant height mode. In the first of these methods the force acting on the cantilever is kept constant, this is achieved by changing the height of the sample to keep the cantilever in the same place. Alternatively the height of the sample can be kept static and the measurements in the *z* plane determined based on the degree of cantilever deflection. In both methods the result is a three dimensional image based on the surface topography of the sample. This data can be used to construct an image based on the *xyz* coordinates recorded at each data point and this has been used many times to image microbial cells.[123]

This AFM technique was employed in a limited capacity to provide an image of the artificially engineered spin-coated biofilm (figure 2.3). The resolution of the image is not high enough to make out much surface detail of the cells; however it does highlight another important ability of AFM, the force measurements from which it gets its name. The way in which AFM works enables molecular forces to be measured by recording how much force is required to move the cantilever away from a surface, for example the micro-tip can be chemically modified/functionalised and the attractive and repulsive forces between the tip and a sample recorded. In this way cell surface charge and hydrophobicity can be investigated and ligand and receptor interactions can be probed.[124] For example, Benoit *et al.* measured cell-cell interaction forces by immobilising a single cell onto an AFM cantilever and approaching the tip towards surface-bound cells.[125]

Figure 2.3 highlights the different adhesive forces present within a SCB with the brighter colours indicating stronger adhesive forces between the AFM tip and the surface. The darker coloured cells can be seen to be surrounded by a highly adhesive region, presumably EPS components. Using this method the changing adhesive properties of the SCB as it matures were monitored. Spin-coated

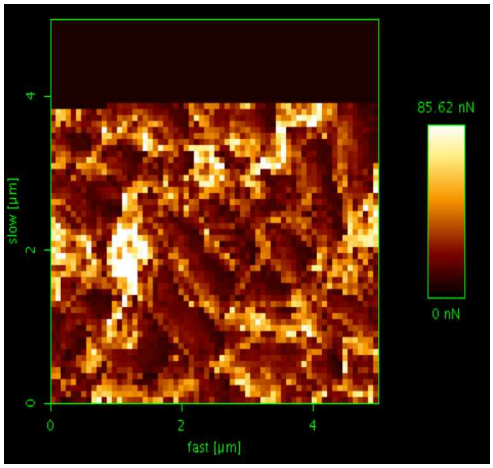


Figure 2.3: AFM image of a spin coated engineered biofilm. At low resolution individual cells of *E. coli* PHL644 can be seen clustering together on the glass microscope slide following spin-coating. Each pixel represents a single AFM measurement. The relative 'stickiness' of different regions of the biofilm is also highlighted with the lighter colours indicating stronger adhesive forces to the tip. It can be seen that surrounding the darker shapes of the cells are areas of higher adhesion. This possibility represents the EPS surrounding the individual organisms. AFM analysis carried out by Dr James Bowen of the University of Birmingham.

biofilms were matured in M63 media and removed after 3-10 days. A 7 mm square section of the spin coated biofilm was excised and placed into an AFM BioCell which contained M63 media, the cell was equilibrated to 30°C and mounted into the AFM. Inside, the cantilever and the biofilm were submerged in the M63. This is known as wet-mode AFM and was performed because it preserves the biofilm interactions and avoids surface tension forces between the tip and wet bacterial surfaces, this provides more accurate force measurements. The AFM instrument allowed a lateral scan range of 100 µm in the *x* and *y* direction and a vertical range in the *z* axis of 90 µm. Five different biofilm regions on each sample were tested with a minimum of 10 readings and three independently grown biofilms were tested for day 3 to day 10 (the results can be seen in figure 2.4).

The calculated adhesive forces between the AFM tip and the surface of the SCB on days 3 through 5 were very low (less than 1 nN). Between days 5 and 6 however there is a dramatic change in the forces detected, with adhesion forces showing an average reading of around 40 nN, a value which remains more or less constant within biofilms up to the tenth day. There is clearly something changing in the structure of the spin-coated biofilm between the fifth and sixth day. Correlation of the AFM results with the ESEM images highlight that this change in adhesive forces occurs at the same time as mass production of EPS components (figure 2.2, E). The biofilm allowed to deposit naturally showed no significant surface adhesion properties even after 7 days in maturation media. However for thorough comparison between the spin coating approach and the more traditional flow

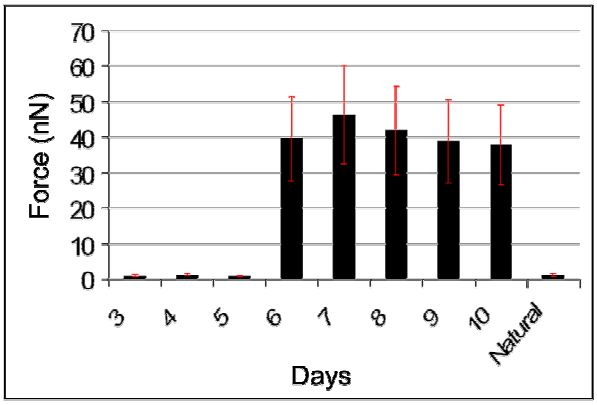


Figure 2.4: AFM calculated mean peak forces for the engineered spin coated biofilm over 10 days and peak forces within a 7 day mature naturally deposited biofilm. There are no significant adhesion forces acting between the AFM cantilever and the biofilm within the first 5 days of maturation or with a 7 day mature biofilm left to deposit naturally. On the 6th day the attractive forces within the SCB increase significantly and remain high for the other four days of study. AFM analysis was conducted by Dr James Bowen of the University of Birmingham.

cell grown natural biofilms, AFM analysis of a 7 day old biofilm of *E. coli* PHL644 formed within a flow cell should be performed.

As well as adhesion forces, distinctive phenomena known as 'snap-off' events were detected using the AFM. These events are caused during contact mode AFM when the cantilever is lowered onto the surface of the cells and exerts a known force onto it. As the tip is retracted the force acting on the tip decreases until the tip has no contact with the surface and the overall force acting on the tip is zero. Figure 2.5A shows how this process looks when plotted onto a force curve. The x axis shows the distance of the cantilever from the cell surface, at distance 0 the tip is in contact with the cells. As the tip is retracted, the force being applied on the cantilever from the surface decreases accordingly. This is illustrated in figure 2.5A as a red line (1) with decreasing gradient. As this line approaches 0 nN it begins to show negative force, this can be explained by attractive forces acting on the tip from the surface (these can range from relatively small atomic forces such as van der waals or large adhesive forces originating from a cell surface). The retracting force begins to compensate for these various attractive forces and the tip begins to retreat from the substrate until all attractive and repulsive forces have been annulled and the cantilever returns to a relaxed state. This can be seen in figure 2.5A as the gradually curving blue line (2) as the force acting on the tip returns to the zero, resting state (green line, 4)

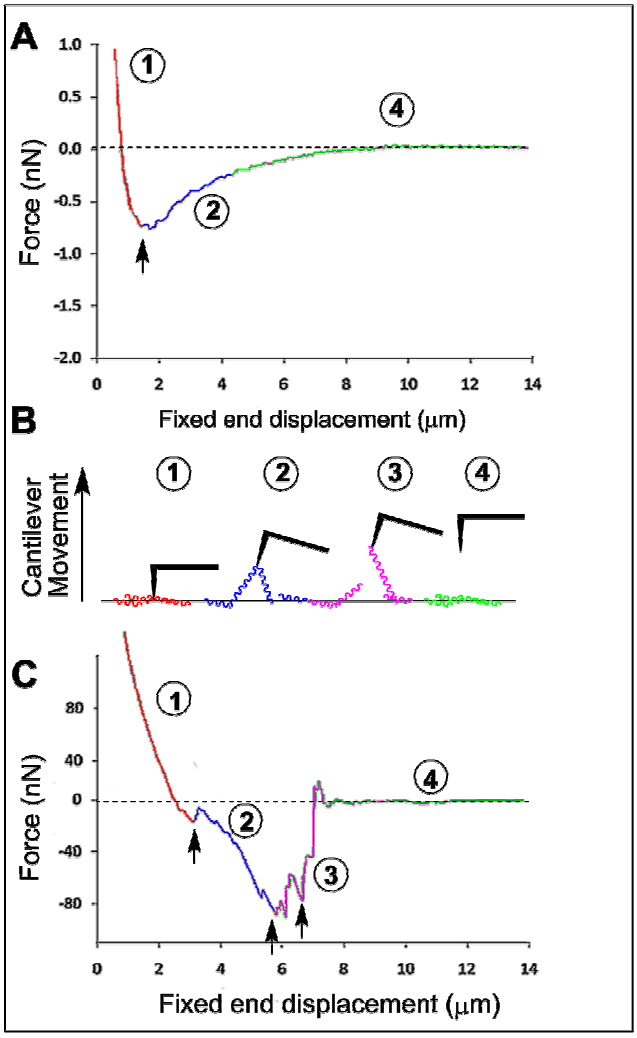


Figure 2.5: Examples of AFM force curves and a cartoon showing different stages of cantilever attraction and retraction. (A) Example of a simple retraction force curve. At stage (1), highlighted in red, the AFM tip is being retracted from the surface. At the end of this stage the tip looses contact with the surface (shown with an arrow). At stage (2), highlighted in blue, various attractive forces between the tip and the surface cause the force acting on the cantilever to decrease slowly. By stage (4), shown in green, the tip has lost contact with the surface forces previously acting upon it and total force has dropped to 0 nN. The total binding force between the surface and the tip is represented by the peak area under the dotted line. **(B)** Cartoon showing the different stages in tip retraction. (1) The tip is in contact the surface. (2) EPS components restrict the vertical movement of the tip. (3) The connections between adhesive fibres and the tip 'snap-off' one by one causing a large change in cantilever force. (4) The cantilever is in a rest state and has no external forces acting on it. **(C)** Force curve showing multiple 'snap-off' events during tip retraction. These 'snap-off' events cause distortions in the curve under the dotted line as a sharp change in force is detected as adhesive EPS fibres detach from the AFM tip. Examples of 'snap-off' events are highlighted with arrows. AFM analysis performed by Dr James Bowen of the University of Birmingham.

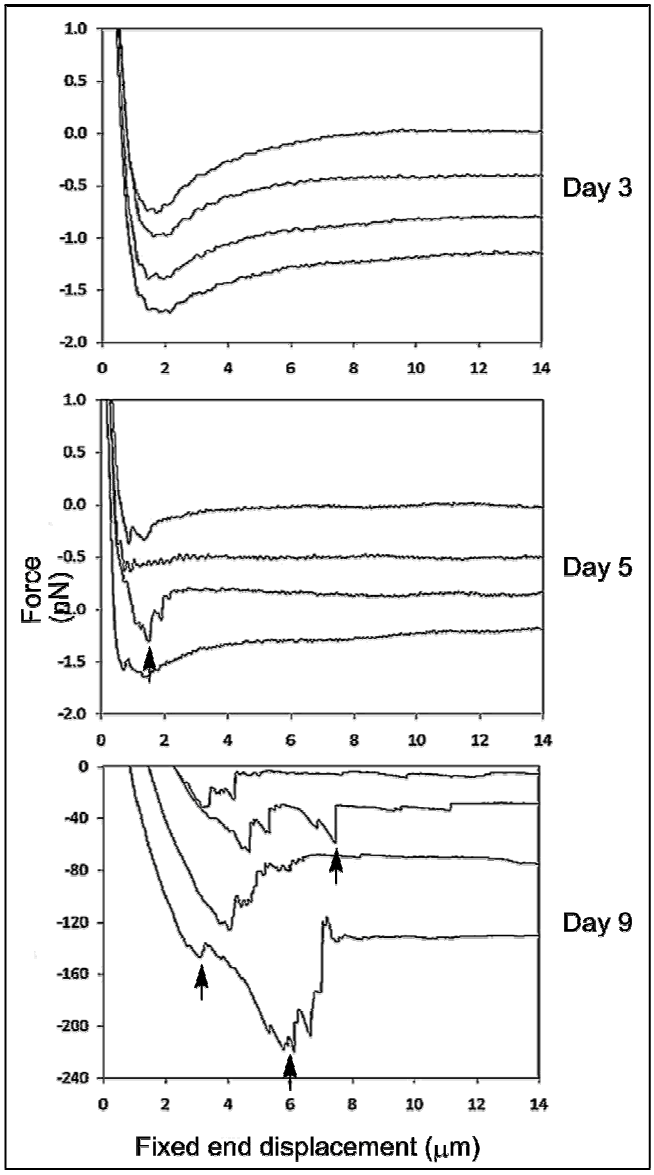


Figure 2.6: Examples of retraction events measured using AFM for spin-coated biofilms over several days. Repeats of four retraction experiments are presented on a single graph. At day 3 a smooth retraction force curve is seen that shows no significant 'snap-off' events. At day 5 more events are seen suggesting increased EPS production. At day 9 multiple 'snap-off' events can be seen suggesting a well developed EPS structure. Arrows indicate examples of such events. AFM analysis was performed by Dr James Bowen of the University of Birmingham.

The distinctive 'snap-off' events occur when AFM is used in this way to study cell surfaces. When the tip pulls away from cells, biomolecules from the cell surface can be pulled off with the tip. As the tip is retracted these biomolecules are stretched and eventually dissociate, which produces a very sudden change in force. These events are marked with an arrow in figure 2.5C. When multiple biomolecules are attached then they will dislocate at different times, meaning several sudden changes in force measured before the curve returns to zero.

When the maturing biofilm (3-10 days old) was analysed by AFM an increasing number of 'snap-off' events occurred with increasing age of the biofilm (figure 2.6). Table 2.1 shows that with a 3 day old biofilm there were only typically 1 or 2 snap-offs events after a minimum of 30 measurements. At day 6 (when EPS production had been shown to begin) this number increased to 3-7. Between days 7 and 9 there were between 5 and 12 events and even greater numbers seen with a 10 day old spin-coated biofilm. Figure 2.6 shows some example force curves from 3, 5 and 9 day old mature SCBs. Day 3 shows a smooth transition between the tip leaving the surface and the return to the resting cantilever state. 5 day old biofilms show a few 'snap-off' events but the 9 day old biofilm shows a largely distorted curve and a large peak area below 0 nN. This increasing peak area indicates an increasingly adhesive cell surface.

Maturation Time (days)	Typical number of 'snap-off' events
3	1-2
4	1-2
5	1-2
6	3-7
7	5-10
8	5-10
9	6-12
10	>12

Table 2.1: **Typical number of 'snap-off' events in spin-coated biofilms of different ages**. A minimum of 30 measurements recorded across three biofilms for each day

Similar multiple events have been reported in studies of polysaccharides on the surface of yeast cells and in the comparison of lipopolysaccharides on the surfaces of gram-positive and gram-negative bacteria.[126, 127] Since the EPS secreted by biofilms is also mainly composed of polysaccharides it seems likely that the events detected in this study are a consequence of increased EPS production and matrix formation. The increasing number of these events is further evidence that there are large changes occurring on the surface of the spin-coated biofilm during the maturation period and that these changes are making the surface more adhesive, linking the cells together into a community.

The production of the extra-cellular matrix surrounding the cells is one of the main requirements in defining a biofilm. The other defining feature is the three dimensional structure, particularly the

characteristic mushroom shaped colonies that emerge as a biofilm matures. Interferometry was utilised to follow the biofilm as it matured and allowed the measurement of biofilm thickness and surface roughness as the maturation period increased (figure 2.7). Interferometry analysis was performed by Dr Tsoligkas using biofilms made by myself and Dr Tsoligkas.

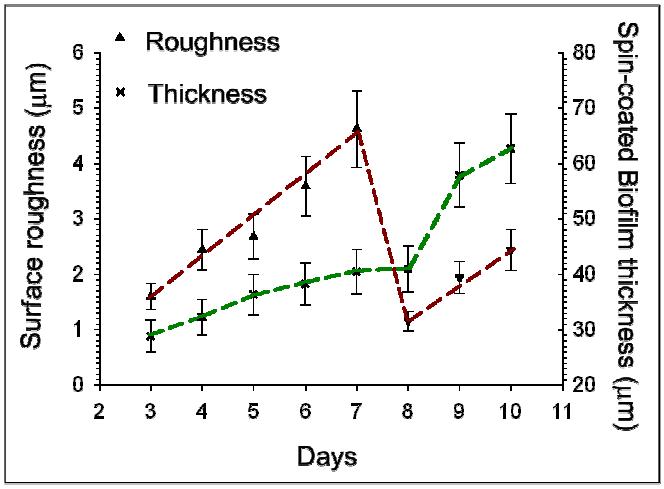


Figure 2.7: Changes in the overall thickness of the spin-coated biofilm and the surface roughness at different stages of biofilm maturation, as measured by interferometry. The engineered biofilm shows increasing thickness (green) and roughness (red) as the biofilm matures up to day 7. At this point a sharp decrease in surface roughness is accompanied by no overall change in biofilm thickness. After day 8 both the thickness and roughness of the biofilm begin to increase again. Error bars represent 3 repeats.

As the biofilm matured from day 3 to day 7 the overall thickness increased from around 30 µm to 40 µm. The surface roughness of the biofilm also saw an increase from just over 1.5 µm on the third day to around 4.5 µm on the seventh. This change can be rationalised by the change in surface features of a biofilm as it develops. Figure 2.8 highlights the changes that the surface of the biofilm goes through as it matures. After 5 days of maturation mushroom colonies start to appear on the surface of the biofilm with channels forming between the caps. It is thought that these distinctive shaped structures are produced by motile bacteria attaching and re-attaching as a result of substrate and mass transfer limitations.[24] The formation of channels within the surface of the biofilm aids the distribution of resources down into the lower echelons of the film. These mushroom shaped colonies continue to grow and as they do the overall surface corrugation increases, this effect is detected by the interferometry as increasing biofilm roughness. Figure 2.7 however shows that after 7 days the surface roughness of the film dropped back to less than 1 µm, while the overall thickness of the film remained constant. This occurs because the formed mushroom colonies have continued to grow larger and larger. Once the mushroom caps have grown large enough they make contact with the

similar colonies around them (figure 2.8, day 7) and eventually merge. The fused mushrooms lose the channels between them and the irregular, bumpy surface is lost but the overall thickness of the film has not changed. This results in a loss of surface roughness as measured by the interferometer.

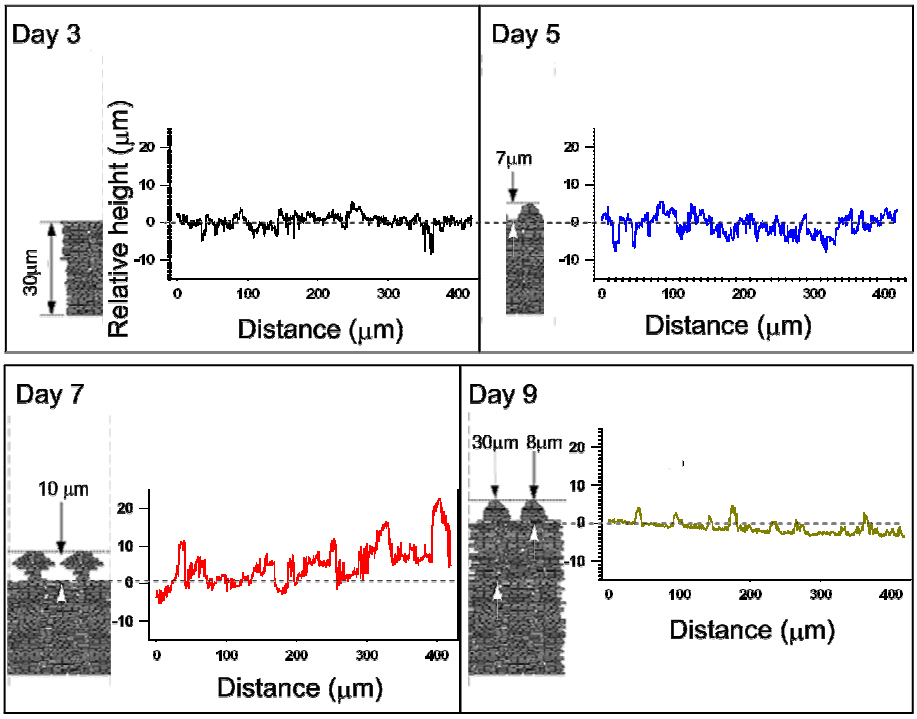


Figure 2.8: Cartoon showing formation of the biofilm three dimensional structure and emergence of mushroom shaped colonies together with the corresponding interferometry data. As a biofilm matures cells at the surface migrate to form mushroom shaped protuberances on the surface. As the mushroom-like caps continue to grow, they fuse together causing the overall thickness of the biofilm to increase. On top of the new biofilm surface fresh mushroom shaped colonies begin to appear. Interferometry analysis of the spin-coated biofilm shows surface roughness increasing with the appearance of mushroom colonies. Interferometry performed by Dr Andreas Tsoligkas of the University of Birmingham.

The loss of the channels in the biofilm once again introduces mass transfer limitations into the biofilm. This triggers the same cell movement as before and further mushroom shaped colonies begin to form on the surface of the biofilm. This increases the overall thickness of the film and starts to intensify the surface roughness once again. This effect was seen by analysing 9 and 10 day old spin-coated biofilm slides by interferometry (figure 2.7).

The individual interferometer readings for days 3,5,7 and 9 are shown in figure 2.8 along with cartoons showing how the formation of surface structural features like mushroom colonies affect the surface topography. At day 3 the interferometry results showed a fairly level surface without

much variation in vertical height. The interferometry data for the fifth day displayed increasing roughness symptomatic of channels forming between the peaks. The seventh day showed the greatest surface roughness with large peaks separated by large troughs consistent with the theory that at this stage large mushroom colonies have formed with channels flowing between them. The surface of the interferometer studied biofilm at day 9 shows much reduced surface features and is more similar to the biofilm on the third day than the seventh. This data strongly correlates with the theory that as the biofilm grows it is going through cycles of mushroom colony formation followed by fusion and increased biofilm thickness.

The total biomass that had formed onto the slides that had been matured for 7 days was calculated using dry cell weight calculations. The total biomass on 14 different 7 day old spin-coated biofilms was re-suspended into sterile water and heated at 100°C. The mass of the cells was measured until a constant mass was reached. The total dry biomass was measured to be approximately 80 mgs on each slide. In previous biocatalysis work conducted with biofilms it has been noted that producing biomass consistently between batches is very difficult using conventional methods.[81] If too much biomass is produced then mass transfer limitations begin to affect the organism and sloughing of cells from the surface can begin, which can lead to the clogging of reactor systems. On the other hand if too little biofilm is produced then catalytic rates will be affected. The consistency of biofilms produced using this spin-coated method is attractive for biocatalysis as reliability between multiple batches is an important starting point for industrial processes.

Immediately following the spin-coating procedure the film of bacteria does not really fulfil the criteria of being a biofilm (i.e. a community of attached cells surrounded by an extracellular matrix) and is in actuality just a layer composed of compressed *E. coli* cells. The analytical results suggest that after 6 days in minimal media at low osmolarity there has been a change in the cell behaviour and that extracellular matrix components have been produced (as seen as white fibrous material on the ESEM and as increased adhesion forces measured by AFM). The three dimensional structure of the engineered biofilm is also behaving in a dynamic fashion with mushroom shaped colonies forming and fusing (as seen by interferometry).

All this suggests that despite its humble beginnings the spin-coated engineered biofilm has undergone a shift from simply being a layer of compressed cells to a matrix enclosed community of bacteria that strongly resembles a biofilm grown in traditional flow cell apparatus. With the increased control over deposition and reproducible incorporation of biomass this method seems to be able to produce biofilms in a manner that has advantages over traditional methodologies.

Chapter 3:

Performance of the Spin Coated Engineered Biofilm as an Immobilised Catalyst and Comparison to the Activity of Planktonic Bacteria and Traditionally Immobilised Enzyme

3.1 Introduction

3.1.1 Biofilms as biocatalysts

In the clinical environment biofilms are exclusively viewed as being a menace and are known to play a considerable role in a variety of persistent infections, including dental caries (e.g. *Streptococcus* infection) and cystic fibrosis pneumonia (e.g. *P. aeruginosa*).[53, 128] Industrially however, biofilms are not seen in such black and white terms. Unwanted biofilms are undesirable in industry because they can, amongst other things, foul reactor systems, contaminate food processing equipment and reduce heat transfer in heat exchangers or cooling towers.[70] The increased robustness of biofilms compared to planktonic cells has meant that mixed species biofilms have found some industrial applications in bioremediation and waste water treatment.[129] Single species biofilms are used in the production of fine chemicals such as ethanol or 2,3-butanediol through fermentations or biotransformations.[64, 70] There are fewer reports of recombinant strains of bacteria being utilised to over express a specific protein and perform desired biotransformation reactions. In the last few years there has been a surge of interest in the potential of biofilms to perform biotransformation reactions and the new method of engineering spin-coated biofilms presented here opens up the possibility to form *E. coli* biofilms that are easy to make, resistant to harsh conditions and capable of being produced alongside an enzyme of choice.

3.1.2 Traditional methods of enzyme immobilisation

The need for long lived, stable biocatalysis has driven the development in enzyme immobilisation techniques. In addition to making enzymes easier to handle this enables the separation of catalyst and product to ensure no protein contamination appears in the final material. Immobilisation also enables the recycling of enzymes that can be time consuming and expensive to produce. In general there are three traditional methods of enzyme immobilisation; binding to a support, entrapment and cross-linking (for extensive review see Sheldon 2007).[130]

Enzymes can be bound to supports using physical (hydrophobic, van der Waals), ionic or covalent bonding. Physical bonding is the weakest of the three and is generally not strong enough to maintain an enzyme-support attachment. Covalent and ionic bonding forces are strong enough to prevent enzyme leaching from the support during catalysis.

Entrapment of enzymes involves capturing them inside a polymer network lattice or inside micro spheres. The forces here are not usually strong enough to totally prevent catalyst leaching, so additional covalent attachment is often used in parallel. This blurs the line between support binding and entrapment techniques.

The problem with the above methods is that the addition of bulky carriers effectively increases the portion of the immobilised catalyst that is not active, resulting in a less favourable mass to catalytic activity ratio.

Thirdly, enzymes can be cross-linked using bifunctional reagents. This technique arose from the discovery in the early 1960s that the surface NH_2 groups of enzymes can be cross-linked together with a chemical joiner such as glutaraldehyde to form enzyme crystals.[131] The draw backs of this technique however were low activity retention, poor reproducibility and low mechanical stability. Since then improvements made to the technique have seen the emergence of cross-linked enzyme aggregates.[132] In this process enzymes are precipitated by increasing the salt concentration or with the addition of organic solvents or non-ionic polymers. The resulting enzyme aggregates are then cross-linked to produce a permanently insoluble mass of enzyme that still maintains a pre-organised structure. This technique can be mixed with conventional ammonium sulphate precipitation to combine the purification and immobilisation steps of an enzyme preparation.

There is no clear answer as to which method is the best one for a given system, comparison of different immobilisation methods for a single enzyme are not easy to find. Although there are thousands of papers on enzyme immobilisation the comparison in each is usually between immobilised and free enzyme and not between different immobilisation techniques.

3.2 Biocatalytic performance of the SCB with tryptophan synthase

To probe the biocatalytic potential of the spin-coated biofilm the enzyme tryptophan synthase was selected. This enzyme mediates the PLP (pyridoxal-phosphate) dependent biotransformation between indole and L-serine to form L-tryptophan. This enzymatic mechanism has previously been harnessed to produce enantiomerically pure L-halo-tryptophan by providing the enzyme with

commercially available halogenated indole analogues which are taken up into the active site and converted to the corresponding tryptophan.[116, 117]

For this study, three different substrates were chosen: 5-fluoroindole, 5-chloroindole and 5-bromoindole. Indole analogues halogenated at the 5 or 6-position had previously been shown to be better substrates for the enzyme (contained with a cell free extract) compared to analogues substituted at the 4 or 7 position [116] and they also compare favourably in terms of cost and commercial availability. Fluorine containing analogues are also better substrates than either the chlorinated or brominated equivalents with tryptophan synthase. A range of indole analogues substituted at the same position but containing different halogens will enable the effect of the different halogens on the spin-coated biofilm catalyst to be studied.

3.2.1 Process development

3.2.1.1 Development of analytical methods to assess tryptophan synthase yield

In the previous studies with tryptophan synthase by Goss *et al.* (2006), the yield of the tryptophan product was directly determined quantitatively by ethyl acetate extraction of the un-reacted indole starting material and subsequent purification of the aqueous fraction on reverse phase C18 silica to access the pure tryptophan.[116] To facilitate the assessment of the ability of the biofilm to mediate this biotransformation a higher throughput analysis was needed that could detect formed product at low concentrations in small sample volumes. Reverse phase HPLC with PDA detection (190 to 800 nm) was chosen as the most suitable method for analysis. Tryptophan has UV absorbance properties that make it easy to detect with a PDA device (photodiode array).

Samples of pure 5-fluoro-L-tryptophan, 5-chloro-L-tryptophan and 5-bromo-L-tryptophan were obtained by following the procedure of Goss *et al*. (2006)[116]. A cell free extract of *E. coli* containing the tryptophan synthase enzyme (commercially available strain, ATCC 37845) was added to a solution of halo-indole, L-serine and PLP, in potassium phosphate buffer, and incubated for 3 days. Following this, the unreacted indole was extracted and the tryptophan purified by reverse phase C18 silica chromatography, yielding samples of all three products, the purifty of which was confirmed by NMR and LCMS analysis (see Appendix 1). These were used as HPLC standards to ensure accurate identification of product from fractions of the biofilm-catalysed biotransformation reactions.

A reverse phase HPLC method (LCgrad_MeOH) was developed using a ZORBAX SB-C18 column (4.6 mm x 15 cm). The mobile phase was a gradient of methanol versus water (with the addition of 0.1%

formic acid to both) at a flow rate of 0.7 ml min^{-1}. After an initial 30 seconds at 10%, the methanol concentration was gradually increased to 95% over 12 minutes. This was followed by a 95% methanol wash for 2.5 minutes before the concentration of methanol was returned to the starting level of 10% and held for a further 5 minutes.

Using this method, the 5-halo-tryptophan analogue standards were seen to elute after 6-10 minutes depending on the hydrophobicity of the substituent (unsubstituted tryptophan or fluorinated tryptophan eluted early in this time range, chlorinated substituents later and brominated tryptophan eluted latest). The differing retention times of the three analogues enables the identification of the formation of the different tryptophan products. Using this method 5-fluoro-tryptophan demonstrates a retention time of around 7.7 minutes, 5-chloro-tryptophan around 9.4 minutes and 5-bromo-tryptophan elutes from the column at around 9.8 minutes. The different PDA spectra of the three products also enables identification of product formation as the λ_{max} (wavelength of maximum absorbance) is different for each analogue: 278 nm for 5F-trp, 280 nm for 5-Cl-trp and 287 nm for 5Br-trp. The haloindole starting materials show a much later retention time of around 12-13 minutes. Comparison of the retention time and λ_{max} values of new peaks formed during the biotransformation reactions allows assignment of these new peaks as halo-tryptophan product.

Analysing samples of halo-tryptophan standards made up to specific concentrations allowed the construction of a standard curve where tryptophan concentration was plotted against HPLC peak area for different tryptophan analogues. Figure 3.1 is an example of such a standard curve

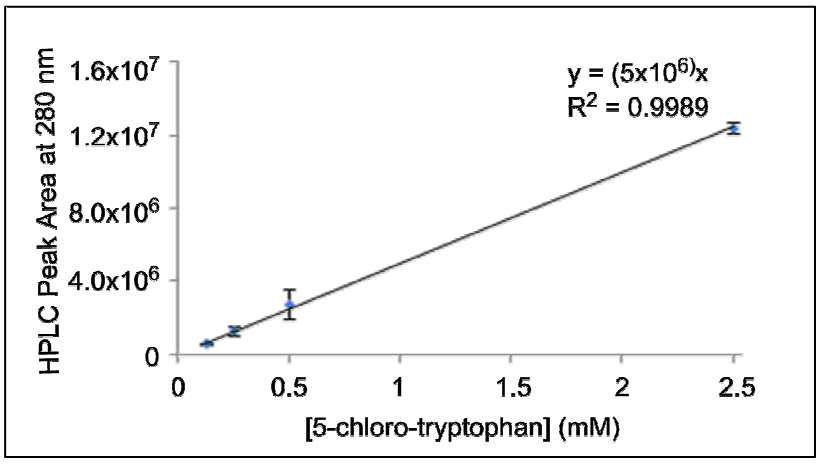


Figure 3.1: Standard curve of HPLC peak area at 280 nm plotted against concentration of 5-chloro-tryptophan. Plotting HPLC peak area against concentration of sample illustrates the linear relationship between the two variables. The resulting gradient of the curve can be used to calculate unknown concentrations based on the peak area of a sample. Data based on calculations from 4 separate samples per data point, standard deviations shown.

constructed by measuring the HPLC peak area of samples containing a range of concentrations of 5-chloro-tryptophan. Peak areas were calculated at the λ_{max} for each tryptophan analogue to obtain the most accurate results (e.g. 278 nm for tryptophan and 280 nm for 5-chloro-tryptophan) (Appendix 1 contains calibration curves for each of the 5-haloindole substrates). Figure 3.1 illustrates the linear relationship between peak area and concentration. At the λ_{max} value the gradient of the plot was the same ($y=(5x10^6)x$) for all of the halogenated tryptophan analogues tested. This gradient of the curve can be used to calculate unknown tryptophan concentration from the HPLC peak area of a sample. The concentration of tryptophan can be used to calculate the number of moles present in a solution of known volume and therefore the overall yield.

To test the accuracy of this yield estimation, two identical biotransformation reactions were performed with 5-bromoindole and a cell free extract of the tryptophan synthase producing *E. coli* MW-043, contained within a length of dialysis tubing, using the standard published procedure.[116] The biotransformation reactions were incubated for 2 days, after which the cell lysate was removed. A 0.5 ml aliquot of each crude reaction was taken and analysed via the HPLC. Any unreacted indole in the reactions was extracted into ethyl acetate and the aqueous layer purified with reverse phase C18 silica. The silica was washed with water and the eluent monitored by TLC until all the PLP and unreacted serine had been washed off. The tryptophan was eluted in methanol. The solvent was removed under reduced pressure and the final yield of the resulting dried tryptophan obtained by weighing. Peak areas were obtained from the aliquots analysed via HPLC and the concentration of 5-bromo-tryptophan estimated using the standard curve of concentration versus peak area. The comparison of 5-bromo-tryptophan yield calculated using both methods is shown in table 3.1. The dry weights of the two reactions assigned the final yields as 15 and 23% respectively. The HPLC estimations were almost identical at 17 and 21%. This strongly indicated that determining the yield of the tryptophan synthase biotransformation reactions using the HPLC offered a fast, reliable, high-throughput and most importantly accurate way of following the biotransformation reactions.

5-Br-tryptophan biotransformation reaction	Dry weight of 5-Br-tryptophan obtained from column	Yield estimated from dry weight of product	Yield estimated from HPLC peak area
1	6 mgs	15%	17%
2	9 mgs	23%	21%

Table 3.1 Comparison of calculated yield of tryptophan as obtained by weight of dry product and by estimation of concentration based on HPLC peak area.

3.2.1.2 The activity of tryptophan synthase in *E.coli* PHL644 planktonic cells

The cell lysate based tryptophan biotransformation published in the literature[116] is based on the activity of cell free extracts of *E. coli* CB149 (in this study named MW-043), a species that is commercially available from the American Type Culture Collection (ATCC 37845) and contains the pSTB7 plasmid that encodes both subunits of the tryptophan synthase enzyme.[113] pSTB7 is an 8.2 kb derivative of pBR322 that contains parts of the *trp* operon from *Salmonella typhimurium* TB1533 including the promoter region and the structural genes *trpA* and *trpB* that code for the α and β subunits of the tryptophan synthase enzyme. The plasmid also contains a truncated version of *trpC* and the terminator sequence. The trp repressor and attenuator regions that normally negatively regulate the operon are absent; meaning that production of genes under the trp promoter is constitutive and does not need induction.

If the overall goal of producing this enzyme and performing this biotransformation within the spin-coated engineered biofilm was to be fulfilled then evidence that the biotransformation could be carried out with whole planktonic cells of the biofilm forming strain of *E. coli* PHL644 had to be obtained. The pSTB7 plasmid was purified from cultures of *E. coli* MW-043 and transformed into competent cells of PHL644. The resulting recombinant strain was labelled as *E. coli* MW-002.

The PHL644 biofilm forming strain of K-12 *E. coli* had not been used previously to perform any biocatalysis so the activity of this strain with tryptophan synthase was assessed both as whole planktonic cells and as a cell free lysate, this allowed direct comparison between PHL644 and the original CB149 strain.

Two 500 ml cultures of *E. coli* MW-002 were grown overnight using the standard method of tryptophan synthase cell lysate preparation (see chapter 7.10.2.3). The bacteria were collected *via* centrifugation and the pellets washed with brine. Following an additional centrifugation the two cell pellets were re-suspended in tryptophan synthase lysis buffer (containing 100 mM KH_2PO_4, 5 mM EDTA and 0.1 mM PLP). The first of these cell suspensions was sonicated and cleared by centrifugation to produce the cell free lysate. The second was used directly in the biotransformation reaction as a planktonic cell suspension.

5 ml of either the cell lysate or the cell suspension was used to catalyse the conversion of 5-bromoindole to the corresponding tryptophan. The effect of supplementing the reaction with 2.5% organic solvent (either DMSO or acetonitrile) was also trialled to see if enhancing the solubility of

the indole in the buffer could improve the overall conversion. The reactions were carried out in triplicate and the average yields are shown in table 3.2.

The mean yield of the whole cell catalysed biotransformation compares favourably with the activity of the cell free lysate. Approximately twice the conversion was seen with the planktonic cells under all conditions tested. The addition of 2.5% DMSO to the reaction was beneficial to the final yield in both the planktonic cells and the cell free lysate reactions. The addition of acetonitrile had the opposite effect and reduced the final yield.

The sonication process undergone by the cell free lysate produces high levels of energy and the cultures must be kept on ice throughout to avoid warming the protein solution. Despite the precautions this process of lysing cells is an aggressive one and there is evidence that with some proteins activity can be lost as a result.[133] The proteins contained within intact cells are protected from external environmental conditions and proteases that may denature or hydrolyse them. Therefore it can perhaps be suggested that the increased activity of the planktonic cell catalyst may be due to increased enzyme stability and improved longevity of catalysis compared to the isolated and more vulnerable proteins present in the cell lysate. Although with the whole cells there are more substrate transport and product removal issues to consider.

	% Yield	% Yield with addition of 2.5% DMSO	% Yield with addition of 2.5% ACN
E. coli MW-002 planktonic cells	18.5 ± 1.4	22.8 ± 0.2	16.9 ± 0.8
E. coli MW-002 cell free lysate	9.9 ± 0.5	13.5 ±3.6	5.0 ± 1.9
E. coli CB149 cell free lysate (ATCC 37845)	26[116]	-	-

Table 3.2: Comparison of the same 5-bromo-tryptophan biotransformation reaction performed with equal volumes of either a suspension of MW-002 planktonic cells or with the equivalent cell free lysate. The effect of 2.5% organic solvent was also examined. Reactions were performed in triplicate. The yield of the same reaction catalysed by the original CB149 strain from the literature is shown for comparison.

The effect of different buffers on the biotransformation reaction was also studied. In previously published work, the tryptophan synthase biotransformation reactions had been carried out at pH 7.8 in 100 mM mono-basic potassium phosphate buffer (KH_2PO_4). The effect of lowering the pH to 7 and reducing the phosphate concentration to 10 mM was investigated. The biotransformation efficiency

of the biofilm maturation M63 media itself was also tested. The cells from a 20 ml overnight *E. coli* MW-002 culture grown in M63 media were collected and re-suspended into 20 ml of tryptophan synthase biotransformation buffer of the appropriate pH or phosphate concentration being tested. 2 ml of this suspension was added to 18 ml of the relevant reaction buffer which was supplemented with serine (5 mM), 5-bromoindole (5 mM), PLP and 5% DMSO (v/v). After three days of incubation at 37°C, 1 ml aliquots were taken and analysed by HPLC. The resulting yields of 5-bromo-tryptophan can be seen in table 3.3.

Buffer	pH	% yield of 5-bromo-tryptophan (duplicate)
100 mM KH_2PO_4	7.8	5.6 ±0
100 mM KH_2PO_4	7.0	4.6 ±0.21
10 mM KH_2PO_4	7.8	2.7 ±0.14
10 mM KH_2PO_4	7.0	2.5 ±0.07
M63 Media	7.0	3.7 ±0.14

Table 3.3: Comparison of 5-bromo-tryptophan yield performed with different reaction buffers

The results illustrated that 100 mM phosphate buffer at pH 7.8 gave the best yield with the biotransformation reaction (5.6%). This is not surprising as these conditions had been used previously because they showed optimum activity with the cell free lysate. Interestingly the reduced pH does not seem to be as important to the final yield as the phosphate concentration. Reduction of phosphate to 10 mM resulted in the overall yield of the reaction dropping by half while the reduction in pH to 7 decreased the yield by about 20%. The M63 media performed fairly well as the biotransformation medium. This media consists of a 100 mM phosphate buffer at pH 7 plus other nutrients, the extra media components obviously having an effect on the yield as there is a slight drop in activity (about 20%) compared to the straight buffer at the same pH.

3.2.1.3 Selecting biotransformation conditions optimal for biofilm catalysis

For the biotransformation reaction to occur in the optimal conditions seen above, the biofilm would have to be removed from the M63 maturation media and placed into a buffer solution. The stability of the artificially spin-coated biofilm in reaction buffer was determined.

Biofilm coated glass slides were produced using the spin-coating method as detailed in chapter 2 and matured for 7 days, this time period had been selected as the optimum maturation period for the spin-coated biofilm slides. At this age of maturation the cell stickiness as measured by the AFM was at its maximum and therefore EPS production was judged to be at the highest observed level. Therefore this was the earliest time period that a fully mature biofilm had formed that could be utilised for biocatalysis.

The stability of the biofilm after being transferred from one media to another and the effect of adding 5-bromoindole (5 mM) was examined. The M63 media was gently removed by a syringe from a SCB (spin-coated biofilm) slide. This was then immediately replaced with fresh M63 media and the flask returned to incubation for 4 hours. 5-bromoindole (5 mM) was then added and the biofilm incubated for a further 3 hours. The refractive index of the solution relative to water was taken as an estimation of the concentration of cells that had sloughed off the slide (used because the concentration of cells was thought to be too low for accurate OD_{600} measurements). In the same way, the M63 media was replaced by the tryptophan synthase 100 mM, pH 7.8 potassium phosphate biotransformation buffer. The effects of lowering the potassium phosphate concentration ten times and reducing the pH to 7 were also examined as before. These experiments were conducted with the assistance of Dr Andreas Tsoligkas of the University of Birmingham.

Following the addition of indole the refractive index of the M63 media relative to water was measured to be 0.314. The biotransformation buffer under the same conditions gave a refractive index of 0.496, suggesting that the biofilm was more unstable under these circumstances. Halving the potassium phosphate concentration to 10 mM had limited effect, resulting in a similar index of 0.341. The biofilm incubated in medium at lowered pH exhibited greater stability, resulting in a refractive index of 0.126. The experiments with the M63 media and pH 7 biotransformation buffer were repeated but this time OD_{600} measurements were taken and 5-bromoindole was not added. As expected the detected OD measurements were very low but values were obtained. Three biofilm slides transferred into fresh M63 media gave OD_{600} readings of between 0.06 and 0.08 after 4 hours. Slides transferred into the pH 7 biotransformation buffer gave OD readings between 0.03 and 0.035. These readings confirmed that the biofilm showed the highest stability when transferred into 100 mM postassium phosphate reaction buffer at pH 7 compared to fresh M63 media or other phosphate buffers at higher pH.

Since the pH 7 biotransformation buffer resulted in the most stable biofilm and had also showed good activity in the planktonic cell biotransformation experiments, this buffer (100 mM KH_2PO_4, pH 7.0) was selected as the biofilm biotransformation buffer.

During biofilm maturation a pseudo laminar flow was achieved by agitating the slides in an orbital shaker incubator, set at 70 rpm to reduce sheering forces during the formation of the adhesive matrix. The stability of the SCB in this biotransformation buffer at higher agitations was tested by monitoring the OD_{600} of the biofilm media at 70 rpm for 5 hours before increasing the agitation to 100 rpm for an additional hour before further increasing the agitation to 150 rpm. The stabilities of three separately formed biofilms were measured (figure 3.2).

During the initial 5 hours of testing the optical density of the buffer did not increase by a significant amount within any of the triplicate repeats. Total re-suspension of another biofilm slide into the same volume of biofilm biotransformation buffer showed an OD_{600} reading of 0.532. This optical density value was used to represent the density of the buffer if the entire biofilm had been sloughed off the slide (this value is marked by a horizontal black line in figure 3.2). Once the shaking was increased to 100 rpm the optical density of all three triplicates started to increase but with one of the repeats showing considerably greater cell dissociation than the other two. Increasing the rpm to 150 resulted in almost total sloughing of the biomass of two of the triplicates. The third triplicate

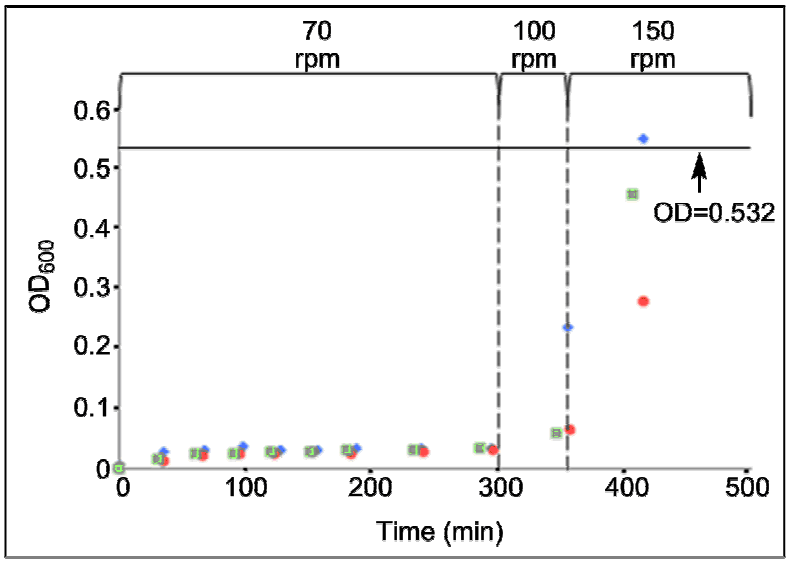


Figure 3.2: Tolerance of three separate spin-coated biofilms in biofilm biotransformation buffer to increasing incubation agitation. The optical density of the buffer containing a spin coated biofilm was measured for 5 hours at 70 rpm. The optical density did not significantly change over this period. The rpm was increased to 100 rpm. The increase in OD_{600} values in this period indicate cells sloughing off from the slide. After the increase to 150 rpm the optical density values rose further as more and more cells detached from the film. The optical density that corresponds to a fully re-suspended SCB (0.532) is highlighted with a horizontal blank line. Squares, circles and diamonds represent results from three separate biofilms.

showed approximately 50% of the biomass was still entrapped after one hour at the higher rpm. However even this lowest value demonstrates a loss of biomass that is unacceptable, therefore the biotransformations would also have to be performed at low rpm to maintain the stability of the biofilm.

3.2.1.4 Determining period of biotransformation based on period of maximum biofilm stability

Measuring the optical density of the buffer to calculate the number of cells that had sloughed off the biofilm was used to judge a sensible time frame for biocatalysis. Biofilm coated slides were submerged into the biotransformation buffer as before but incubated at a low level of agitation, 70 rpm. OD_{600} measurements were then taken on aliquots of the culture over a 40 hour window. The results are displayed in figure 3.3. The biofilms for this experiment and the analysis was conducted by Dr Andreas Tsoligkas.

Within the first 10 hours of submersion in buffer, the optical density of the solution increased slowly. The biofilm therefore showed only slight dissociation into the medium over this period. As time increased the biofilm dissociated further. Between 25 and 30 hours, the mean optical density of a number of biofilm containing solutions showed values around 0.05. As seen before, a totally re-suspended biofilm provided the OD_{600} value of 0.532 (shown by a solid black line in figure 3.3). This value represents total dissociation of the biofilm. The mean OD_{600} value collected for 30 hour submerged biofilms therefore represents a loss of about 10% of the biofilm cells (threshold shown by a line of short dashes in figure 3.3). The variation in OD_{600} values between the different biofilms tested at this time period increased, with some biofilms showing up to 20% cell dissociation (shown as longer dashed line in figure 3.3). After 40 hours the average OD_{600} measurement showed a loss of more than 20% of the original biomass. The variation in results at this later stage, though, was greater than the variation seen at the earlier time points; some biofilms showed only a 10% biomass loss after 40 hours, while the biofilms worst affected had lost nearly 50%.

Due to this increasing inconsistency the 30 hour window was selected as the best period for testing biofilm biocatalysis. After this period over 20% of the biofilm may be dissociated and any catalytic rate at this point may be attributed to planktonic free cells that had detached into the solution and not the surface attached biofilm cells. Cell detachment may also be caused by cell death and this could be triggered by the lack of essential metabolic substrates within the biotransformation buffer. If the cells have started to die at this point then a corresponding drop of reaction rate may be expected and would unfairly influence the reaction kinetics.

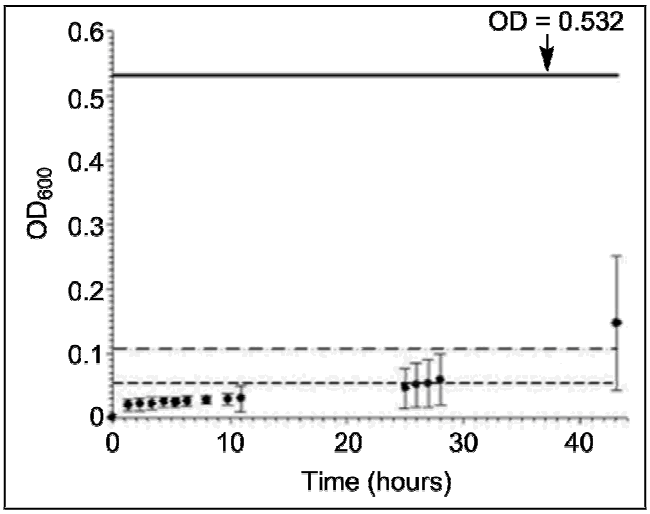


Figure 3.3: Optical density of SCB-containing biotransformation buffer over 45 hours. The maximum optical density value obtained by totally re-suspending the contents of a SCB into the buffer is shown by a solid black line at 0.532. This value represents a 100% dissociated biofilm. The threshold representing 20% cell dissociation is shown as a long dashed line at around 0.1. The OD value representing 10% cell dissociation is shown as a shorter dashed line at around 0.5. All biofilms tested retained at least 80% biomass up to 30 hours of incubation. Measurements taken by Dr Andreas Tsoligkas of the University of Birmingham.

3.2.2 Biocatalytic activity of the spin-coated biofilm (SCB)

3.2.2.1 Performance of SCB with the tryptophan synthase biotransformation

Having determined the stable reaction window with the SCB to be 30 hours, the activity of the biofilm to catalyse the tryptophan synthase biotransformation was explored. 7 day mature spin coated slides were transferred into 70 ml of the biofilm biotransformation buffer (supplemented with 5% DMSO to aid indole solubility). To this haloindole was added to a final concentration of 2 mM. Serine was then added in excess (7 mM) to ensure the concentration of this was not limiting and to compensate for the possibility that viable planktonic cells may consume it for other purposes. The biotransformations were then incubated at the biofilm optimum temperature of 30°C and at the low sheering speed of 70 rpm to maintain biofilm stability during the reaction.

5-fluoroindole was chosen as the substrate for the initial testing of catalytic activity. Fluoroindole is more water soluble than the other 5-haloindoles and, at the 2 mM concentration used, can be totally dissolved in the buffer system. This prevented lumps of indole from floating around in the buffer, which may introduce sheering forces on the surface of the SCB. Previous studies with the cell

free lysate illustrated that 5-fluoroindole is the best haloindole substrate for tryptophan synthase, usually presenting a yield in excess of 80 percent.[117] This high yield was another reason for selecting this indole as the first test of biofilm activity; the reaction rate in this case would be limited by the biofilm catalyst and not by the substrate, therefore acting as a good indicator of activity. An aliquot of the reaction was taken after 24 hours, the reaction stopped by centrifugation to collect bacterial cells, and analysed using the HPLC method established previously (LCgrad_MeOH). The results can be seen in table 3.4.

Haloindole	Individual reaction yields (%) performed in triplicate	Average yield (%) after 24 hours
5-fluoroindole	58,32,26	39 ± 16
5-bromoindole	24,13,10	11 ± 2
4-bromoindole	64,41,20	42 ± 22

Table 3.4: Percentage yield of several haloindoles following initial trials of SCB catalytic ability over 24 hours.

The HPLC analysis revealed the formation of a new peak at around 7.5 minutes that was identified by retention time and PDA spectrum as 5-fluoro-tryptophan, indicating that something inside the reaction, presumably the biofilm, was capable of catalysing the biotransformation reaction. The reaction was conducted in triplicate with three different biofilm specimens but there was very little consistency between the separate films, with the different biofilms converting the 5-fluoroindole in 58, 32 and 26 percent yield.

With the general concept proved, other halogenated indoles were trialled as well. 5-bromoindole was converted to the corresponding 5-bromo-tryptophan in an average yield of 11% (table 3.4). The variation in this reaction was considerably less than with the fluoroindole, showing a standard deviation of only 2%. As a further test, 4-bromoindole was selected as an additional substrate. 4-bromoindole is a liquid at room temperature and this was chosen for a very similar reason as the fluoroindole, the liquid state of this indole meant that no lumps of insoluble indole were present in the reaction that could possibly knock the biofilm off the slide. 4-bromo-tryptophan was produced with an average yield of 42% but again this set of repeats also lacked consistency across the different experiments (± 22%). The 4-bromo-tryptophan yield was intriguing as the only previous paper reporting this biotransformation listed the overall yield of this reaction as 3% (the highest yielding biotransformation with the biofilm in this study reached 64%). It is unclear whether this drastic difference in yield is somehow due to the biofilm or if other factors are responsible. The enzymatic mechanism of the tryptophan synthase αβ dimer suggests that the increased steric hinderance of a 4 or 7-substituted indole ring should mean that indole substituted at these positions yields extremely

poorly, particularly when incorporating a bulky bromine group. Previous work has backed up this theory, which makes this high 4-bromoindole yield particularly anomalous. No further study was conducted with this substrate and the reason for this anomaly was not investigated.

The general lack of consistency with the biofilm biotransformation reactions may be down to several factors. Although different biofilms were calculated to contain roughly the same amount of dry weight there are going to be slight changes in surface structure. The surface of a biofilm is a dynamic system and constantly able to change, this may result in drastic surface area changes that could impact the rate of the biotransformation reaction. The fact that the standard deviation is significantly less with the 5-bromoindole biotransformation may suggest that the faster rate of the 5-fluoroindole reaction is exaggerating small differences in substrate concentration or catalyst amount. To try to alleviate the inconsistent conversions in the initial trials, indole quantities were carefully weighed out on a balance accurate to $1/10^{th}$ of a milligram for the subsequent biotransformation reactions.

A total of 8 spin coated biofilms were used to conduct identical biotransformation reactions using 5-chloroindole as the substrate. Aliquots of the reaction mix were taken over the 30 hour period. Aliquots were centrifuged to remove any cells and frozen at -20°C prior to HPLC analysis. The reactions were conducted on two different preparations of biofilms (4 in each), which were spin-coated on two different days approximately one month apart. The data obtained by HPLC analysis of the results of the two sets of experiments was pooled and the mean, standard deviation and standard error of the mean calculated. The percentage yield of the reaction was plotted against time figure 3.4b. These repeats showed much less variation between the yields with the standard deviation of the 8 samples being at most 10%. This increased reliability may be due to the more accurate measurement of starting materials but can also be related to increased experience of preparing the spin coated biofilms. When initial tests were carried out the spin coating concept was in infancy. By the time these sets of experiments were run the prepared biofilms were more consistent and did not show as much instability as biofilms prepared early in the project.

The 5-chloro-tryptophan time course showed a constant rate of reaction over the entire 30 hours tested. This demonstrates that the biofilm was just as active after the 30 hour time period as it was when it started. This suggests a level of catalytic longevity that would be important if the biofilm concept were to be adapted to a continuous flow system rather than batch.

A repeat of the biotransformation reactions carried out with 5-fluoroindole as the substrate worked much better than the initial tests. Again, this may be due to inconsistency within biofilms generated

early in the project. 4 different biofilms were submerged into identical biotransformation reactions containing 5-fluoroindole and good levels of reproducibility were seen (average standard deviation at each time point was 11%). In this biotransformation almost total conversion of the fluorinated indole to the equivalent tryptophan was seen. The initial reaction rate (figure 3.4a) was observed to be faster than with the 5-chloroindole but the rate decreased after around 24 hours. This is most likely due to the almost complete consumption of starting material during the reaction which lowers the rate. The increased reaction rate of the fluorinated indole could be due to the increased buffer solubility of the fluorinated indole compared to the chlorinated. This would suggest that the solubility of the indole is actually rate limiting with this biotransformation. It is also possible that because the fluorine substituent is significantly smaller than the chlorine, meaning that the width of the resulting tryptophan is less, that the 5-fluoroindole travels more easily through the tunnel between the two tryptophan synthase subunits into the active site than the 5-chloroindole. This faster transfer into the active site could also account for the increased reaction rate.

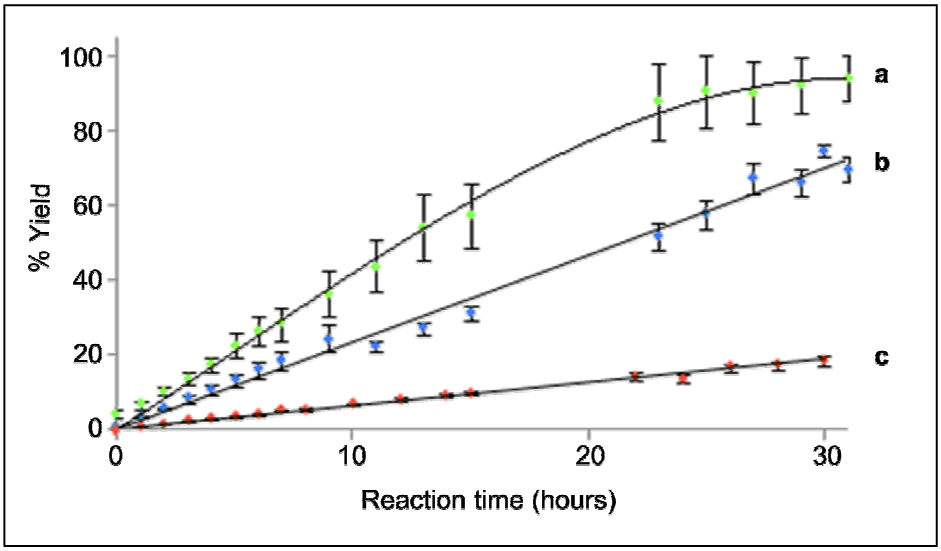


Figure 3.4: Mean yields of biofilm mediated tryptophan synthase biotransformations of the 5-haloindole substrates into the corresponding 5-halo-L-tryptophan. (a) Reaction profile for 5-fluoroindole. Results are the mean of 4 separate biotransformation experiments with individual biofilms. (**b)** Reaction profile for 5-chloroindole. Results are the mean of 8 separate biotransformation reactions with separate biofilms. **(c)** Reaction profile for 5-bromoindole. Points indicate mean of 7 separate biofilm reactions. Error bars indicate the standard error of the mean.

Identical biotransformation experiments were carried out with 5-bromoindole as the substrate (figure 3.4c). In this instance 7 different biofilms grown on two different days were used to catalyse individual reactions. Overall the mean of the biotransformation reactions showed a significantly lower overall yield than the fluorinated or chlorinated indoles tested. This lowering of yield with increasing size of the indole substituent is consistent with previous work conducted with this enzyme.[116] The reaction rate was also much slower. However, like the 5-chloroindole biotransformation, the overall reaction rate did not decrease during the 30 hour window of the biotransformation, suggesting that if left longer the overall yield could be increased. However because of the increased uncertainty about biofilm stability after this time frame and the earlier proof that planktonic cells could also act as effective biocatalysts, any yield achieved after detachment of cells could be the result of planktonic cells rather than the immobilised biofilm bacteria. For this reason the biotransformations remained limited to this 30 hour window.

The seven separate biotransformations with 5-bromoindole showed very low variability between batches. The average standard deviation at each time point was only slightly over 1%. This shows remarkable consistency between different biofilms batches. The formation of biofilm surface topology is dynamic and hard to control; therefore different biofilm batches should show variation in surface area and diffusion rates from the surface of the biofilm into the lower strata of cells. Consistency of the biofilm catalysed yields seems to be related to the total overall yield. As the bromine substituent is exchanged for chlorine and then to fluorine the aqueous solubility of the haloindole increases, this seems to also increase the overall yield and the standard deviation of the repeats. This gives support to the idea that the rate of this biotransformation is limited by the substrate and poorer substrates obscure the differences between biofilm batches.

3.2.2.2 Ability of the biofilm to be recycled

The fact that the spin coated biofilm showed no loss of activity over the period tested suggested that the biofilm biocatalyst could be recycled from one batch reaction to the next without any observed loss in activity. To determine this, a 5-chloroindole biotransformation was performed in the same way as before but for the reduced time of 12 hours. After this time the biofilm coated slide was removed from the reaction and gently washed with buffer before being re-submerged into a fresh reaction of identical set up. This new biotransformation was monitored for an additional 12 hours before a third recycle was initiated. This allowed the potential of recycling the biofilm to be examined without moving significantly outside of the biofilm stability window of 30 hours.

As before, the recycling experiment was performed with multiple biofilms prepared on two separate days, in this case 3 biofilms prepared on one day and 3 on another. Figure 3.5 shows the results of all the repeats combined together. The plots have been lined up to demonstrate how the rate of reaction varies over time.

Remarkably, the rate of reaction does not seem to alter throughout the different recycles. The biofilm catalysed reaction progresses at the same rate in the third reaction as in the first. This conservation of catalytic rate is consistent with the earlier 30 hour biotransformation experiments but is still unexpected as the biofilm slide has undergone two different wash steps before catalysing the final biotransformation. If this is a genuine result then the spin coated biofilm could have great potential for continuous flow processing.

In the earlier biotransformation experiments, however, it was noted that the reactions may be limited by indole solubility, which could be masking changes between biofilm batches. There is a possibility that the same thing is occurring with the recycled biofilms and that any loss of catalytic activity as the biofilm is being recycled may be obscured by the rate limiting indole. Further experiments would be needed to determine this to satisfaction.

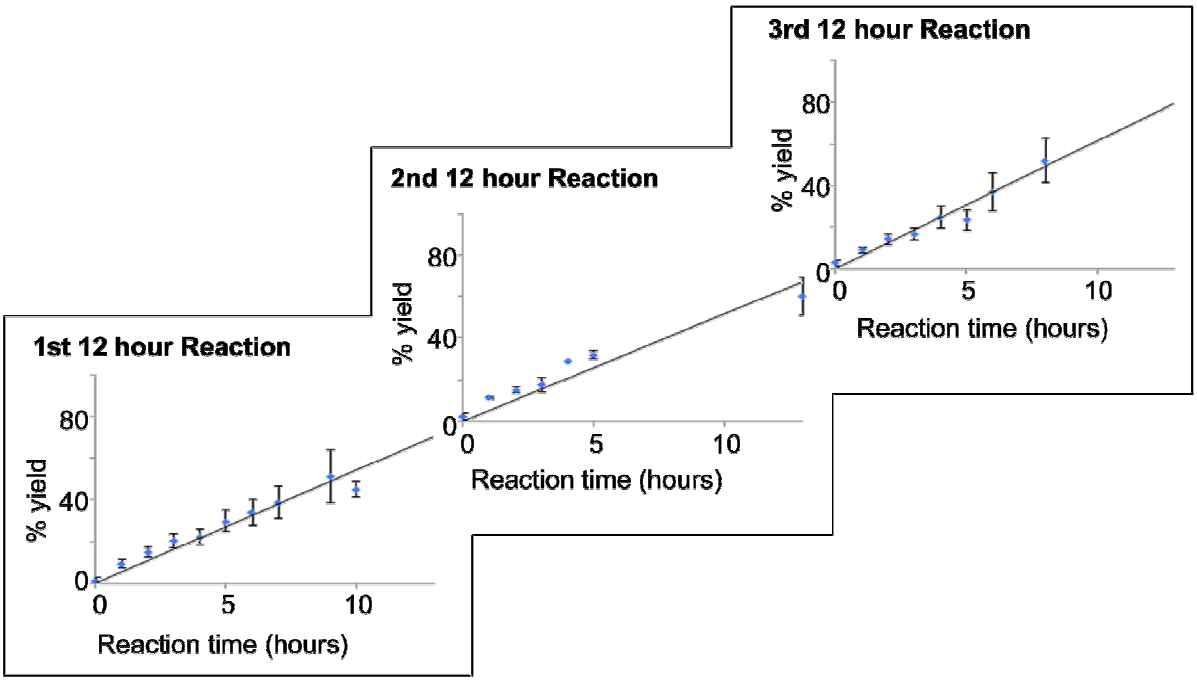


Figure 3.5: Plots showing three sequential 12 hour biofilm biotransformation reactions with 5-chloroindole performed by recycling the same biofilm. Recycling the biofilm into three identical reactions leads to little change in catalytic rate over 36 hours. The error bars indicate standard error of the mean.

3.3 Comparing biofilm biocatalysis to planktonic cells

Before conducting the biofilm reactions, it had been confirmed that planktonic cells were able to catalyse the biotransformation but there was still some doubt over whether the biofilm was doing the same or whether it was in fact sloughing off planktonic cells that were performing the reaction in solution. To determine this and to quantify if the biofilm acts as a superior catalyst compared to planktonic cells, an identical 5-chloroindole biotransformation was performed but with planktonic cells. Prior to doing this, however, the amount of planktonic cells equivalent to the biomass coated on to the slides had to be determined.

3.3.1 Determination of SCB equivalent quantities of planktonic cells

The amount of biofilm biomass present on the surface of a spin coated slide had already been determined to be 80 milligrams by measuring the dry cell weight of SCB slides by heating to a constant mass. This method of biomass determination has been widely used with biofilm research (for a review on biofilm quantification see Lazarova *et al.* (1995)).[134] A planktonic cell suspension was prepared by re-suspending a main culture of *E. coli* PHL644 MW-002 into one tenth of the original culture volume of biofilm biotransformation buffer. The quantity of planktonic cells in this cell suspension was calculated by drying various volumes of this cell suspension until a constant dry mass was achieved.

Figure 3.6a demonstrates the expected linear relationship between volume of cell suspension and dry cell weight. According to the mass results the cell suspension was equal to approximately 10-12 mg of dry cell weight per ml of suspension. Therefore 6-8 ml of cell suspension would contain a biomass approximately equivalent to the amount of immobilised bacteria in the spin coated biofilms.

To make sure the estimation of the quantity of planktonic cells equal to 80 milligrams was as accurate as possible an alternative method was also used. Another common method used to estimate cell mass is to measure the total protein amount, a commonly used estimation is that the total protein content inside cells contributes to almost 50% of the weight of dry cell masses of bacteria.[135] Aliquots of the cell suspension were taken in the same range of volumes as were used for dry cell weight experiments. These aliquots were centrifuged to collect bacterial cell pellets and then the cellular proteins released using a modification of the NaOH protocol used by Li *et al.* (2006).[64]

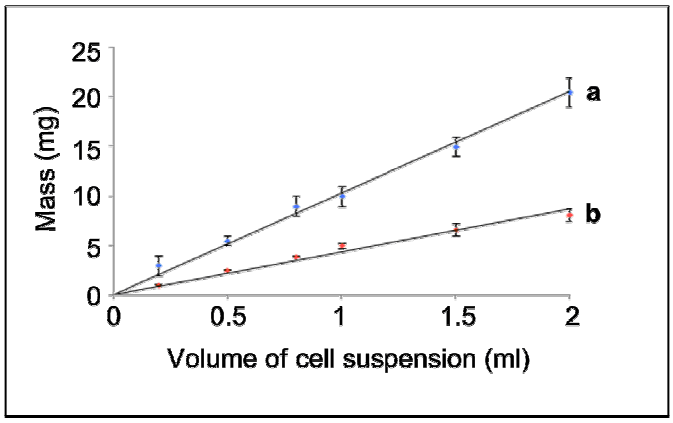


Figure 3.6: Dry cell weight and total protein content of different volumes of *E. coli* PHL644 MW-002 planktonic cell suspension. (a) Mass of various volumes of the planktonic cell suspension, achieved by drying the biomass in an oven until a constant mass was reached. **(b)** For the same volume of cell suspension the bacteria were lysed and the total concentration of contained proteins measured by Bradford type assay. This was used to give the mass of the protein content in the cell suspension. Comparison of the dry cell weight and total protein content showed that the total protein content of *E. coli* contributes roughly 50% of the total dry weight of cells. This is consistent with the literature.[135]

The cell debris was removed using centrifugation and the protein content estimated using a Bradford type assay.

The Bradford type assay relies on the properties of the dye Coomassie Brilliant Blue G-250 (see figure 3.7 for structure). The dye is commercially available premixed with phosphoric acid and ethanol. In these acidic conditions the dye is a brownish colour. With the addition of a protein-containing sample the dye donates its free electron to ionisable groups on the protein. This change in surface charge causes the buried hydrophobic residues to become exposed. These hydrophobic regions interact with the non-polar regions of the dye *via* van der Waals forces. The resulting ionic interactions caused by the proximity of the positive groups of the protein with the negative charge of the dye further strengthens ionic bonding between the protein and the dye. The protein-dye complex causes a shift in the absorbance spectrum from the red/brown form to blue. This can be followed by monitoring the optical absorbance of the solution at a wavelength of 595 nm. Therefore the amount of this complex in solution can be used as a measure of protein concentration.

20 µl of each of the cell free lysates was added to 1 ml of Bradford reagent. Based on the colour change of the samples the lysates were diluted 10x to bring the colour into the linear range of a set of bovine serum albumin (BSA) standard solutions of known concentration (0.2, 0.4, 0.6, 0.8 and 1 mg ml^{-1}) and the absorbance of the unknown lysates and the BSA samples of known concentration

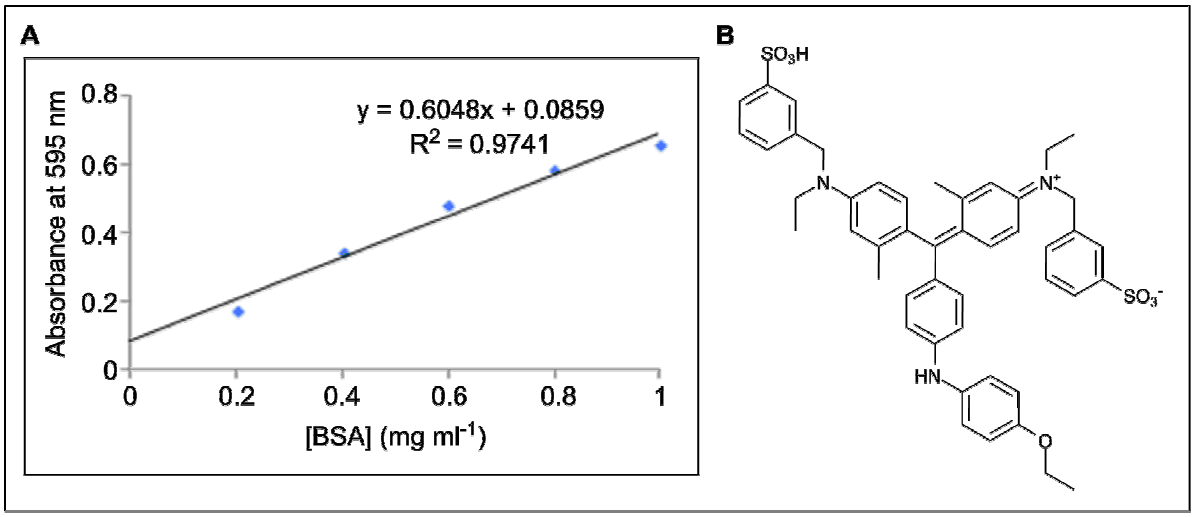


Figure 3.7: Calibration curve of bovine serum albumin concentration versus absorbance recorded at 595 nm after treatment with Bradford reagent. (A) Calibration curve constructed from known concentrations of BSA mixed with Bradford reagent produces increased absorbance in the 595 nm region dependant on protein concentration. **(B)** Structure of the Coomassie Brilliant Blue G-250 dye that is the basis for the bradford type protein assay.

measured at 595 nm. This data was used to construct a calibration curve of BSA concentration versus absorbance. The gradient of this calibration curve was used to determine the unknown protein concentration in milligrams per millilitre for each of the cell lysate aliquots. This concentration was then used to determine the actual mass of protein in each aliquot of original cell suspension with the dilution factor taken into account.

Using this Bradford method it was determined that the total protein mass in 1 ml of cell suspension was 5-6 mg. As can clearly be seen in figure 3.6, this figure is half the amount of mass calculated using the dry mass method. This ratio of protein mass versus dry mass is consistent with values found in the literature.[64]

3.3.2 Tryptophan synthase biotransformation with planktonic cells

The conclusion reached from both complementary methods of dry cell mass determination was that in order to use an equivalent quantity of biomass to the biofilm (80 mgs) with a planktonic cell catalysed reaction then 6-8 ml of cell suspension should be used.

Quantifying the biomass using the dry mass of the biofilm however is a little tricky. As seen earlier when a biofilm forms it encases itself in EPS components (mainly long chained polysaccharides). These EPS components are taken into account when the total dry mass is taken, therefore 80 mgs is probably not an accurate estimation of the actual number of cells. Since the object of using the

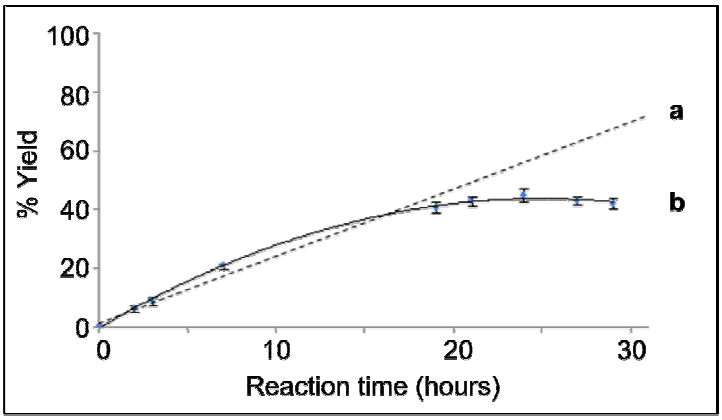


Figure 3.8: Rate of 5-chloro-tryptophan production using 2x equivalent planktonic cells compared to SCB. (a) Dotted line indicates the reaction profile for 5-chloroindole biotransformation carried out with the SCB (see figure 3.4). **(b)** Reaction profile for twice the biomass of planktonic cells compared to the biofilm coated slides. The initial rate is faster with the planktonic cells but all activity ceases after approximately 15 hours. Error bars show the standard error of the mean.

planktonic cells was to demonstrate the improvements offered by biofilms, twice the calculated quantity of planktonic cells were used in the subsequent biotransformations in order to compensate for any errors within the mass quantifications. Due to the presence of EPS components, this was considered to be a very generous approximation.

A 70 ml biotransformation reaction was set up using identical conditions to the biofilm mediated biotransformations. 12 ml of an *E. coli* PHL644 MW-002 planktonic cell suspension, in biofilm biotransformation buffer, was added to a further 58 ml of the buffer. This was supplemented with 2 mM 5-chloroindole, 7 mM serine and 5% DMSO and incubated under identical conditions to the biofilm mediated reactions. Samples were taken every hour for the first 7 hours and then periodically thereafter, any reaction was halted by centrifugation to remove any cells. The aliquots were analysed by HPLC to follow the production of halotryptophan during the reaction. The reactions were carried out in triplicate.

The reaction profile for the planktonic catalysed biotransformation is shown in figure 3.8b, the previous reaction profile for the same biotransformation reaction performed using the SCB is highlighted with a dotted line (figure 3.8a). The initial rate of reaction with the planktonic cells is slightly greater than the equivalent rate with the biofilm, particularly within the first 10-15 hours. After this point, however, a sharp decline in reaction rate can be seen, and no further reaction seems to have occurred after 20 hours. This contrasts with the biofilm, where no loss in catalytic activity of the cells could be seen over the entire 30 hour range. Despite the presence of at least twice the biomass, the planktonic catalysed reaction did not produce a corresponding doubling of

the reaction rate. In fact the initial rate of the planktonic cell reactions was not significantly higher than with the biofilm which was an unexpected result. This could indicate a rate limiting factor that is independent from catalyst concentration, the poor aqueous solubility of the indole substrate may be responsible.

Overall the evidence from these biotransformations suggest that the spin coated engineered biofilm is capable of functioning as a much better immobilised biocatalyst than planktonic cells. The source of this advantage comes from the increased duration of catalysis with no observable loss of activity over the time window. Questions remain, however, as to the exact cause of this longevity. It may simply be down to the fact that biofilm entrapped cells are better able to protect a fragile protein from harsh environments. In the case of this biotransformation the lack of available carbon sources in the reaction mixture is a possible reason for a drop in bacterial viability. The viability of the planktonic cells following the biotransformation is something that should be studied in the future to ascertain if this is responsible for the dropping catalytic rate.

The longevity seen with the spin-coating technology demonstrates the potential for transferring the biofilms from batch processes to a continuous flow. Further study is required to demonstrate whether biofilm mediated biocatalysis would confer similar advantages to other biotransformations in addition to the tryptophan synthase work shown here.

3.4 Comparison to traditionally immobilised enzyme

3.4.1 Immobilising tryptophan synthase using Ni-NTA resin

To fully justify the advantages of using a spin-coated biofilm as an immobilised biocatalyst it was deemed necessary to compare the activity of the SCB to a more traditional form of catalyst immobilisation. There are many different methods of enzyme immobilisation but a quick and simple method of immobilising the tryptophan synthase enzyme onto a support was to utilise an affinity tagging approach. Affinity tagging is an initial method of protein purification and a common technique is to install a hexa histidine-tag on the N- or C- terminus of the target protein. The resulting protein can then be isolated by mixing with Ni-NTA (nitrilotriacetic acid) resin.

Ni-NTA resin consists of a nitrilotriacetic acid support (a tetradentate chelating absorbent developed at Hoffmann-La Roche) that can chelate around ions of nickel. The NTA occupies four of the six available ligand binding sites in the coordination sphere of the ion (see figure 3.9) which forms a very stable connection that can retain the ion under a wide variety of conditions. The remaining two

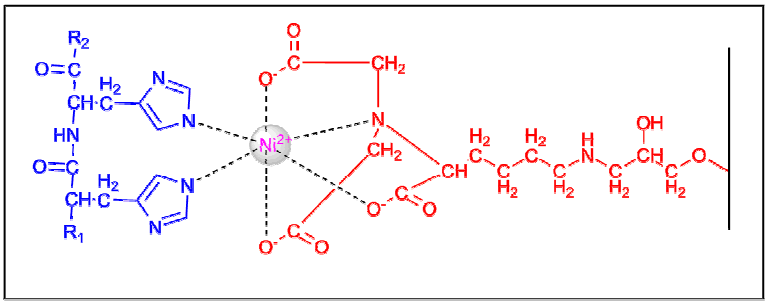


Figure 3.9: Mechanism of histidine chelation to Ni-NTA resin. The Nitrilotriacetic acid (NTA) support coordinates with the Nickel at four of the six available ligand binding sites (in red). The two remaining ligand binding sites in the coordination sphere of the nickel ion are free to interact with histidine residues contained within the hexa histidine-tag engineered onto the target protein (in blue). Other non tagged proteins can then be washed from the resin leaving only the protein of interest.[136]

ligand binding sites are free to interact with the multi-histidine tag present on the target protein. The general protocol for Ni-NTA protein purification involves the addition of Ni-NTA resin to a solution containing a mixture of cellular proteins and the required tagged protein. The removal of any unwanted proteins in solution is achieved by washing the resin with an increasing concentration of imidazole which helps to remove any additional proteins that bind weakly to the chelating resin (the extra imidazole competes with histidine for binding). Once all unwanted proteins have been removed a wash containing high concentration of imidazole (around 300 mM) elutes the protein from the Ni-NTA to leave a solution containing the target protein in fairly high purity. If this elution step is omitted this technique can be used to simply immobilise a target protein onto a solid support (the Ni-NTA). This technique is not widely used as the expense of Ni-NTA purification resin makes it impractical to use on a large scale as well as the environmental impacts of using large quantities of nickel. There is however some precedent of immobilising enzymes onto functionalised surfaces using the properties of histidine tags.[137, 138]

A potential problem with tryptophan synthase, however, is the fact that it consists of two distinct subunits (α and β). Immobilisation of both may restrict the association between the two but on the other hand the association between the two subunits may not be strong enough that both will be immobilised if only one is tagged.

3.4.1.1 Ni-NTA immobilisation strategy 1

Previous experiments had used the commercially available tryptophan synthase containing plasmid pSTB7. To enable effective cloning and the straightforward insertion of a poly histidine tag it was

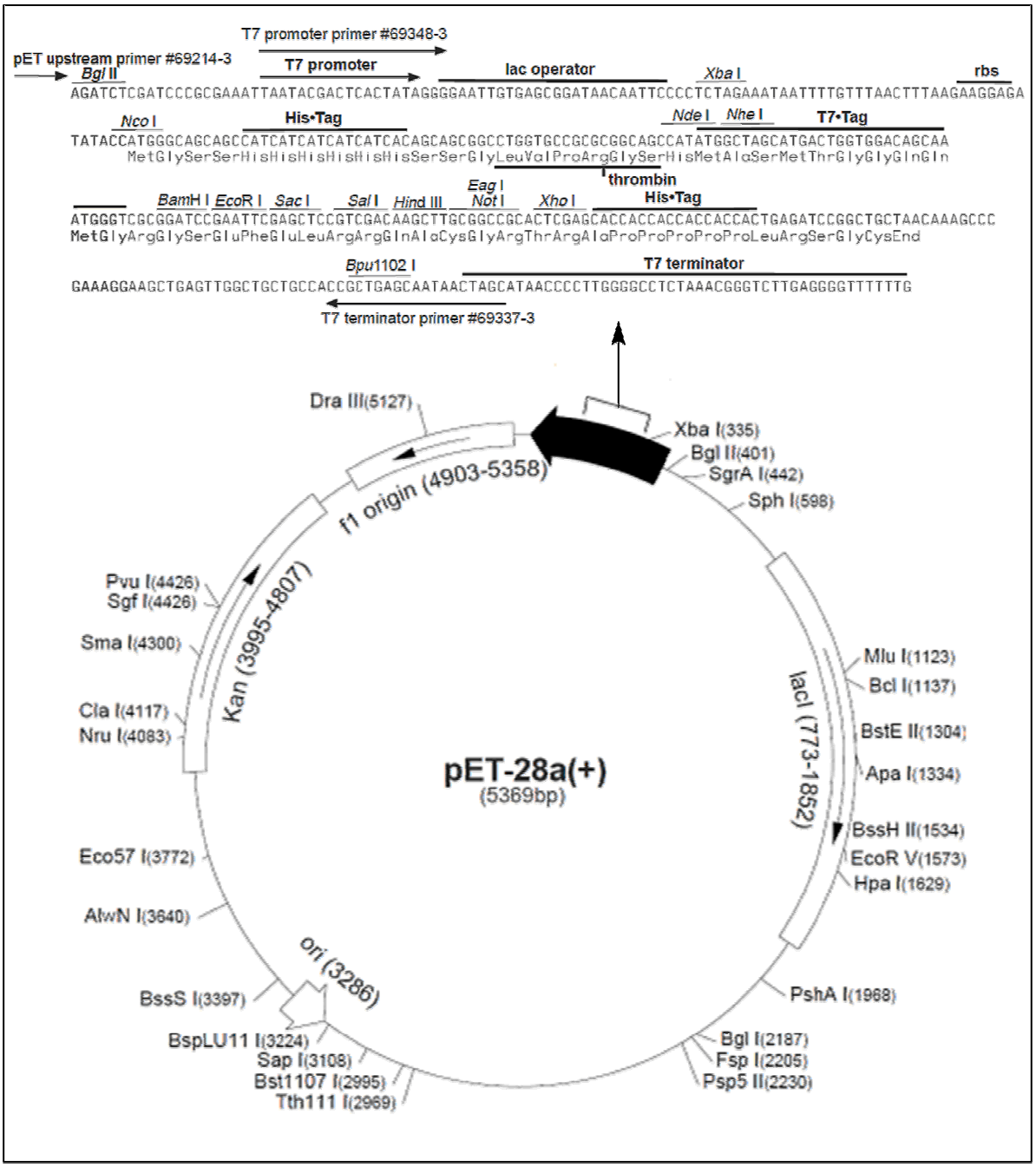


Figure 3.10: Plasmid map and annotation of multiple cloning site of pET28a(+). Diagram of the main features of the pET28a(+) plasmid (Novagen) including cloning/expression region. The MCS region features an N-terminal 6-His tag and an optional C-terminal 6-His Tag (out of frame). The insertion of a gene product of choice is aided by a selection of restriction enzyme sites. Protein production is controlled by the T7lac promoter.

decided to place the tryptophan synthase enzyme into the commercially available translational vector pET28a(+) (Novagen). This plasmid contains a multiple cloning site to enable the insertion of target genes easily and an N-terminal poly histidine tag to facilitate protein purification (see figure 3.10). A C-terminal tag is also available but is presented outside the frame of translation meaning extra bases have to be inserted to use it. Several cloning strategies were then undertaken to attempt to produce histidine-tagged tryptophan synthase to facilitate immobilisation.

3.4.1.1.1 N-terminal Ni-NTA immobilisation of tryptophan synthase β-subunit

The first strategy attempted was a quick and easy restriction digest and ligation approach. The pSTB7 plasmid contains several parts of the trp operon. The important parts are the promoter region and the structural genes *trpA* and *trpB* that code for the α and the β subunits of the enzyme. The trp repressor and attenuator regions that normally negatively regulate the operon are absent, meaning that production of genes under the trp promoter is constitutive. The pSTB7 plasmid map (figure 3.11) highlights that the two tryptophan synthase structural genes are flanked by an EcoRI restriction site and a HindIII site. These sites can be used to remove a 1.7 kb fragment containing

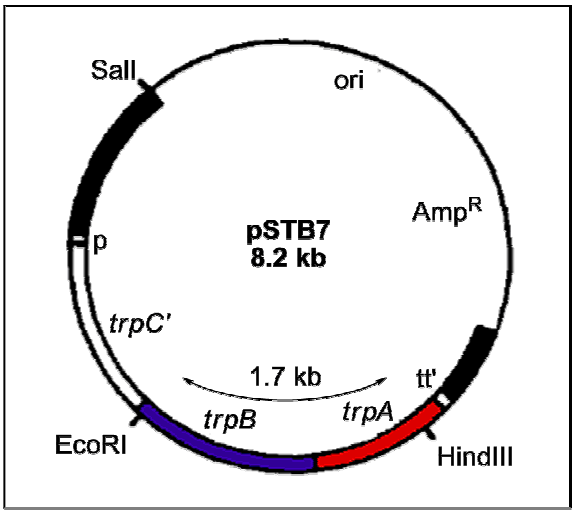


Figure 3.11: pSTB7 plasmid map. Plasmid pSTB7 is a derivative of pBR322 containing part of the tryptophan operon: the trp promoter (P), truncated *trpC* (C'), the *trpB* (shown in blue) and *trpA* genes (shown in red), and the terminator (tt'). The trp attenuator, *trpE*, *trpD*, and part of *trpC* genes have been deleted. The lack of trp repressor and attenuator sequences removals all negative regulation from the operon, effectively making translation constitutive. Plasmid map adapted from Kawasaki (1987).[113]

most of the two tryptophan structural genes. The two sites however fall slightly internal to the genes.

Analysis of the sequence of the two genes (figure 3.13) reveals that the EcoRI site falls 31 base pairs into the *trpB* gene. The protein secondary structure was predicted using the UCL bioinformatics online server (PSIPRED) and the published tryptophan synthase crystal structure studied[95] (this information is featured in figure 3.14). This revealed that the restriction site on *trpB* translates into E11 and F12 of the protein, which occur before any emergence of secondary structure. This means

that the first ten amino acids that are lost when the gene is digested with EcoRI are unlikely to play an important role in the structure and function of the subunit.

The restriction site within the *trpA* gene however appears 595 base pairs into the sequence and over 200 base pairs before the end. The site translates into K199 and L200 which in both the secondary structure predictor and the crystal structure can be seen to form part of an α-helix. Large amounts of secondary structure would be lost as a result of this digestion. However the biotransformation step is performed by the beta subunit and the alpha subunit is not used; therefore any loss in activity of this subunit may not be important. The two subunits are allosterically linked however, with binding of substrates to one subunit triggering conformation changes in the other. The loss of significant portions of the C-terminal region of the alpha subunit may interfere with this pattern and reduce activity. The crystal structure shows that the C-terminus of the alpha subunit is away from the interface of the two subunits so binding of the two subunits should remain undisturbed

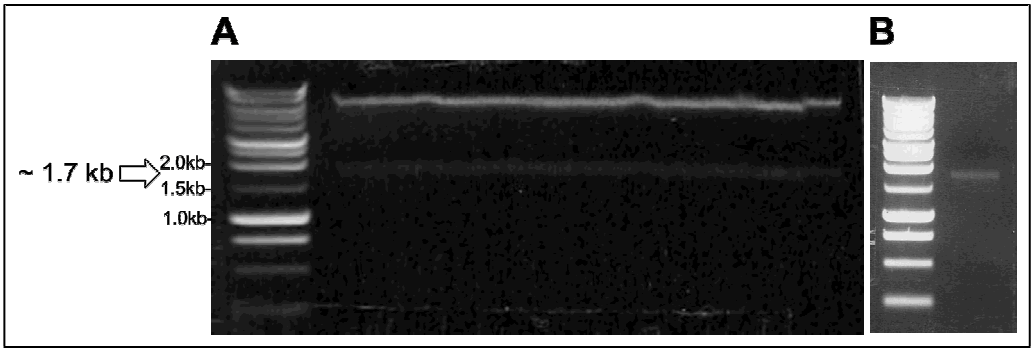


Figure 3.12: DNA gel electrophoresis of pSTB7 and tryptophan synthase structural genes *trpA* and *trpB*. (A) Purification gel of EcoRI/HindIII digested pSTB7 shows original 8.2 kb plasmid separated into the removed 1.7 kb section of *trpB* and *trpA* and the remaining 6.5 kb linearised plasmid. (B) Electrophoresis gel showing purified *trpB/trpA* section. Band sizes from top: 10000,8000,6000,5000,4000,3000,2500,2000,1500,1000,750,500,250bp.

Figure 3.13: DNA sequence for the *trpB* and the *trpA* gene inside pSTB7. The restriction sites used to remove the genes from pSTB7 are highlighted. (EcoRI site inside *trpB* and a HindIII site inside the *trpA* gene). The position of the two genes in relation to each other in pSTB7 and the direction of translation (highlighted with black arrows) is shown.

Despite the possible draw backs the method was trialled regardless due to the quick and easy methodology. Plasmid DNA was purified from the plasmid containing *E. coli* CB149 host strain (MW-043) and restriction digested with EcoRI and HindIII enzymes (incubated at 37°C overnight). The resulting mixture was run on a purification agarose gel and the band that corresponded to the 1.7 kb size of the excised fragment was purified from the gel (figure 3.12).

The excised gene fragment contained DNA overhangs that were used to ligate the *trpB/trpA* fragment into corresponding overhangs of EcoRI/HindIII digested pET28a(+) to form the recombinant plasmid pMW17. This plasmid was then transformed into *E. coli* BL21 to form strain MW-030.

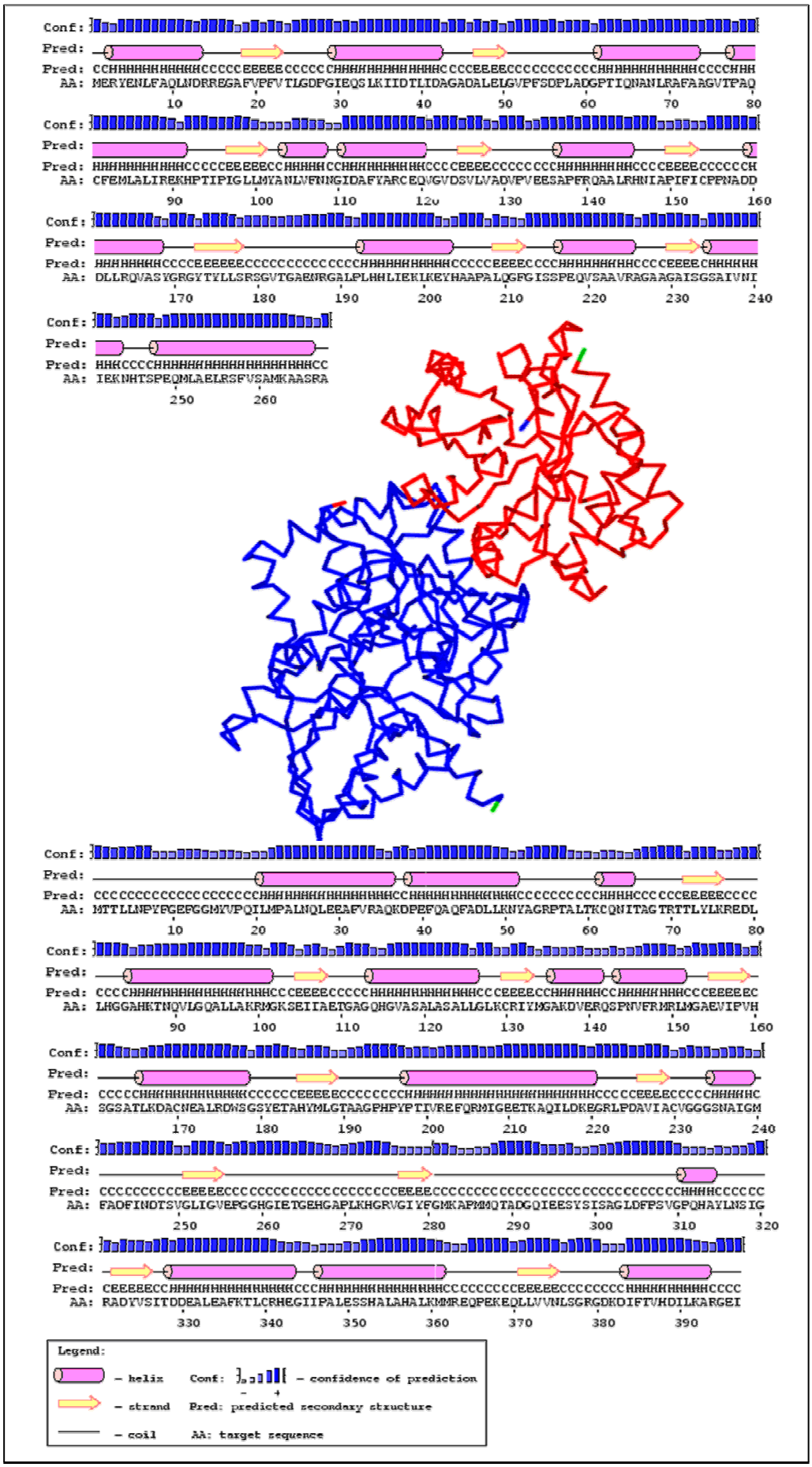


Figure 3.14: Prediction of secondary structure based on amino acid sequence and crystal structure of the tryptophan synthase αβ dimer. Secondary structure prediction based on amino acid sequence is shown. The crystal structure of tryptophan synthase dimer is also shown. N terminal regions are shown in yellow. C-terminal regions are shown in green. (structure from Hyde 1988)[95]

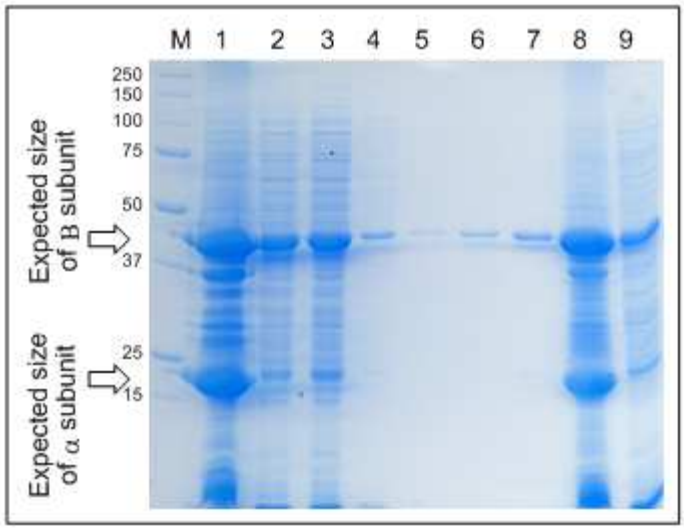


Figure 3.15: SDS-PAGE analysis of production of tryptophan synthase from *E. coli* BL21 MW-030 induced at 37°C and tryptophan synthase production and Ni-NTA purification at 16°C. (M) Precision plus protein marker (Bio-Rad) sizes shown in KDa. **(1)** Cell pellet of culture grown at 16°C. **(2)** Cell lysate of 16°C culture. **(3)** Ni-NTA resin flow through. **(4)** Ni-NTA lysis buffer wash 1. **(5)** Ni-NTA lysis buffer wash 2. **(6)** Ni-NTA wash buffer wash. **(7)** Ni-NTA elution. **(8)** Cell pellet grown at 37°C. **(9)** Cell lysate grown at 37°C. Expected trpA gene product at 21.7 kDa. Expected trpB gene product at 42.9 kDa.

Two 10 ml LB cultures of this strain were grown at 37°C until the OD_{600} reached 0.6 (mid log phase of growth). Then protein production under the control of the T7 promoter was induced with the addition of 0.1 mM (final concentration) IPTG (isopropyl β-D-1-thiogalactopyranoside). One of the cultures was then switched to 16°C and incubated for 24 hours, while incubation of the other was continued at 37°C for 4 hours. Following incubation the bacteria from both samples were pelleted and lysed using chemical cell lysis buffer containing lysozyme (2 mg ml^{-1}). The histidine tagged tryptophan synthase produced in the culture grown at 16°C was purified from the resulting supernatant using the Ni-NTA purification system and analysed by SDS-PAGE (figure 3.15).

At both temperatures tested the tryptophan synthase enzyme was produced in good yield. Both showed reasonable levels of soluble protein in the cell lysate, although the lower temperature did produce more protein over all. The 42.9 kDa beta subunit appeared to be produced at a higher level than the smaller 21.7 kDa alpha subunit (the full size of the alpha subunit before truncation would have been 28.6 kDa). Ni-NTA purification on the cell lysate produced at 16°C showed the beta subunit being purified independently of the alpha subunit (figure 3.15, lanes 4-7). It had been hoped that the attachment between the two subunits would be strong enough to co-purify both parts of the enzyme together.

The expression plasmid pET28a(+) produces recombinant proteins with an attached N-terminal poly histidine tag. Closer examination of the crystal structure of tryptophan synthase (figure 3.14) shows that the N-terminal region of the beta subunit (shown as a single red residue in the figure) is angled upwards towards the alpha subunit. Either the attachment of the two subunits is not strong enough to enable co-purification or it is possible that by installing extra sequence at this end the binding of the two subunits is being inhibited.

3.4.1.1.2 Biocatalytic activity of Immobilised β-subunit

To test activity of the purified beta subunit, a larger culture of *E. coli* BL21 MW-030 was grown, protein production under the T7 promoter induced with IPTG and incubation continued at 37°C for 4 hours. The cell free lysate was generated by sonication and 5 ml was kept aside. The remaining lysate was subjected to Ni-NTA purification and the beta subunit purified as before.

Figure 3.16 shows the Ni-NTA purification of the beta subunit as followed by SDS-PAGE analysis. The end result looked on the SDS-PAGE to be large amount of the tryptophan synthase β-subunit in fair purity. Consistent with the previous small scale studies, no real evidence of any associated alpha subunit was seen. Estimation of the protein concentration provided by a Bradford type assay suggested that after buffer exchange into HEPES protein storage buffer the concentration of the beta subunit was 6.26 mg ml^{-1} (146 µM).

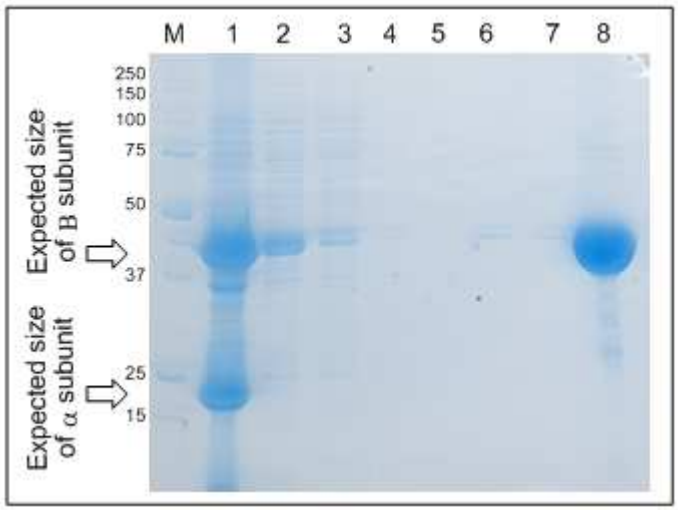


Figure 3.16: SDS-PAGE analysis of Ni-NTA purification of tryptophan synthase from *E. coli* BL21 MW-030 induced at 37°C. (M) Precision plus protein marker (Bio-Rad) sizes shown in KDa. **(1)** Cell pellet. **(2)** Cell lysate. **(3)** Ni-NTA column flow through. **(4)** Ni-NTA lysis buffer wash 1. **(5)** Ni-NTA lysis buffer wash 2. **(6)** Ni-NTA wash buffer wash 1. **(7)** Ni-NTA wash buffer wash 2. **(8)** Ni-NTA resin elutant. Expected trpB gene product at 42.9 kDa.

The activity of the isolated beta subunit, in catalysing the tryptophan synthase biotransformation, was compared to the activity of the 5 ml of cell lysate that had been kept aside before purification. The cell lysate should contain both subunits, while the purified protein just contains the beta subunit. In this set of experiments the aim was to determine whether the binding of the alpha and beta subunits was being prevented by the histidine tag or whether the two subunits were just not associated strongly enough to co-purify on Ni-NTA. If the enzyme within the lysate is active but the purified beta subunit shows no activity then this would suggest that both subunits are required for catalysis and that the histidine tag is not preventing the association of the two subunits within the lysate. If neither show activity then the extra affinity tag may be destroying activity.

Two sets of identical biotransformation reactions were performed in duplicate. The reactions contained 5-chloroindole (2 mM) in 50 ml of biotransformation reaction buffer (containing 7 mM serine and PLP), supplemented with 5% DMSO. To the first set of experiments, 2.5 ml of the cell lysate was added, 70 µl of the purified beta subunit was added to the second (the protein content calculated to be roughly similar). The biotransformation reactions were incubated at 37°C for 24 hours. Aliquots of both reactions were analysed by HPLC. No 5-chloro-tryptophan production was seen with either set of experiments.

This lack of activity with both the pure enzyme and the lysate suggested that it was not the purification process that was destroying activity. The most likely explanation is that the attachment of the poly-histidine tag is responsible. The cause of this could be that the long tag interferes with the interface site of the two subunits and that the multi-subunit enzyme (tryptophan synthase is generally found as an $\alpha_2\beta_2$ tetramer) cannot form, resulting in subunit monomers and a drop of catalytic activity as a consequence. Alternatively the affinity tag may be directly influencing activity by interfering with the conformation of the protein or by blocking the indole substrate tunnel that is nearby. The lack of activity may also be due to the truncation of the N-terminal region of the beta subunit, although the likelihood of this explanation is reduced due to the small amount of sequence lost and the lack of secondary structure at this position. If the alpha subunit is essential to activity of the enzyme then the more significant loss of C-terminal sequence information from the alpha subunit may be to blame.

To test whether the activity of the enzyme was just being reduced rather than destroyed the same biotransformation experiments were performed again but with 5-fluoroindole. Previous work with the cell lysate and in this study with the biofilm demonstrates that this indole acts as a very good substrate with yields as high as 90%. However after 24 hours no trace of product formation could be seen with HPLC analysis.

This rapid method of tryptophan synthase purification did not work. Therefore alternative methods of installing a histidine tag onto the beta subunit were explored.

3.4.1.2 Ni-NTA immobilisation strategy 2

3.4.1.2.1 Cloning and histidine tagging of the α and β-subunits of tryptophan synthase

The beta subunit had proven to be inactive after the installation of an N-terminal histidine tag; therefore the option of a C-terminal tag was explored. To do this another pET vector (Novagen) was used: pET21a(+) (see figure 3.17 for plasmid map and MCS diagram). This plasmid uses the same T7lac promoter as pET28a(+) but incorporates a C-terminal hexa-histidine tag instead of an N-terminal.

Primers were designed that flanked the start and the end of the *trpB* gene (TRPSYN-B-F1 and TRPSYN-B-R1) and introduced a HindIII restriction site at the start and an XhoI site at the end of the gene. The gene encoding for the beta subunit was amplified using PCR with purified pSTB7 plasmid acting as the template. The resulting PCR product was digested and ligated into pET21a(+) using HindIII and XhoI restriction enzymes. The resulting recombinant plasmid was named pMW18 and for protein production was transformed into *E. coli* BL21 to form strain MW-031.

As seen previously the alpha subunit potentially plays an important role in the activity of the enzyme, a theory that is backed up by the literature.[107] Therefore two cloning strategies were employed to amplify the *trpA* gene out of pSTB7. The first involved designing primers to flank the *trpA* gene (TRPSYN-A-F1 and TRPSYN-A-R1) and install NdeI and XhoI restriction sites at either end of the PCR product. This was digested and ligated into NdeI/HindIII digested pET28a(+) to form pMW19. This plasmid was transformed into *E. coli* BL21 to form the strain MW-032. The plasmid was also transformed into competent *E. coli* BL21 MW-031 cells to form strain MW-034 which expressed both tagged subunits in different plasmids.

The second approach used a different forward primer (TRPSYN-A-F2) to install an NcoI restriction site at the start of the *trpA* gene. In pET28a(+) the NcoI site falls before the start of the histidine tag (see figure 3.10), therefore when this second version of the gene was ligated into NcoI/HindIII digested pET28a(+) the region of the plasmid containing the histidine tag was lost, producing a protein that is un-tagged. This plasmid was named pMW21 and was transformed into chemically competent *E. coli* BL21 MW-031 (already containing the *trpB* encoding pMW18) cells to form the strain MW-036 that could produce both subunits.

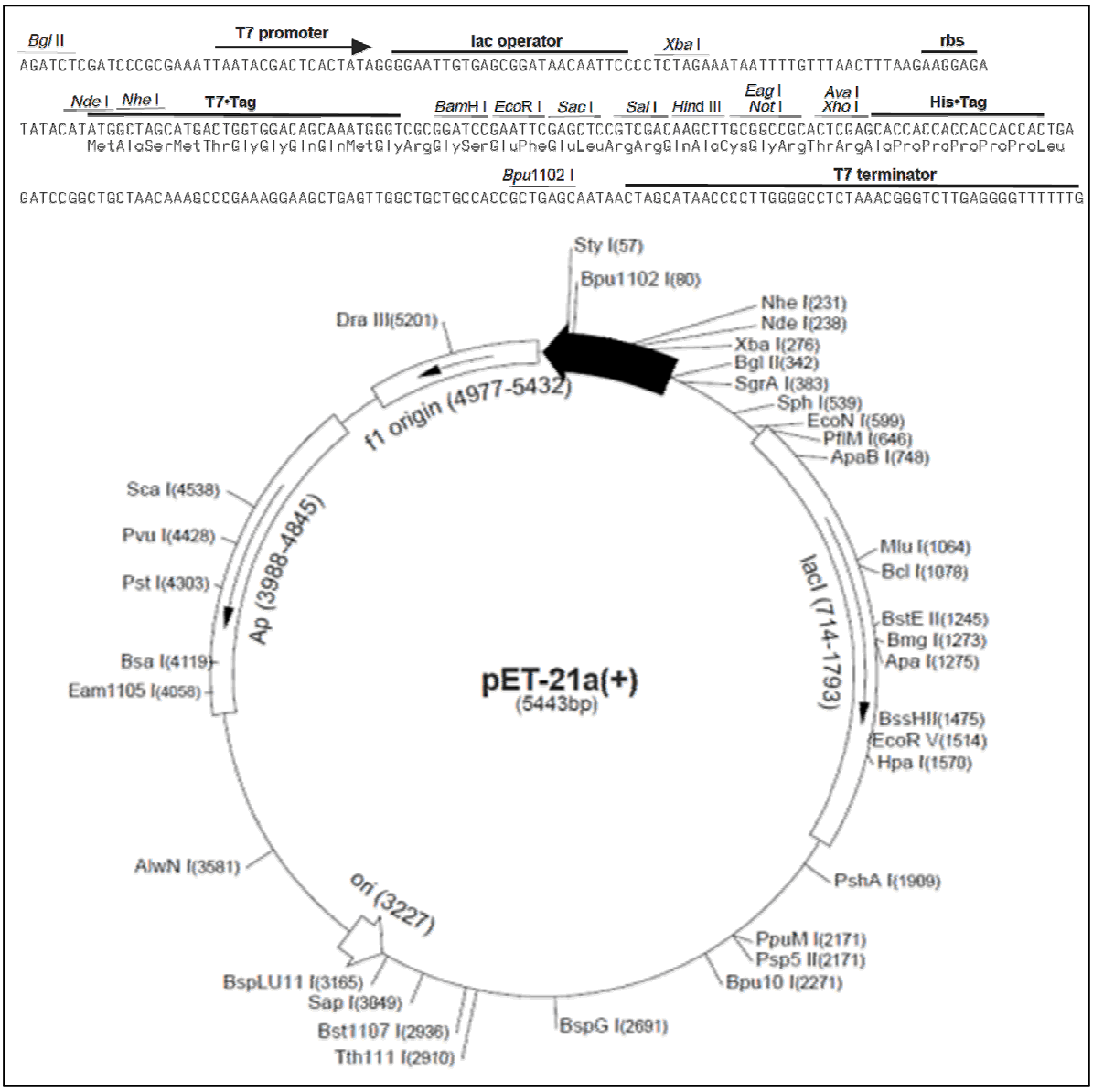


Figure 3.17: Plasmid map and annotation of multiple cloning site of pET21a(+). Diagram of the main features of the pET21a(+) (Novagen) including the cloning/expression region. The MCS region features a C-terminal 6-His tag. The insertion of a gene product of choice is aided by a selection of restriction enzyme sites inside the multiple cloning site (black arrow on plasmid map). Protein production is controlled by the T7lac promoter.

The pMW19 plasmid containing *trpA* with an N-terminal histidine tag could be used to purify the alpha subunit separately if it was not possible to co-purify the alpha and beta subunits together.

3.4.1.2.2 Performance of the isolated C-terminally immobilised β-subunit

Biotransformation reactions performed with the N-terminally immobilised β-subunit had failed to demonstrate any catalytic activity. As explained earlier this may have to do with the localisation of the histidine tag. The strain containing the new C-terminally tagged construct (*E. coli* BL21 MW-031) was grown in 10 ml of LB and protein production under the T7 promoter of pET21a(+) induced in the usual way. Following chemical lysis, small scale Ni-NTA spin purification was carried out on the supernatant (see figure 3.18).

Visualised on SDS-PAGE, the eluent showed two bands at the appropriate size (figure 3.18, lane 5). This anomaly could be due to the appearance of a truncated form of the tryptophan synthase. Tryptophan synthase naturally exists as a tetramer so potentially there is some binding of the native *E. coli* tryptophan synthase which is being carried through the purification. The native *E. coli* beta subunit will not have the additional C-terminal region installed with the tag so would appear to run shorter on the SDS-PAGE. There is also an extremely faint almost invisible band that appears to run slightly above the 25 kDa marker. This could be evidence of some association of native alpha subunit (around 28 kDa) associating with the tryptophan synthase complex.

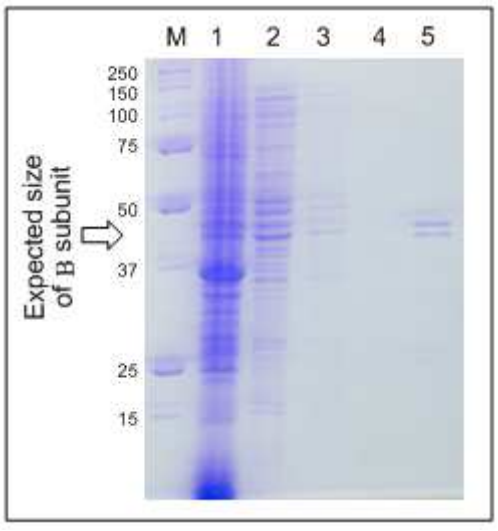


Figure 3.18: SDS-PAGE analysis of small scale NiNTA purification of the tryptophan synthase β-subunit from *E. coli* BL21 MW-031 induced at 37°C. (M) Precision plus protein marker (Bio-Rad) sizes shown in KDa. **(1)** Cell pellet. **(2)** Ni-NTA column flow through. **(3)** Ni-NTA lysis buffer wash. **(4)** Ni-NTA wash buffer wash. **(5)** Ni-NTA resin eluent. Expected trpB gene product at 42.9 kDa.

A 500 ml culture of MW-031 was grown and the cell free lysate incubated with 2 ml of Ni-NTA resin in the usual way. Rather than being purified on a plastic column, the resin was collected via centrifugation following the binding incubation period. The supernatent was removed and the resin washed with Ni-NTA lysis buffer. The resulting suspension was centrifuged again and this time the resin was washed with Ni-NTA wash buffer. A final collection and wash was then performed with the tryptophan synthase biotransformation buffer.

To test the strength of the immobilisation method the resulting resin was re-suspended in 50 ml of biotransformation buffer supplemented with 2 mM of 5-fluoroindole. The suspension was then incubated at 37°C for 24 hours. After this the resin was collected again and washed twice more with Ni-NTA wash buffer before being re-suspended into an identical biotransformation for 24 hours. After this final reaction the resin was collected, washed with Ni-NTA wash buffer and the protein eluted from the resin using Ni-NTA elution buffer. The rate of conversion for these biotransformations was not monitored. The object was to determine whether repeated washing and re-suspending was capable of removing the protein from the resin or whether high indole concentration would have an effect on binding. Samples were taken of all the wash and suspension steps and analysed via SDS-PAGE (figure 3.19).

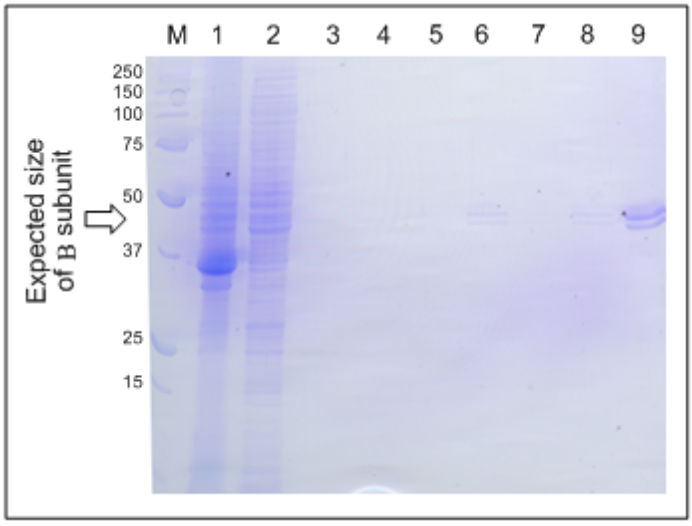


Figure 3.19: Stability of Ni-NTA immobilised tryptophan synthase β-subunit (MW-031) to biotransformation conditions monitored by SDS-PAGE. (M) Precision plus protein marker (Bio-Rad) sizes shown in KDa. **(1)** Cell pellet. **(2)** Cell lysate. **(3)** Biotransformation buffer wash of Ni-NTA resin bound tryptophan synthase. **(4)** Ni-NTA resin wash following 5-fluoroindole biotransformation reaction 1. **(5)** Ni-NTA wash buffer wash. **(6)** Ni-NTA wash buffer wash 2. **(7)** Wash following biotransformation reaction 2. **(8)** Ni-NTA wash buffer wash 3. **(9)** Ni-NTA elution. Expected trpB gene product at 42.9 kDa.

SDS-PAGE analysis of the resin immobilised enzyme showed very little loss of immobilised enzyme during the biotransformations and multiple washes. Some enzyme was lost during the Ni-NTA wash buffer washes (lane 6 and 8, figure 3.19) but the SDS-PAGE is relatively clean until the protein is eluted off the resin (lane 9, figure 3.19). This all indicates that the Ni-NTA resin functions well as an immobilisation scaffold under the conditions of the tryptophan synthase biotransformation.

To make a comparison to the biofilm immobilised biocatalyst the amount of beta subunit on the Ni-NTA resin needed to be quantified. Previous results with the Ni-NTA bound N-terminally tagged subunit had shown a 6 mg ml^{-1} concentration of protein on 1 ml of nickel resin. The quantification was repeated for the C-terminally tagged subunit.

A 500 ml culture was grown again and the cell free lysate incubated with 1 ml of Ni-NTA resin as before. The resin was washed with Ni-NTA lysis and Ni-NTA wash buffers to remove contaminant proteins. The beta subunit was eluted off the resin in 13 ml of Ni-NTA elution buffer. A Bradford type assay and a standard curve of known BSA concentration were used to calculate that the eluent contained 0.532 mg ml^{-1} of protein. This corresponded to 6.917 mg of protein in total eluting from the 1 ml of nickel resin.

3.4.1.2.3 Biocatalytic performance of the isolated C-terminally immobilised β-subunit and comparison to the SCB

The only real way to compare the activity of the immobilised pure enzyme with the biofilm was to make the assumption that 100% of the protein present in the biofilm was tryptophan synthase, obviously an overestimate. The average total protein content of the biofilm was previously calculated to be approximately 40 mg. Therefore 6 ml of Ni-NTA resin would contain equivalent amounts of protein. The cost of Ni-NTA resin is fairly high so a scale down biotransformation was used to keep the resin volume to 1 ml.

A duplicate of biotransformation reactions were performed that had been scaled down ten times compared to the original biofilm biotransformation experiments. 1 ml of Ni-NTA resin (containing approximately 6 mg of tryptophan synthase beta subunit) was prepared as before. The resin was then added to 7 ml of biotransformation reaction buffer. The reactions were supplemented with 350 µl of a solution of 5-chloroindole in DMSO (5.7 mg ml^{-1}). This gave a 2 mM final concentration of indole and a final 5% DMSO concentration.

Aliquots of the duplicates were analysed with HPLC to calculate the rate and yield of the beta subunit catalysed reactions (figure 3.20).

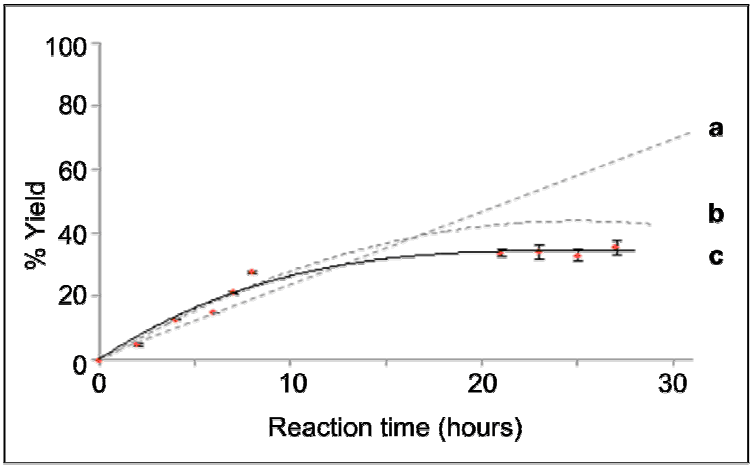


Figure 3.20: Rate of 5-chloro-tryptophan production using comparable amounts of immobilised tryptophan synthase β subunit compared to total protein content of SCB. (c) Profile for the biotransformation reaction as catalysed by the C-terminally immobilised tryptophan synthase beta subunit. Protein concentration (mg ml^{-1} in buffer) is equivalent to the total protein content of the spin coated biofilm. For comparison, traces of the reaction profiles for the same biotransformation carried out by the biofilm **(a)** and the 2x equivalent mass of planktonic cells **(b)** are shown with grey dotted lines.

In contrast to the N-terminally tagged protein, the C-terminally immobilised beta subunit catalyses the biotransformation reaction without the addition of the alpha subunit as can be seen in figure 3.20. This suggests that the localisation of the tag is important to activity. In figure 3.20 the previous data relating to the biocatalytic rate of the spin-coated biofilm and the 2x equivalent dry weight of planktonic cells are shown with dotted grey lines. While the immobilised enzyme seems to initially perform at a similar rate to the 2x equivalent amount of planktonic cells (faster than with the biofilm immobilised catalyst) activity is lost after a fairly short time (approximately 10 hours) and a lower overall yield than the planktonic cells is achieved. Although the total protein concentration of the planktonic cells is roughly twice as much, the problem of direct quantification is again difficult as the assumption is made that all protein inside the planktonic cells is tryptophan synthase, which is clearly false.

After the completion of the biotransformation the Ni-NTA resin was washed with Ni-NTA lysis buffer and the protein was eluted in 20 ml of Ni-NTA elution buffer. The total protein concentration of the eluent was calculated to be approximately 0.37 mg ml^{-1} which made a total of 7 mg of protein on the resin. This value was consistent with the average quantity on the resin before the biotransformation. This demonstrates that the observed loss of activity with the immobilised enzyme was not caused by the enzyme dissociating from the resin. Enzyme denaturation is probably to blame. Another

explanation could be possible product inhibition; in the whole cell systems the product of the biotransformation may be exported from the cells, keeping the internal concentration of halotryptophan low which is not the case for the purified enzyme. Recycling the enzyme into a fresh biotransformation would help to identify whether inhibition or denaturation is to blame for the loss of activity.

The immobilised enzyme seems to be out performed by the biofilm immobilised catalyst. The problem with the direct comparison of the immobilised beta subunit to the biofilm, however, is that the absence of the alpha subunit may be limiting the effectiveness of the enzyme due to the reported drop in catalytic rate experienced when using the isolated beta subunit.[107] Therefore the comparison needed to be done with both subunits of the tryptophan synthase.

3.4.1.2.4 Biocatalytic performance of immobilised α/β subunits of tryptophan synthase and comparison to the SCB

Previously two different plasmids containing variations in the α-subunit had been constructed (pMW19 and pMW21).The first contained the *trpA* gene with an N-terminal histidine tag, the second without a tag. Both plasmids had been transformed into competent cells containing the C-terminally tagged beta subunit. This produced two strains (MW-034 and MW-036) that contained both the alpha and beta subunits inside separate plasmids; one with both subunits histidine tagged and the other with only the tagged beta subunit.

Protein production of both strains was performed in 10 ml cultures which were lysed with lysozyme (2 mg ml^{-1}) following incubation. The cell free extracts were mixed with Ni-NTA resin and small scale spin purification carried out on each. The amount of purified alpha and beta subunits were assessed by SDS-PAGE (figure 3.21).

In both cases the alpha and the beta subunits of tryptophan synthase were purified with the nickel resin. In the case of MW-034, with both of the subunits tagged, significantly higher levels of the alpha subunit were purified than the beta subunit (figure 3.21, lane 9). This could be because this smaller subunit is being produced at a faster rate than the larger beta subunit. As the ratio of α to β is high this means that the majority of the alpha subunit is not bound to a beta subunit making the majority of the purified protein useless. More encouragingly the un-tagged alpha subunit present in cultures of MW-036 seems to be co-purifying with the tagged beta subunit (figure 3.21, lane 4), suggesting that the two subunits are bound together. Similar quantities of both subunits appear to be purified. This makes this construct much more likely to be catalytically active. Both subunits exhibit the same double banding phenomena that were observed with the isolated pure beta

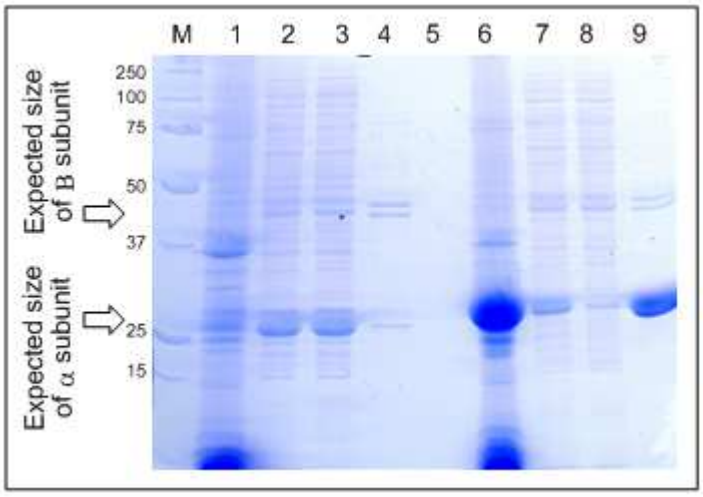


Figure 3.21: SDS-PAGE analysis of production and Ni-NTA purification of both subunits of tryptophan synthase from *E. coli* BL21 MW-034 (lanes 6-9) and MW-036 (lanes 1-4). (M) Precision plus protein marker (Bio-Rad) sizes shown in KDa. **(1)** MW-036 cell pellet. **(2)** MW-036 cell lysate. **(3)** MW-036 Ni-NTA column flow through. **(4)** MW-036 Ni-NTA elution. **(5)** blank. **(6)** MW-034 cell pellet. **(7)** MW-034 cell lysate. **(8)** MW-034 Ni-NTA column flow through. **(9)** MW-034 Ni-NTA elution. Expected trpA gene product at 28.6 kDa. Expected trpB gene product at 42.9 kDa.

subunit. As seen before, it is possible that native *E. coli* tryptophan synthase subunits are binding with the over-expressed recombinant proteins to form the tetramer and are being co-purified with the histidine tagged beta subunit.

The catalytic potential of the immobilised tryptophan synthase containing both subunits was explored in the same way as with the isolated beta subunit; with biotransformations in duplicate. A Bradford type assay was used to determine that approximately 3.7 mg of the tryptophan synthase enzyme was immobilised onto 1 ml of the Ni-NTA resin. The biofilm catalysed biotransformation reactions contained a maximum of 40 mg of protein in 70 ml of reaction buffer leading to a 0.5 mg ml^{-1} concentration of protein. To keep this concentration of catalyst the same with 1 ml of Ni-NTA resin a 10x scale down of the biofilm reactions was prepared with 2 mM 5-chloroindole. The rate of halotryptophan production was followed with HPLC in the standard way.

The rate of production of 5-chlorotryptophan is shown in figure 3.22. The biotransformation with the α/β immobilised enzyme shows an initial rate that is far greater than any of the other systems tested. The multi-subunit immobilised enzyme performs significantly better than the beta subunit immobilised on its own, providing further evidence that although the alpha subunit does not contribute directly to the biotransformation reaction, the allosteric coupling that links the two

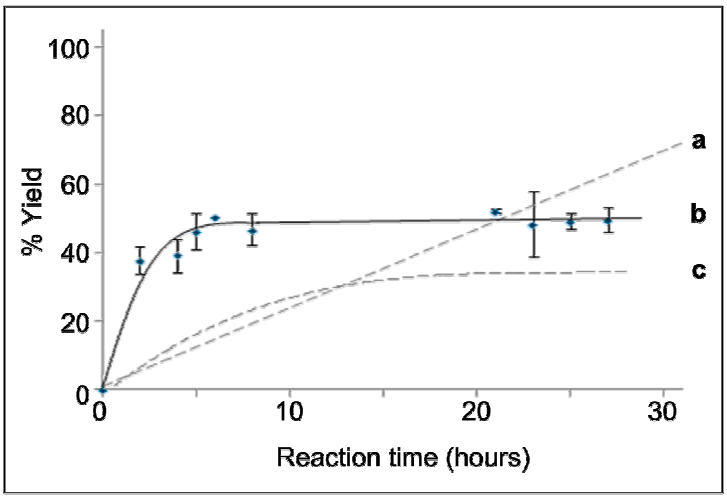


Figure 3.22: 5-chlorotryptophan production mediated by multi-subunit immobilised tryptophan synthase using comparable amounts of immobilised tryptophan synthase subunit compared to total protein content of SCB. (b) Profile for the biotransformation reaction as catalysed by the α/β tryptophan synthase immobilised by the C-terminus of the beta subunit. Protein concentration of 0.6 mg ml^{-1} is equivalent to the total protein concentration of the spin coated biofilm. For comparison the trace of the reaction profile for the same biotransformation carried out by the isolated immobilised beta subunit is shown by the dotted line **(c)** and for the biotransformation performed by the spin coated biofilm **(a)**.

subunits is very important to activity. The initial rate of reaction with the immobilised enzyme however is not long lasting. The rate falls to zero before 10 hours have elapsed, the same finding that occurred with the beta immobilised enzyme. The final yield reached was 49%, better than with the planktonic cells and with the immobilised beta subunit.

As before dissociation of immobilised catalyst from the resin before and during the biotransformation reactions was followed by SDS-PAGE analysis and the amount of protein still contained within the resin following the reactions assessed by a Bradford assay of the eluent (figure 3.23).

Following the biotransformation reaction the buffer was removed and the resin washed with fresh biotransformation buffer. The resin was collected by centrifugation following each wash and the supernatant removed. Following a further two washes with Ni-NTA lysis buffer and Ni-NTA wash buffer, the total protein remaining immobilised to the resin was eluted using Ni-NTA elution buffer. The total quantity of protein remaining on the resin following the biotransformation and subsequent washes was calculated to be an average of 2.5 mg (± 0.1 from the two biotransformation reactions). This is a loss of about 1.2 mg from the average quantity immobilised onto the resin prior to the

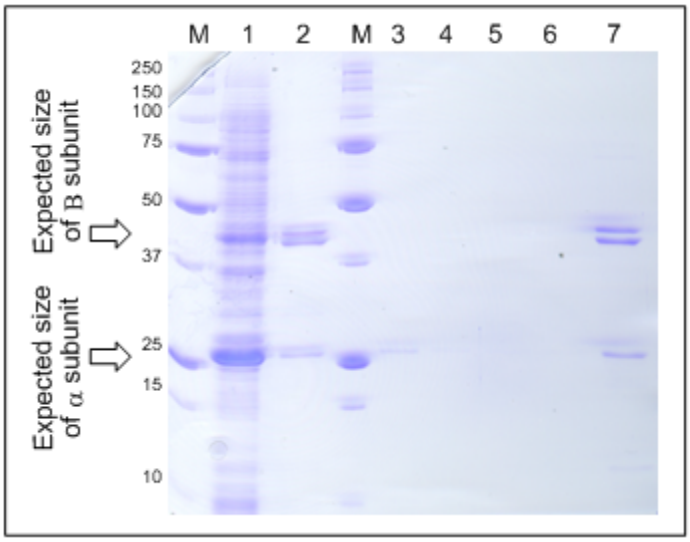


Figure 3.23: Analysis of dissociation of immobilised α/β tryptophan synthase from the Ni-NTA resin during different stages of biotransformation. (M) Precision plus protein marker (Bio-Rad) sizes shown in KDa. **(1)** MW-036 cell pellet showing β subunit expression and significant α subunit production. **(2)** Standard MW-036 Ni-NTA purification elution fraction for comparison shows both the α and β subunits co-purifying on the resin. **(3)** Analysis of biotransformation buffer following the reaction. **(4)** Resin washed with fresh reaction buffer. **(5)** Resin washed with Ni-NTA lysis buffer. **(6)** Resin washed with Ni-NTA wash buffer. **(7)** Post biotransformation elution. Expected trpA gene product at 28.6 kDa. Expected trpB gene product at 42.9 kDa.

reactions. The SDS-PAGE analysis of the reaction and washes showed that some of the alpha subunit was dissociating from the resin during the biotransformation reaction (figure 3.23, lane 3). As already discussed, the association of the alpha and beta subunits is vital to optimum activity of the enzyme. A rough estimation was made, based on the molecular weights of the different subunits that in the enzyme complex the beta subunit represented 60% of the total mass. Using this estimation the loss of 1.2 mg of protein from the resin equates to around 50-60% of the original amount of alpha subunit being lost (assuming no loss of beta subunit). If this is the case then it could explain the loss of tryptophan yield during the reaction. However the counter argument is that at least 40% of the enzyme remained associated in the α/β state and although the beta subunit had been shown to lose activity as an isolated enzyme it would not be enough to explain the total loss of activity after such a short period of time. It is much more likely that the enzyme became denatured and unable to continue to catalyse the reaction. This improved performance of the α/β dimer compared to the isolated β-subunit also makes substrate inhibition a less likely cause of the sudden drop in rate seen with the immobilised enzyme after less than 10 hours as you would expect this to affect both examples equally.

3.5 Biofilm mediated biocatalysis discussion

3.5.1 Analysis of the biocatalytic potential of the SCB

The performance of the artificially engineered, spin-coated biofilm in catalysing the tryptophan synthase mediated biotransformation between indole and serine was assessed and compared to the activity of planktonic cells and purified tryptophan synthase immobilised onto Ni-NTA resin.

Comparison of the catalytic activity of the different methods can be seen in figure 3.24 and table 3.5. It was observed that the biofilm catalysed reactions exhibit much greater longevity than either planktonic cells or immobilised enzyme. This is not without president; biofilms of *Neurospora crassa* involved in the bioremediation of phenol demonstrated a catalytically active period that was eight times longer than the equivalent planktonic cultures.[75] The initial rate of reaction with the immobilised enzyme and the planktonic cells was much greater but during the course of the reaction the biofilm is capable of maintaining a constant rate for much longer.

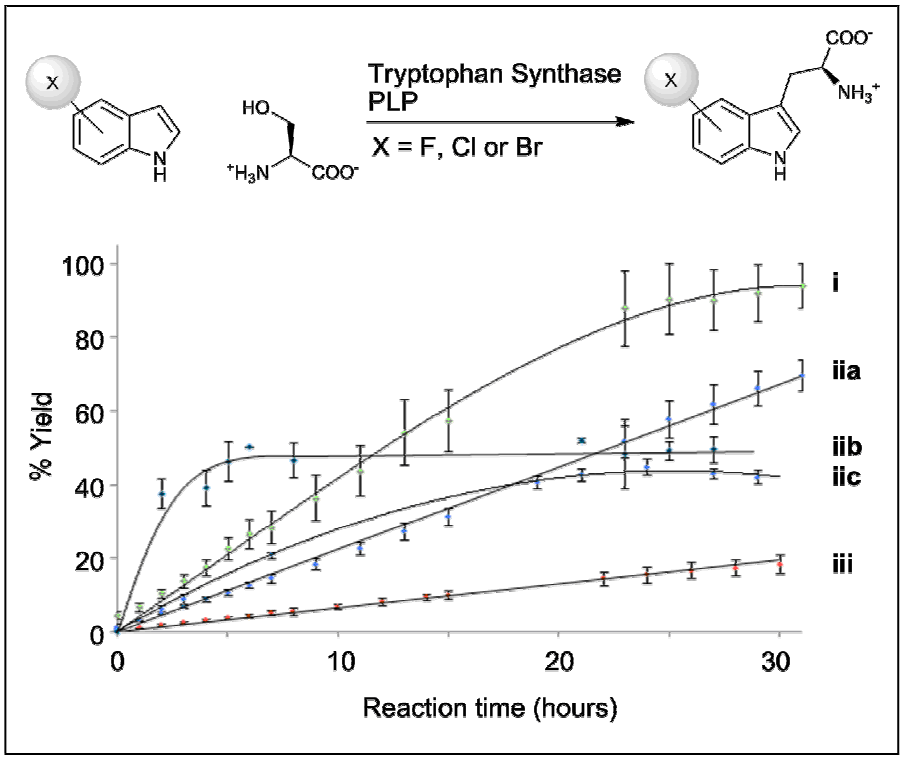


Figure 3.24: Overlay of reaction profiles for tryptophan synthase biotransformations performed by a variety of methods. Reaction profiles for: **(i)** 5-fluoroindole (biofilm catalysed), **(iia)** 5-chloroindole (biofilm), **(iib)** 5-chloroindole (immobilised α/β subunits of trp synthase), **(iic)** 5-chloroindole (incubated with twice the biomass of planktonic cells), **(iii)** 5-bromoindole (catalysed by the spin coated biofilm).

	Catalyst			
	1	**2**	**3**	**4**
Conditions	30 hours, 30°C	30 hours, 30°C	30 hours, 37°C	3 days, 37°C
Relative Protein Concentration	0.6 mg ml^{-1} (total biofilm protein content)	1.2 mg ml^{-1} (whole cell protein content)	0.4-0.6 mg ml^{-1} (tryptophan synthase purified)	0.4 mg ml^{-1} (total protein content)
Tryptophan generated				
5-F	93% conversion			63% conversion[116]
5-Cl	78% conversion[a]	40% conversion[b]	49% conversion[c]	50% conversion[116]
5-Br	18% conversion[a]			16% conversion[116]

[a] Reaction still proceeded at initial rate following 30 hours. [b] No further reaction was observed after 24 hours. [c] No further reaction was observed after 6 hours.

Table 3.5: Comparative yields of 5-halotryptophans generated by different methods. (**1**) spin coated engineered biofilm, (**2**) planktonic cells, (**3**) Ni-NTA resin immobilised α/β subunits, (**4**) enzyme as component of cell-free lysate as demonstrated by Goss *et al.* (2006).[116]

The increased longevity of the biofilm reaction compared to either the planktonic cells or immobilised enzyme is consistent with the evidence that bacteria clustered inside these structures protect themselves from external stresses. The reaction medium itself does not provide any food sources for the bacteria, the drop in catalytic rate with the planktonic cells may simply be caused by loss of cell viability or death. However there are many features of biofilms that could potentially minimise their effectiveness as biocatalysts compared to individual cells and immobilised enzymes, the most obvious is perhaps the dramatic loss of surface area exposed to the substrates.

Biofilms rely on efficient diffusion of substrates through the surface into the lower layers of cells and the removal of waste products in the opposite direction. The chemical heterogeneity that arises inside a biofilm leads to individual member cells existing in one of three environments. Cells on the surface find themselves with an excess of substrates. At the base poor diffusion leads to an accumulation of products, while cells in the middle experience various levels in between.[15] As explained previously, it has been suggested that the formation of the distinctive mushroom shaped colonies often seen on the surface of biofilms are caused by the concentration gradient of substrates present at the interface. The resulting channels between the mushrooms act as substrate highways to move substrates and products from deeper cells. This essentially means that the access of biofilm-bound cells to the haloindole starting material is significantly reduced compared to its individual planktonic counterparts. The same is therefore true of the halo-tryptophan product which has to find its way out of the cell and into the reaction media for collection.

If a sample were taken of the bacteria contained within the planktonic cell suspension a high degree of homogeneity would most likely be seen, with all the cells exposed to the same substrate concentration and oxygen levels, theoretically making them capable of performing the biotransformation reaction at the same level. Studies of biofilms however show great variation in physiological state and the genotypes and phenotypes that are expressed.[15] This is understandable as location within a biofilm will expose cells to position dependent conditions. For instance positions that are replete with substrates and oxygen, replete with substrates but not oxygen and positions neither replete with substrate or oxygen. This lack of homogeneous conditions within the biofilm leads to different growth rates, with surface cells growing more rapidly than buried ones. This can also lead to the formation of persister cells. These cells are in a state of almost total dormancy and have been implicated as playing a major role in biofilm antibiotic resistance and obviously these dormant cells would be unable to function as biocatalysts.

It is for these reasons that the performance of the biofilm as a biocatalyst is so interesting. However it should be noted that the level of heterogeneity within biofilms prepared using the spin coating technology has not been measured. Applying assumptions based on traditionally grown or wild type biofilms is dangerous without further study as the spin-coating process may produce biofilms that are distinct from examples produced in traditional ways. Cell differentiation and persister cells may not be present in our biofilm. However the substrate transport limitations can still be applied with a fair degree of certainty.

During the biofilm catalysed biotransformation experiments it was noted that the lack of 5-chloroindole and 5-bromoindole solubility may have made the reactions substrate limiting. The slower than expected rate of the planktonic cell compared to the biofilm catalysed reactions provided evidence that this may indeed be the case. Further experiments with varied substrate/catalyst concentration would be required to prove this. One problem with this theory however is the huge increase in rate when immobilised, pure enzyme is used. If the indole solubility was truly limiting then we would expect to see a much slower conversion with the immobilised tryptophan synthase. Studies with the immobilised enzyme catalysing the other 5-haloindole in this study would help to pin down the potential solubility issues.

If the indole solubility in the buffer is not directly responsible, an alternative explanation may lie in the mechanism of cellular indole uptake and tryptophan efflux. There may be a limiting step with how the different haloindoles are being uptaken into the whole cells, a problem not applicable to the purified enzyme. Indeed Ni and co-workers have shown that the efficiency of whole cell biocatalysts can be improved by reducing the thickness of the outer membrane, therefore improving

the diffusion of the substrates and products.[139, 140] Tryptophan synthase is a cytosolic enzyme, consequently in both the biofilm and planktonic catalysed reactions the substrates must first enter the cells prior to catalysis and the tryptophan product must be excreted into the reaction buffer in order to be detected and purified. Studies with *Pseudomonas aeruginosa* have suggested that the increased antibiotic resistance seen with biofilms may be due to the observed upregulation of drug efflux pumps.[62, 141] If increased efflux is a general feature of biofilms then this may provide a clue as to why the biofilm seems to be performing better than planktonic cells.

Indole is known to play a significant and diverse role in *E. coli*, particularly as a signalling molecule. In this capacity it must be exported into the cells environment before use. The literature supported view is that the Mtr transporter (a permease normally responsible for uptake of tryptophan) and the AcrEF-TolC multidrug exporter are responsible for indole import and export from the cell;[142] however more recent work concluded that indole transport is actually independent of these proteins. Studies with indole and artificial membranes have shown that indole can rapidly cross an *E. coli* lipid bilyaer without assistance.[143]

The transport of tryptophan in *E. coli* is not so well understood. Import of the aromatic amino acids is known to be controlled by a general aromatic amino acid permease transporter as well as individual specific transporters for the three aromatic amino acids.[144] Specific tryptophan uptake is controlled mainly by the high affinity permease Mtr which is the same transporter implicated in indole import.[145] The mechanisms behind efflux of tryptophan however are less clear. While exporters have been found for many of the amino acids, there is not much known about the transport of the hydrophobic amino acids tyrosine, phenylalanine and tryptophan. The tight regulation of the production of these amino acids *in vivo* is such that specific transport mechanisms are thought to not exist. The hydrophobicity of these amino acids is considered to be large enough to allow general diffusion through the cell membrane without assistance. However recent studies have highlighted the possible involvement of an inner membrane protein in *E. coli* encoded by the gene *YddG*. This gene is a homologue of one found in *Salmonella enterica* sv. Typhimurium which encodes a transporter protein implicated in the efflux of the toxin methyl viologen.[146] Over-expression of this gene in *E. coli* led to extracellular accumulation of the hydrophobic amino acids, particularly phenylalanine.[147] If these transporters are being up regulated in the spin coated biofilms compared to the planktonic cells then this could offer another feasible explanation as to why the biofilm is performing better.

The substrate for the tryptophan synthase biotransformation reactions is indole. Indole is known to be an important extracellular signalling molecule in a range of different organisms, including *E. coli*.[50]

It is generally produced during stationary phase and is known to have an effect on gene expression, multi-drug exporters and pathogenicity. It also plays a somewhat ambiguous role in biofilm formation.[25] Studies using global transcriptome analyses, confocal microscopy, isogenic mutants and dual-species biofilms have shown that indole is a non-toxic signal that represses motility in *E. coli via* the autoinducer-I quorum sensing pathway controlled by SdiA.[25] Motility is very important both for initial biofilm formation and for the migration of colonies inside biofilms along concentration gradients (the causal event behind mushroom colony formation). *E. coli* inside flow cells supplemented with 500 μM indole showed reduced biofilm formation and flat colonies compared to scattered towers found without addition of indole. Conversely however it has also been demonstrated that bacterial species that possess the tryptophanase gene (breaks down tryptophan into indole) show enhanced biofilm forming ability. Di Martino and co-workers found that a strain carrying a mutation in *tnaA*, the gene that encodes tryptophanase, presented a decreased tendency to form biofilms, a feature that was attenuated with the addition of physiological concentrations of indole.[51] Transcriptome analysis of this mutant found repression of seven different genes associated with motility. So in this case indole seems to increase cell motility when fed to this mutant strain. Therefore there seems to be evidence that indole can both repress and enhance biofilm formation depending on concentration but in both instances it is decreased cell motility that actually leads to reduced biofilm formation.

In this study however the haloindole substrate was only added after the biofilm had already reached maturity. This perhaps changes the overall influence indole had on the biofilm. If the haloindole functions as above then perhaps the motility of the *E. coli* was reduced once the biotransformation was underway, therefore promoting biofilm stability rather than inhibiting formation. In order to survive biofilms are able to regulate their own size, effectively producing signals that cause cells to detach in times of community stress such as carbon source limitations. The addition of indole to the buffer may have had the affect of actually stabilising the biofilm and preventing cells from dissociating due to lack of energy. The dynamic nature of the biofilm may also have been inhibited as cells with limited motility would not be able to make further adjustments to the topology of the biofilm. Therefore indole may have actually been an essential component for long term stability. SCBs exhibiting different biotransformation pathways may not be as stable as the experiments presented here, or the indole may have reduced biofilm stability and other biotransformations may exhibit longer lives, further study in this area is required.

There are many features of the spin-coated biofilm that are not fully understood. A full range of further experiments dealing with all of them will in the future answer some of the outstanding questions.

3.6 Addressing the questions – future work

This preliminary work into the development of a versatile biocatalytic biofilm highlights the possible advantages such a system possesses. However many questions have been raised during the project that could be addressed with further experiments.

3.6.1 Understanding and improving the SCB as a catalyst

Many questions still remain unanswered; particularly about how the structure of the spin coated biofilm differs from traditional flow cell formed specimens and also how the SCB manages to remain consistently active for extended periods.

One of the key questions is where in the biofilm does catalysis actually occur? Is the indole substrate only penetrating a few layers down into the strata, or are the cells lower down in a state of dormancy and therefore unlikely to perform as a biocatalyst. In these scenarios only the very top layers of cells would be responsible for the biotransformation. When the cells on the surface (exposed to hostile conditions and no food) begin to die they could be dissociating from the biofilm and being replaced with fresher cells from just below. The dead cells could also be cannibalised to provide energy for the rest of the biofilm. Alternatively is catalytic rate kept alive by regeneration of fresh tryptophan synthase enzyme inside the biofilm?

There are a number of further experiments that could address these points, including methods to study protein turnover/regeneration. GFP tagging can be used to identify where in the biofilm protein production is occurring, also the O_2 requiring property of GFP means that combined with flavin-dependent anaerobic fluorescent proteins (AFPs) the diffusion of oxygen through the different biofilm cells can be measured. Propidium iodide staining could also be used to determine the different cell viabilities throughout the biofilm. Confocal microscopy and substrates with fluorescent properties could be used to follow how far through the biofilm they diffuse. Fluorescent products could also help to identify which areas of the biofilm are actually conducting reactions. Tryptophan itself has fluorescent properties which may make it acceptable for this study.

In addition to trying to understand why the biofilm behaves as a longer lasting catalyst than planktonic cells it would be useful to understand why exactly the planktonic catalysed reactions lose

activity. Flow cytometry on the planktonic cells following the biotransformation reaction could provide evidence to support the idea that conditions within the buffer lead to cell death and the corresponding drop in biocatalysis.

The biofilm mediated reactions tested so far have all been batch systems. The raison d'etre of immobilised catalysis is in continuous flow processes. The potential of the SCB as a catalyst has been demonstrated but to truly reach its potential methods of immobilising the biofilm as a component of flow should be explored. Within this study the biofilms used were limited to a 30 hour reaction window. This window would obviously have to be extended to make a long term biocatalytic solution viable. To achieve this, the reason for the biofilm dissociation following 30 hours needs to be explored. The simplest explanation may be that energy sources and metabolic building blocks are largely absent in the reaction buffer. Pulsing in nutrients at intervals may help to maintain a stable biofilm. This would have to be carefully investigated as too many nutrients may switch the biofilm forming genes off and steer the bacteria back towards a planktonic lifestyle. A range of supports should also be investigated to identify a support that is more appropriate to scaled up immobilisation than the spin coated slide methodology.

One of the interesting findings from the biofilm spin coating analysis was the sudden switch in EPS production that occurred after 6 days of maturation. This switch may be mediated by a number of factors including quorum sensing mechanisms. If the reason for this sudden change can be understood then theoretically it may become possible to artificially induce this process earlier to produce biofilms that mature much more quickly.

3.6.2 Addressing the issues raised by the choice of tryptophan synthase

The choice of tryptophan synthase as the benchmark biotransformation has presented some unique challenges which make accurately determining the biocatalytic effectiveness of the biofilm difficult. The first problem was the solubility of the indole. The low solubility of the 5-bromo and 5-chloroindole raised questions about whether it was rate limiting in the biotransformation. Simple kinetics with the biofilm and planktonic cells would help to resolve this problem. Under normal steady state conditions doubling the concentration of the starting material should result in a doubling of the rate. The same is true if the amount of catalyst is doubled while the substrate concentration is kept constant. If the indole solubility is limiting the reaction than doubling the concentration should have less of an effect on the overall rate. This property has perhaps already been seen, as using double the amount of planktonic cells compared to the biofilm did not result in a

corresponding increase in rate, but more accurate experiments are needed to verify this. This simple set of kinetic experiments should also be performed with the immobilised enzyme. If the pure enzyme does not experience the same rate limitations then factors such as cell uptake or product efflux should be looked at.

Secondly the choice of the tryptophan synthase biotransformation introduced indole as the substrate. As discussed earlier indole is known to play a number of important roles in *E. coli*, particularly important to this study is the effect that indole has on biofilm formation. The choice of tryptophan synthase may have been serendipitous; it is possible that indole is having a stabilising affect on the spin-coated biofilm. Another biotransformation reaction may not show the same level of longevity if indole was not added. Therefore the effect indole has on biofilm stability should be studied and the efficiency of other biotransformation reactions tested.

3.7 Work in progress: potential of other biotransformations: Tryptophan-7-Halogenase

The advantages that the SCB presents to tryptophan synthesis has been presented in this study. To improve the impact of the work however it needs to be proved that the biofilm is capable of catalysing a number of different biotransformations and act as an adaptable scaffold that enzymes can be plugged easily into. To this end, the tryptophan-7-halogenase PrnA from the pyrrolnitrin biosynthetic cluster was inserted into the biofilm. This is another potentially useful biocatalyst candidate which is discussed in more detail in the next chapter.

The pSTB7 plasmid used for the tryptophan biocatalysis was constitutively active, meaning that the protein production switch is always set to ‘on’. The pET vectors utilised in the tryptophan synthase immobilisation experiments are typical examples of expression plasmids in that protein translation is under the control of a promoter that needs to be turned on. Commonly used promoter systems are the T7lac promoter (used in the pET system) and the arabinose induced promoter araBAD. The T7 system cannot be used in every species of *E. coli*; the genome of the host organism needs to encode for the viral T7 RNA polymerase in order for the promoter to work efficiently. The PHL644 strain used to form the SCB does not contain this polymerase gene so it was unknown how well plasmids utilising the T7 promoter would work. To this end the propionate induced plasmid pPro-24(*s*)-gfp was also trialled, the additional advantage of this plasmid was the formation of an N-terminal GFP fusion protein which may prove useful in following protein expression inside the SCB.

Two plasmids that had been prepared by a colleague for a separate but related project were transformed into *E. coli* PHL644 competent cells. The first plasmid was designated pSG22 and was a pET21a(+) based plasmid containing the gene for the PrnA halogenase. This enzyme converts tryptophan into 7-chlorotryptophan. The resulting strain was identified as *E. coli* PHL644 MW-040. Despite showing generally better levels of protein production than pET21a(+), pET28a(+) could not be used with the biofilm as PHL644 contains intrinsic kanamycin resistance which is the selection marker for pET28a(+). The other plasmid was designated as pSG49 which consisted of PrnA inside the pPro-24-gfp propionate induced plasmid. The resulting strain following transformation was identified as *E. coli* PHL644 MW-042.

To investigate protein production inside PHL644 under the T7lac promoter a 10 ml culture of MW-040 was grown in LB and induced with 0.1 mM IPTG once the OD_{600} of the culture reached 0.6. As a control the same volume of culture was grown with *E. coli* strain BL21 MW-039. This strain consisted of pSG22 inside the *E. coli* strain BL21. The effect of adding or removing IPTG was also trialled. BL21 contains the necessary genome to be able to produce proteins under the T7lac promoter efficiently. Following an additional 4 hours of growth the bacterial cells were lysed and Ni-NTA spin purification performed on the resulting cell lysate. The SDS-PAGE analysis of this can be seen in figure 3.25.

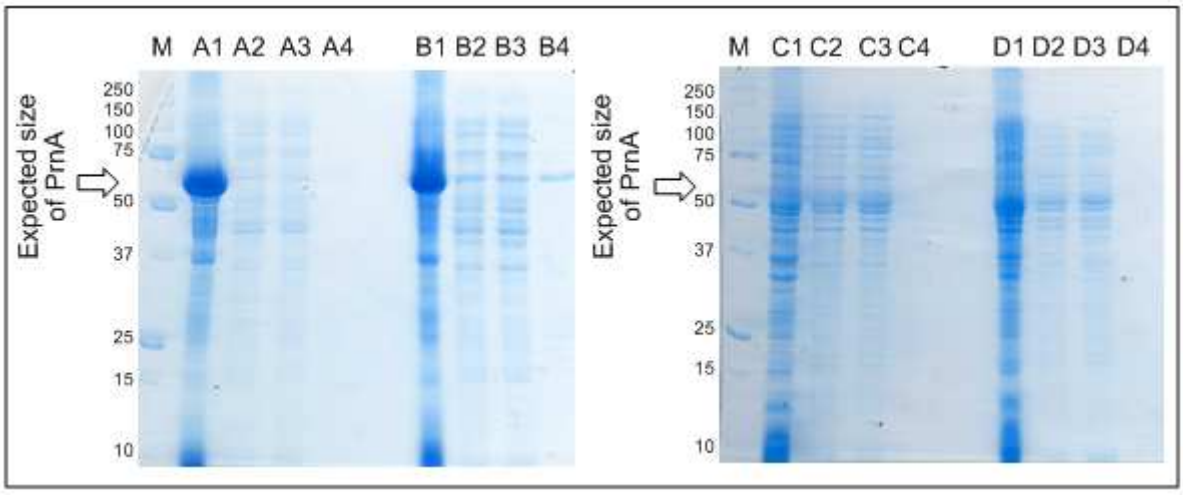


Figure 3.25: SDS-PAGE analysis of PrnA halogenase production under the T7 promoter with *E. coli* BL21 and PHL644 cells with or without the presence of IPTG. (A) Production of PrnA in BL21 cells following induction with 0.1 mM IPTG, **(B)** Production of PrnA in BL21 cells without 0.1 mM IPTG, **(C)** Production of PrnA in PHL644 cells following induction with 0.1 mM IPTG, **(D)** Production of PrnA in PHL644 without 0.1 mM IPTG. **(1)** Bacterial pellet, **(2)** Cell free lysate, **(3)** Ni-NTA column flow through, **(4)** Ni-NTA column elution. PrnA expected size 63.2 kDa.

Protein production using *E. coli* BL21 was consistent with what was expected. Figure 3.25 (A1-4) shows a big band in the pellet at approximately the right size (around 63 kDa). Not much soluble protein is seen and hardly any appears in the Ni-NTA elution (lane A4). The efficiency of small scale chemical lysis is not very high however. The effect of not adding IPTG can be seen to be very minimal. A large band can be seen in the pellet fraction in figure 3.25B1. Interestingly the culture that was not induced with IPTG shows more protein in the elution fraction. This demonstrates that not all promoter controlled protein expression systems are very tight and can leak even when not in the presence of the inductant. The same plasmid inside PHL644 cells (figure 3.25C,D) shows only small amounts of soluble protein being produced with none being visible in the purified fraction. It seems likely on this evidence that the T7 promoter is not ideally used with the biofilm producing strain.

The same experiment was conducted for the propionate induced pSG49 plasmid inside *E. coli* Bl21 and PHL644 cells. The results of PrnA production under the propionate promoter can be seen in figure 3.26. No real protein band could be seen in either strain of *E. coli* in the absence of propionate

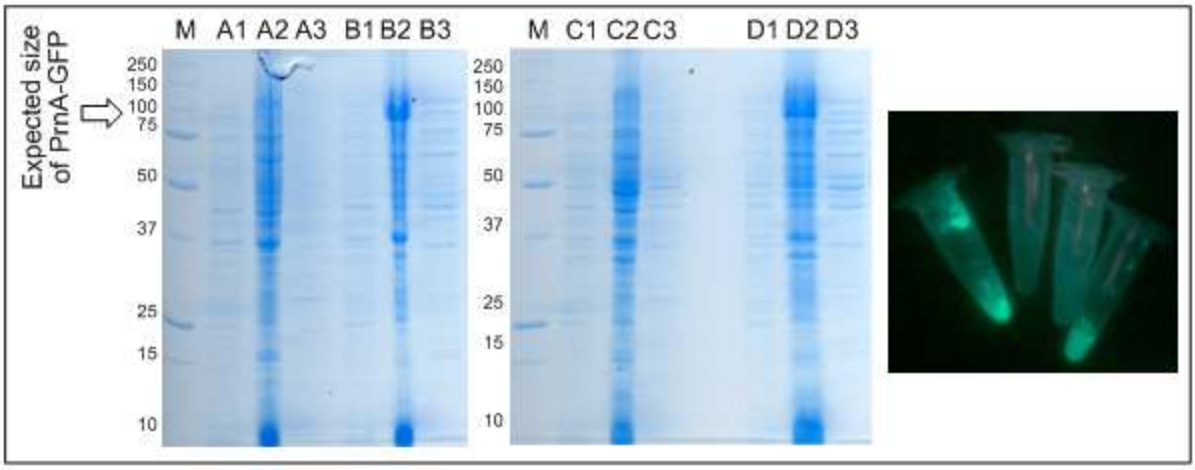


Figure 3.26: SDS-PAGE analysis of PrnA halogenase production under the propionate promoter with *E. coli* BL21 and PHL644 cells with or without the presence of 20 mM sodium propionate. (A) Production of PrnA in BL21 cells without propionate, **(B)** Production of PrnA in BL21 cells following induction with 20 mM sodium propionate, **(C)** Production of PrnA in PHL644 cells without propionate, **(D)** Production of PrnA in PHL644 following induction with 20 mM sodium propionate. **(1)** Pre-induction sample, **(2)** Cell pellet, **(3)** Cell free lysate. PrnA-GFP fusion expected size 88.35 kDa. Position of likely band highlighted with arrow. BL21 and PHL644 Cell pellets and lysates containing GFP can be seen fluorescing under uv light following induction with propionate.

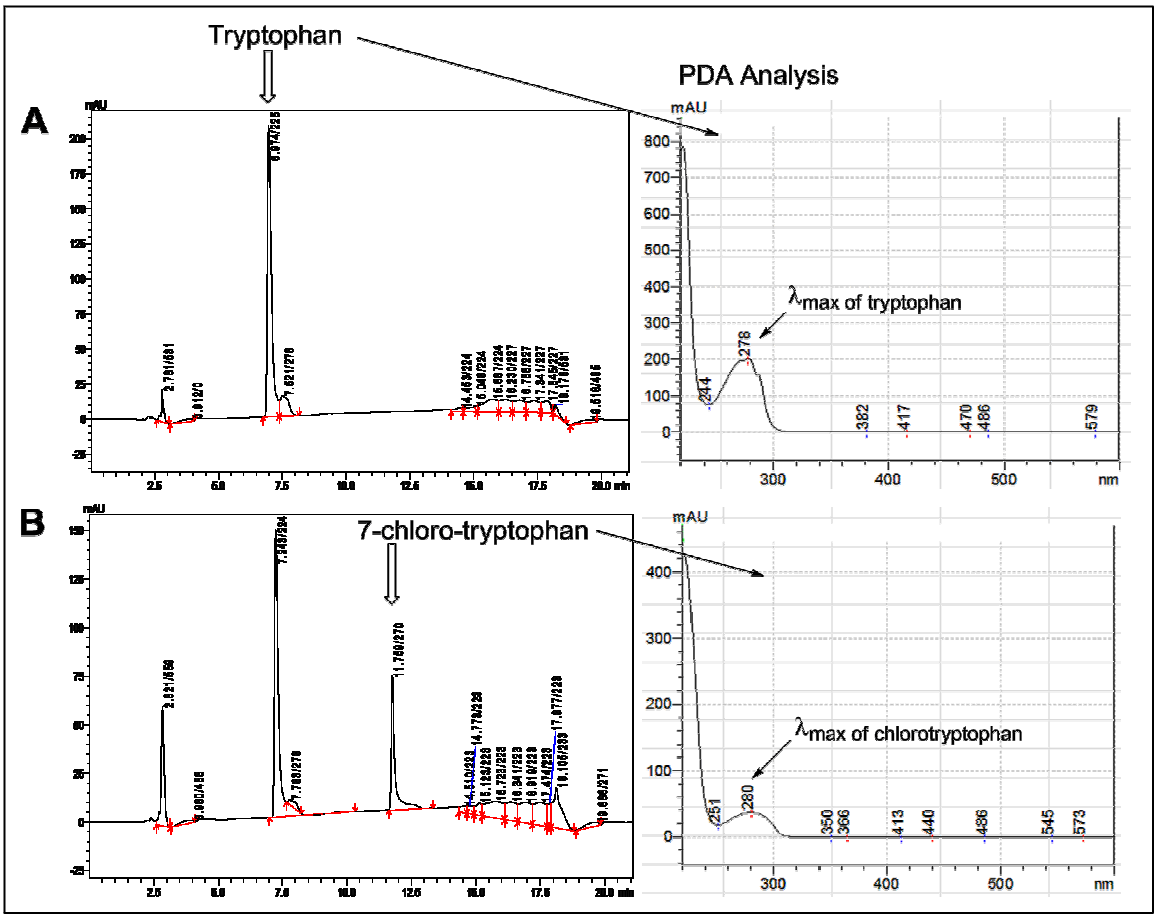


Figure 3.27 HPLC traces of PrnA halogenase activity of *E. coli* MW-040. (A) Biotransformation reaction at time 0 shows single peak of tryptophan at 6.9 mins. **(B)** The same biotransformation 6 hours later shows the emergence of a secondary peak at 11.75 mins that corresponds to the formation of 7-chlorotryptophan (verified by PDA analysis showing λ_{max} of 280 nm).

inductant (figure 3.26A,C). In the cultures that had been induced with 20 mM sodium propionate however a new band could be seen at the approximate size (88 kDa). The presence of induced PrnA-GFP was visualised by holding samples of the cell pellets and cell lysates of BL21 and PHL644 under ultra violet light. The resulting tubes glowed bright green (figure 3.26). This confirmed that PrnA could be produced inside the biofilm producer PHL644 using the propionate promoter. To confirm the activity of the halogenase within planktonic cells some simple biotransformation reactions were set up. The activity of the PHL644 compared to BL21 cells was also investigated. Even though the SDS-PAGE analysis showed no evidence of over production of T7 induced proteins inside PHL644 these strains were tested for activity anyway.

E. coli PHL644 strains MW-040 and MW-042 were grown in 500 ml of LB and a planktonic cell suspension formed in the same way as with the tryptophan synthase planktonic cell experiments. *E. coli* BL21 strains MW-039 and MW-041 planktonic cell suspensions were prepared in the same way.

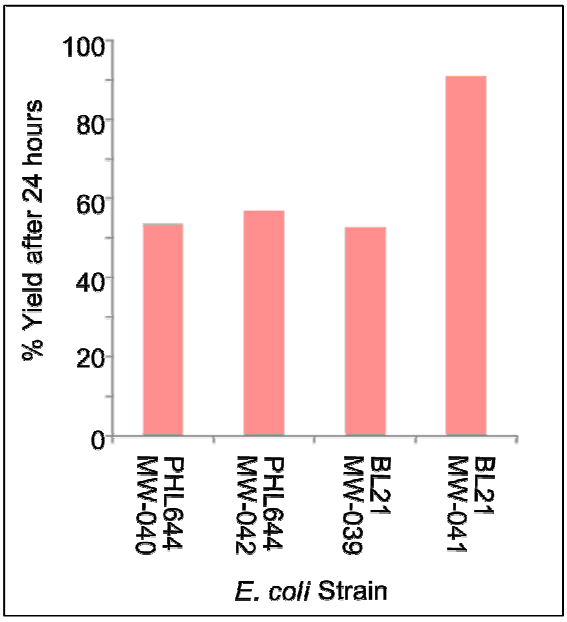


Figure 3.28 Average PrnA mediated conversion of tryptophan to 7-chlorotryptophan after 24 hours using planktonic cell suspensions of various *E. coli* strains.

For each of the strains, duplicate biotransformation reactions were set up containing 50 ml of potassium phosphate reaction buffer, 1 mM L-tryptophan and 25 mM NaCl. To each reaction 10 ml of the prepared cell suspension was added inside dialysis tubing. The reactions were incubated at 28°C for 24 hours. The reactions were then analysed with HPLC using the same method developed for assessing the yield of tryptophan synthase biotransformations. An example of HPLC traces from the reaction can be seen in figure 3.27.

Halogenase activity was present in all of the cell-suspension samples (see figure 3.28), even with the T7 induced pET21 plasmid inside PHL644. This suggests that active protein is still being produced despite no evidence of over production on the SDS-PAGE analysis. The T7 promoter is not negatively regulated so there is potential for unregulated protein production such as seems to be produced in this case.

Of the four planktonic cell suspensions tested three of them (MW-040, 042 and 039) converted tryptophan into the chlorinated product in the same yield (approximately 50%). Interestingly the best biocatalyst was the GFP tagged PrnA inside *E. coli* BL21 (MW-041) which demonstrated a final yield of just over 90%.

These studies prove that different enzymes can be inserted into the biofilm producer *E. coli* PHL644 and that they show biocatalytic activity. Different promoters can be used within the biofilm

producer. Despite the fact that PHL644 does not have the required genes to over express proteins regulated under the T7 promoter there is enough residual activity to show enzyme turnover.

The next step would be to create a biofilm using these strains and see if the resulting immobilised biocatalyst is capable of showing enhanced catalytic longevity compared to the planktonic cell solution shown here. This would help to demonstrate the versatility of the SCB and therefore the wide scope of this biofilm biocatalyst technology to the world of industrial biocatalysis.

Chapter 4:

Background: Natural Products and Halogenation

4.1 Halogenated natural products

The plethora of different secondary metabolites known to be produced by living organisms number in the hundreds of thousands. A diverse sub-set of these metabolites contain carbon-halogen bonds and over 4000 of these naturally produced organohalogens have been identified to date including several well known antibiotics such as vancomycin and chlorotetracycline.[148, 149]

Chlorinated natural products make up the majority of this group while bromine containing compounds are the second most abundant. Compounds containing iodine and fluorine are significantly rarer.[150] These natural products full into a broad range of structural classes from simple phenolic and aliphatic compounds to more structurally complex polyketides and oligopeptides (see figure 4.1 for some examples).[149] Chlorinated natural products are typically produced by terrestrial organisms while compounds containing bromine are more likely to be produced by marine microorganisms.[151] This is most likely due to the relative abundance of chloride and bromide ions in the environment of each organism.

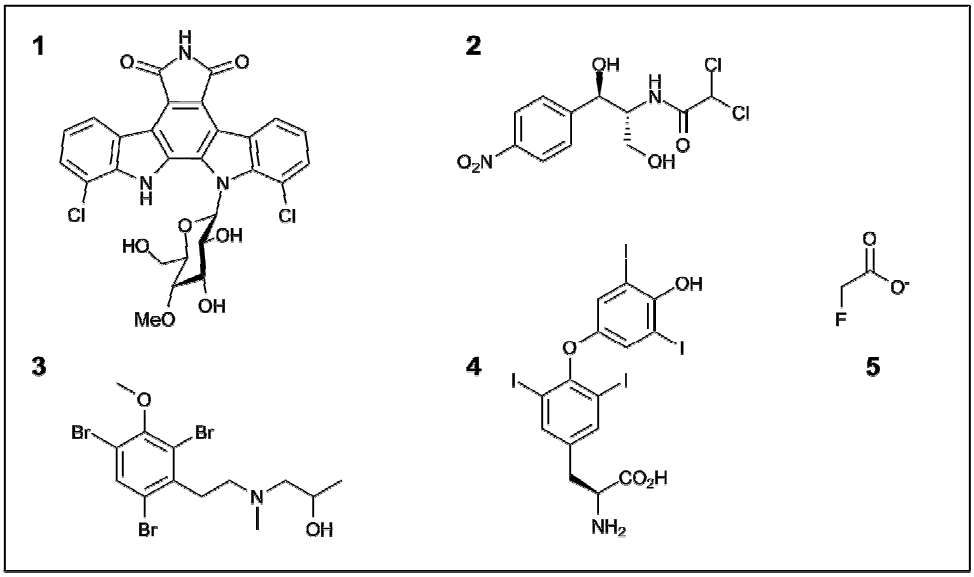


Figure 4.1: Examples of naturally occurring organohalogens. Chlorinated natural products include rebeccamycin (**1**) and chloramphenicol (**2**). Convolutamine A, (**3**), and tetraiodothyronine, (**4**), are examples of brominated and iodinated natural compounds. Though rarer, fluorinated natural products such as fluoroacetate (**5**) are also found in nature.

The biological activities of these halogenated compounds also demonstrate great diversity. The chlorinated chloramphenicol for example **(2)** (figure 4.1) shows antibiotic activity while rebeccamycin **(1)** (figure 4.1) exhibits antitumour effects.[152] The brominated convolutamine family of alkaloids **(3)** (figure 4.1) show cytotoxic properties.[153] Iodinated compounds are much scarcer in nature but examples include calicheamicin (an antibiotic produced by a soil actinomycete)[154] and tetraiodothyronine **(4)** (figure 4.1) which is produced in humans as one of the major forms of the mammalian thyroid hormone.[155] Fluorinated compounds are rarer still but examples include the toxin fluoroacetate **(5)** (figure 4.1) which can be found in the leaves of certain plant species.[156]

Within these classes of halogenated natural product, the incorporation of the halogen is usually essential for activity. This can be illustrated by clorobiocin **(6)** (figure 4.2) which is a potent inhibitor of bacterial DNA gyrase and a member of the aminocoumarin family of antibiotics. Eustáquio and co-workers introduced mutations into the biosynthetic cluster encoding for this antibiotic in order to knock out the *clo-hal* gene responsible for installing the chlorine atom, thus generating an analogue that contained hydrogen at this position instead. In a parallel experiment, insertion of a methyltransferase gene into the mutant enabled the generation of a methylated variant of clorobiocin with the methyl group occupying the same position as the chlorine.[157] The chlorinated antibiotic was 8 times more active compared to the hydrogen variant and 2 times more than the methylated variant.

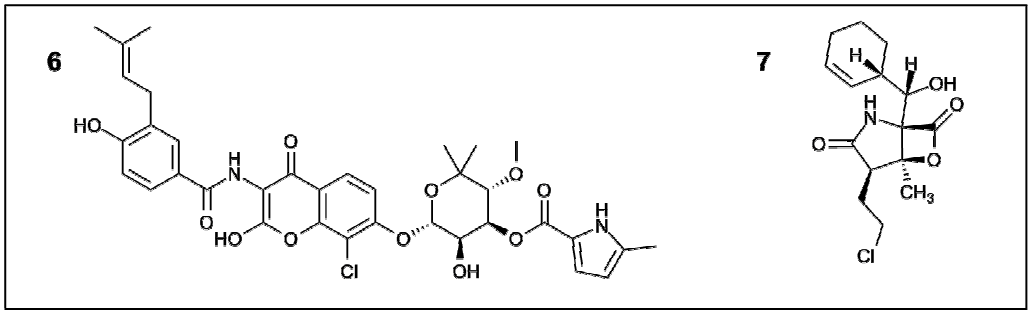


Figure 4.2: Structures of two clinically relevant chlorinated compounds. Clorobiocin (**6**) is an antibiotic of the aminocoumarin family. Salinosporamide A (**7**) is a potential anti-cancer drug.

In another example the activity of the 20S proteosome inhibitor salinosporamide A **(7)** (figure 4.2) is strongly influenced by the presence of the chlorine substituent. The lactone of the salinosporamide is ring opened by an active site threonine hydroxyl present in the proteosome. The resulting free nucleophilic hydroxyl of the lactone then intramolecularly displaces the chlorine resulting in an irreversible covalent bond, preventing the reformation of the lactone which would release the inhibitor. This activity is eight times greater than the deschloro analogue salinosporamide B,

produced by the same marine organism, *Salinispora tropica*, which incorporates hydrogen at the same position as the chlorine (0.7 nM compared to 5.2 nM).[158] The removal of this chlorine leaving group renders a less stable compound that is vulnerable to hydrolysis and therefore represents a less potent inhibitor. Incorporation of fluorine at this position, using mutasynthesis techniques to replace the *salL* gene (responsible for chlorination of salinosporamide) with the fluorinase gene *flA*, results in a compound with slightly reduced proteosome inhibition than the chlorinated salinosporamide A (1.5 nM). This is to be expected as the high energy required to break the C-F bond makes displacement by the active site hydroxyl unlikely, making the binding of fluorosalinosporamide reversible. The fluorinated analogue does however show three fold better activity than the non-halogenated salinosporamide B. The ability of the fluorine to form hydrogen bonds within the active site has been implicated as the reason for this increase.[159]

Rebeccamycin **(1)** is an anti-tumour drug with activity against topoisomerase I. Removal of the chlorine actually increases the *in vitro* anti-tumour activity of the compound which suggests that the halogen is not essential.[160] However cell antiproliferative assays show the opposite result with the chlorinated variant possessing higher activity. These results suggest that the inclusion of the chlorine aids in the transport of the compound across the cell membrane, Rodrigues and co-workers suggested that subsequent cleavage of the chlorine bond may occur inside the cell to improve activity of the final compound.[160]

4.2 Biological Halogenation

We have seen that nature produces a range of halogenated natural products and there are a number of classes of enzyme that are responsible for the introduction of these groups. Selective halogenation of natural products to enhance their activity is attractive for medicinal chemistry; understanding the mechanisms involved in biological halogenation may give rise to new and alternative synthetic catalysts for halogenation.

Many different organisms produce halogenases but they can generally all be grouped into two main categories.[149] Members of the first category require dioxygen for activity and tend to be very substrate specific; these enzymes require either flavin or α-ketoglutarate as a co-substrate. Members of this class are referred to as O_2-dependent halogenases. The second class are less rigid when it comes to choice of substrate and utilise hydrogen peroxide and therefore are known as haloperoxidase enzymes (HPO).[151] Both classes function by first generating a reactive hypohalite species (X^+) by a two-electron oxidation of the halide.[150] The electron deficient nature of this species (in most cases a hypohalous acid) allows the enzyme to halogenate nucleophilic carbon centres.

Enzymes of this type have been found that incorporate chloride, bromide and iodide but neither dioxygen or hydrogen peroxide have the power to oxidise fluorine, therefore no fluorinated compounds are produced using these enzymes. The mechanisms that produce the limited number of fluorinated natural products will be briefly discussed later.

4.2.1 Haloperoxidases (HPOs)

There are two types of haloperoxidase which either contain heme or vanadate as a cofactor. In heme-dependent HPOs the formation of the hypohalous acid is a redox process while the vanadate-dependent halogenases do not change their oxidation state but function as lewis acids.[149] Further classification of the HPO class of enzymes is based on which halides they accept as substrates, their name derives from the most electronegative halide that can be used. Chloroperoxidase enzymes can catalyse the oxidation of chloride, bromide and iodide, bromoperoxidases oxidise bromide and iodide, while iodoperoxidases can only accept iodide.[149]

The proposed mechanism of action for the heme-dependent HPOs involves the binding of peroxide axial to the enzyme bound Fe^{III}-porphyrin complex (figure 4.3). Loss of water results in a Fe^{IV}-oxo species which is intercepted by the halide to from a Fe^{III}-hypohalite species. This enzyme bound species can potentially be used directly for halogenation or free hypohalous acid can be released which can perform halogenation on substrates not adjacent to the active site. The lack of substrate specificity of these enzymes and the fact that these enzymes work best at acidic pH has resulted in the wide belief that the free acid is the primary agent responsible for halogenation.[150] It is an enzyme of this class that produces the iodinated mammalian thyroid horme **(4)** (figure 4.1) in an iterative fashion with several passes of halogenation occurring to yield the final tetra-iodinated product.

The vanadium-dependent enzymes are found widely in marine organisms that produce brominated natural products, although some examples of these enzymes in terrestrial organisms have been found.[161] In a similar fashion to the heme-dependent HPOs, the vanadium centre binds hydrogen peroxide and activates it ready for halide attack (figure 4.3). The difference is that the metal ion does not act as a redox centre but more like a lewis acid. Binding of the peroxide to the vanadium forms a reactive peroxo intermediate which reacts with a halide to form the vanadium-bound hypohalite. As with the heme-dependent enzymes it is unclear whether this enzyme bound intermediate is the active halogenation species or whether hypohalous acid is released,[150] the general lack of substrate specificity with these enzymes would suggest the latter. The formation of free acid is also supported by studies showing that the formation of brominated organic substrates does not occur

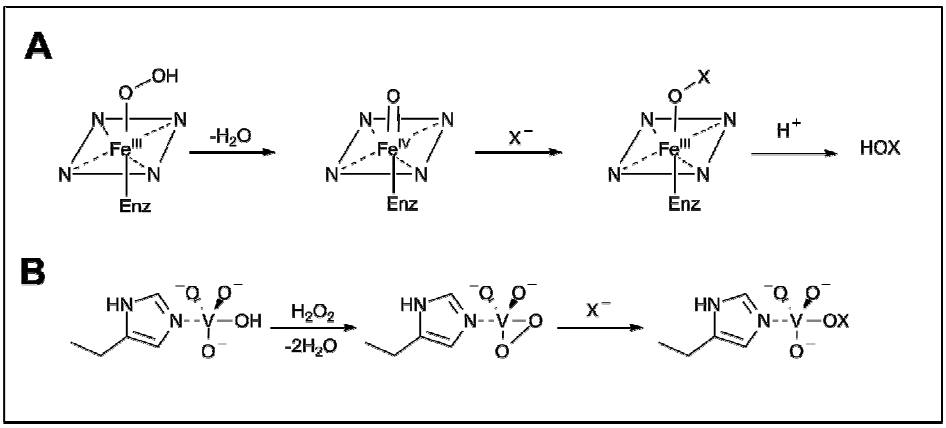


Figure 4.3: Formation of hypohalous acid (HOX) by Haloperoxidase enzymes. Both heme- and vanadium-dependent HPOs generate the X^+ equivalent ion needed to halogenate electron rich substrates. The formed hypohalous acid can either function as an enzyme-bound intermediate or dissociate to act free in solution.[150]

stoichiometrically, with H_2O_2 being consumed faster than brominated product is formed. This strongly indicates that free HOBr is being produced and is then lost from the active site in the presence of a poor substrate.[162]

4.2.2 O_2 dependent halogenases

Although the haloperoxidases were the first class of halogenase enzyme to be discovered two new classes have recently been revealed. The first class require reduced flavin as a cofactor and are therefore classified as $FADH_2$-dependent halogenases. The first of these to be discovered was found to be responsible for installing the chlorine during chlortetracycline **(15)** biosynthesis.[163] Homologues of this enzyme have subsequently been found in dozens of other secondary metabolite clusters, it has been suggested that this prevalence could be used to detect new clusters for halogenated natural products.[151]

The second class of dioxygen dependent halogenase are non-heme Fe^{2+} enzymes homologous to the α-ketoglutarate containing two-histidine, one carboxylate family of non-heme Fe^{2+} oxygenases.[151] These enzymes are capable of chlorinating inactivated carbon centres and like the similar oxygenase enzymes use α-ketoglutarate (α-KG) as a co-substrate and are generally involved in the chlorination of terminal methyl groups of amino acids linked to peptidyl carrier proteins.[149] Such inactivated carbons are not ameanable to attack by electrophiles and so a radical mechanism of chlorination has been proposed for these types of enzyme.[164]

4.2.2.1 Mechanism of non-heme Fe^{2+}/ α-ketoglutarate/O_2-dependent halogenases

The information gained on this class of halogenase enzymes was obtained by studying the enzyme SyrB2. This enzyme is responsible for chlorinating the γ-methyl group of L-threonine during the biosynthesis of syringomycin E in *Pseudomonas syringae* **(8)** (figure 4.4).[165] Just as we have seen with other halogenated secondary metabolites, the phytotoxicity of this compound is strongly dependent on the presence of the halogen and shows three times less activity in the absence of chlorine.[165] It has also been shown that this enzyme will only halogenate peptidyl carrier protein bound L-threonine and can also incorporate bromide in additional to the chloride.[166]

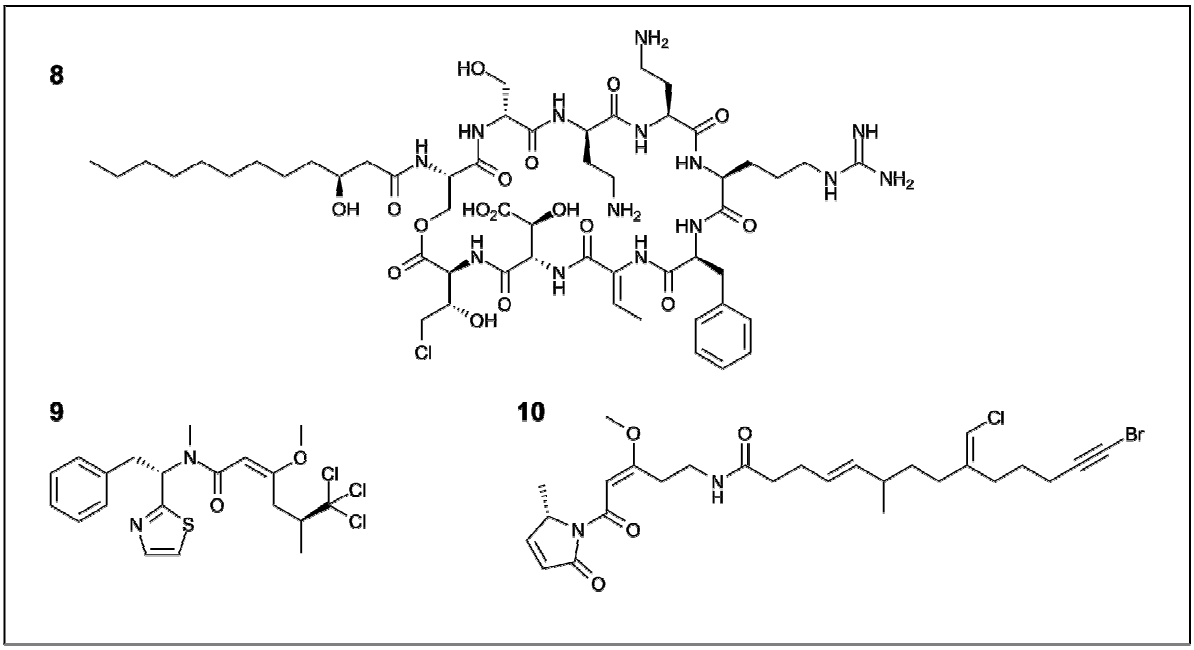


Figure 4.4: Examples of natural products halogenated by non-heme Fe^{2+}/α-ketoglutarate/O_2-dependent halogenases. (8) syringomycin E, **(9)** barbamide and **(10)** jamaicamide.

The crystal structure of SyrB2 shows that as with the α-KG-dependent oxygenase enzymes; the non-heme iron centre is bound by two histidine residues, α-ketoglutarate and water.[167] The feature that sets the halogenase apart, and presumably imparts its specific activity as a halogenase, is that the site that would normally be occupied by a carboxylate with the oxygenases (aspartate or glutamate) is occupied by chlorine which, together with the other ligands, gives the Fe centre an oxidation state of 2^+.

The proposed catalytic cycle for non-heme Fe^{2+} halogenases can be seen in figure 4.5. Attack by dioxygen results in the decarboxylation of the α-KG and forms a Fe^{4+}-oxo-species. This highly reactive

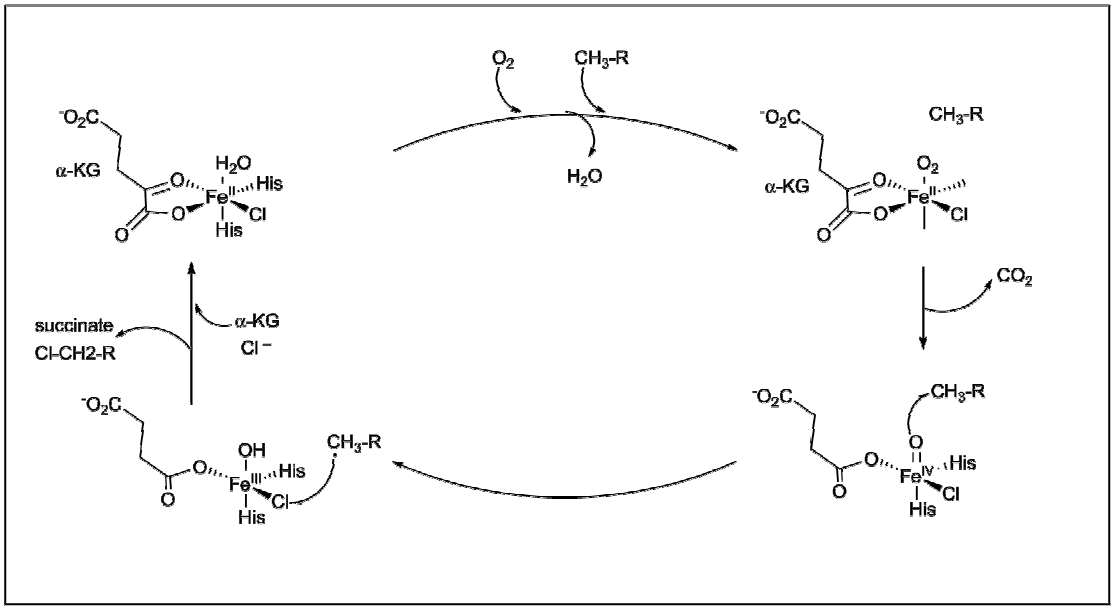


Figure 4.5: Catalytic cycle of the non-heme Fe^{2+}/α-ketoglutarate/O_2-dependent halogenases. R-CH_3 represents the methyl group of the carrier protein-bound substrate.[149] Decarboxylation of the α-KG is catalysed by molecular oxygen. A highly reactive Fe^{4+}-oxo-species forms which abstracts a hydrogen radical from the aliphatic carbon centre of the substrate, which subsequently abstracts the halide to yield the halogenated product.

species abstracts a hydrogen radical from the aliphatic carbon centre of the substrate, which subsequently abstracts the halide to yield the halogenated product.[149] Similar enzymes with identical mechanisms have been proposed for the biosynthesis of several other halogenated natural products including barbamide **(9)** and jamaicamide **(10)**.[168, 169]

4.2.2.2 Mechanism of $FADH_2$-dependent halogenases

The second class of dioxygen dependent halogenation enzymes are the $FADH_2$-dependent halogenases. The first examples of this type were discovered to be involved in pyrrolnitrin **(11)** and chlortetracycline **(15)** biosynthesis.[163, 170]

When Hammer and co-workers identified the pyrrolnitrin gene cluster, four enzymes (PrnA, B, C and D) were shown to be the sole enzymes required for biosynthesis (insertion of these four genes into *E. coli* resulted in the production of pyrrolnitrin).[170] Pyrrolnitrin is a dichlorinated molecule and gene knockouts were used to prove that within the cluster PrnA and PrnC both act as halogenase enzymes to regioselectively install one of the chlorine atoms each.[171] The significant and complimentary work carried out by the groups of van Pée and Naismith on PrnA has resulted in this enzyme becoming the model for how all enzymes in this family operate.

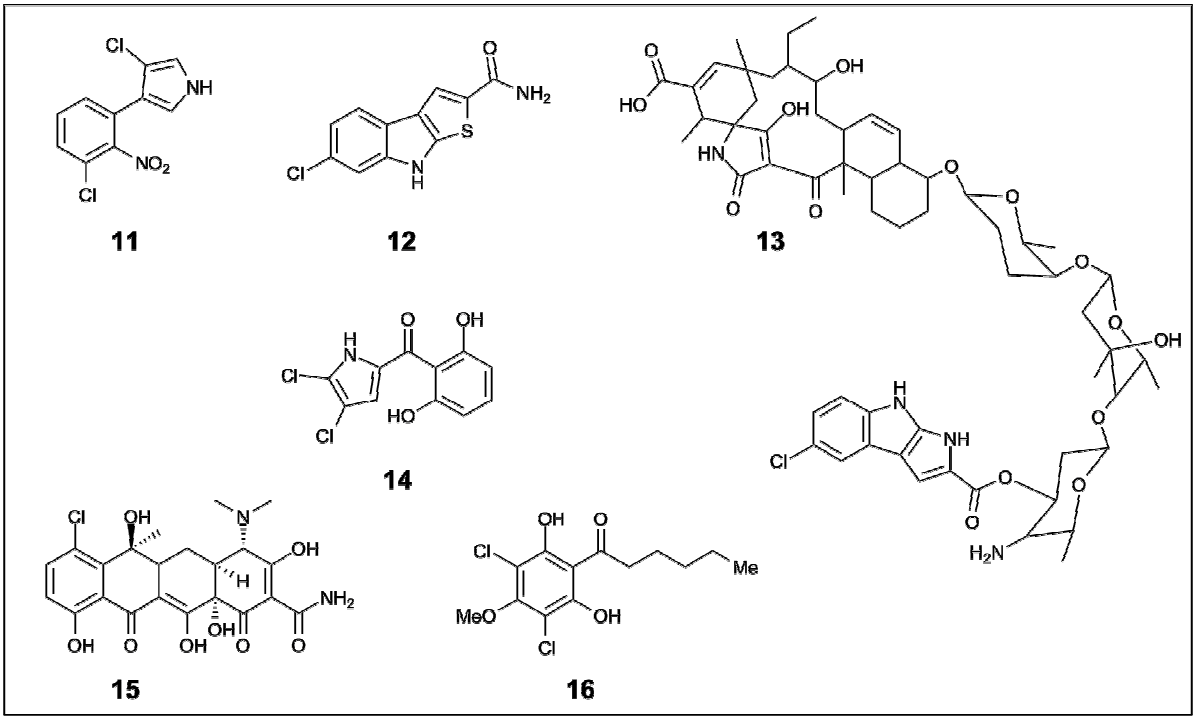


Figure 4.6: Examples of chlorinated natural products produced by $FADH_2$-dependent halogenase enzymes. Pyrrolnitrin **(11)**, thienodolin **(12)**, pyrroindomycin B **(13)**, pyoluteorin **(14)**, chlorotetracycline **(15)** and DIF-1 **(16)**.

As the name of the family suggests, essential to the mechanism of flavin-dependant halogenase enzymes is the presence of reduced flavin. $FADH_2$ is not very stable and tends to be quickly oxidised by O_2 back into FAD, therefore the FAD must be reduced *in situ* by or for the halogenase when it is ready to be used. The purification and characterisation of PrnA was reported in 2000 by van Pée where it was determined that PrnA was able to regio-selectively halogenate L-tryptophan to 7-chloro-L-tryptophan.[172] This protein is one of a group of $FADH_2$-dependent halogenases from different biosynthetic clusters that utilise tryptophan as their substrate. Figure 4.7 shows a comparison of the different products of these enzymes and provides a definition for the numbering system convention around the indole ring. RebH catalyses an identical reaction to PrnA to produce 7-chloro-tryptophan during the biosynthesis of rebeccamycin **(1)** (figure 4.1). Despite performing exactly the same function the two enzymes only share a similarity of 54% which highlights the remarkable variation present between members of this family of enzymes. Two other similar tryptophan halogenases have been found that are capable of chlorinating tryptophan at different positions. PyrH from pyrroindomycin B **(13)** biosynthesis in *Streptomyces rugosporus* chlorinates tryptophan at the 5 position.[173] The enzyme Thal from *Streptomyces albogriseolus* is characterised as a tryptophan-6-halogenase involved in the biosynthesis of the alkaloid natural product thienodolin **(12)**.[174] More recently another tryptophan-6-halogenase, stth, has been isolated from *Streptomyces*

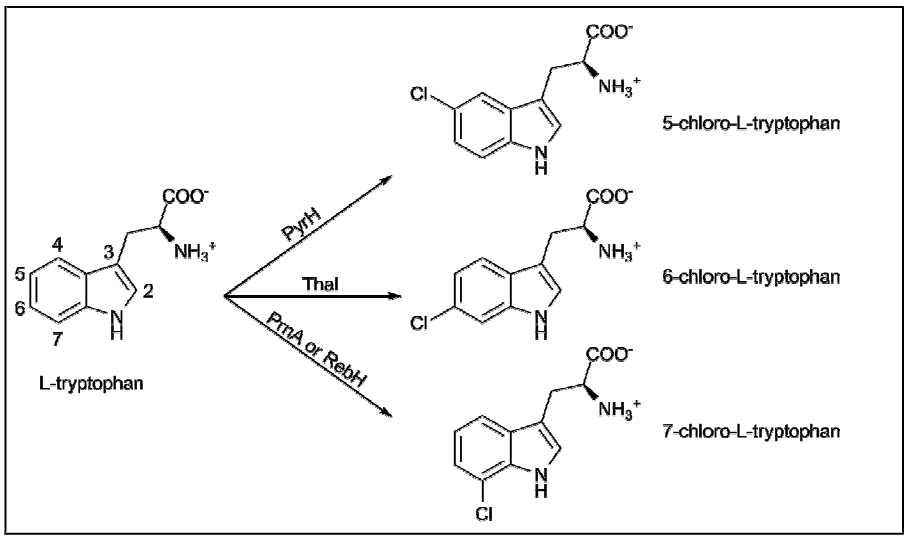


Figure 4.7: Regioselective chlorination of L-tryptophan by different $FADH_2$-dependent halogenases. Different members of the family regio-selectively install a chlorine atom at a different position around the indole ring. The numbering convention for the indole ring is shown.

toxytricini. This enzyme was located by genome analysis inside an NRPS cluster and the function identified by purification and characterisation but the product of the cluster remains unknown.[175] All the tryptophan halogenase enzymes share approximately 40-50% sequence identity, with Thal and RebH showing the most similarity at 63%. The fact that these similar enzymes produce only a single chloro-tryptophan isomer demonstrates that, unlike the HPOs, $FADH_2$-dependent enzymes operate with extremely high selectivity.

During the purification of PrnA, a second protein component was initially co-purified that was found to be essential for halogenation activity *in vitro*.[172] This component was identified as a flavin reductase, a class of redox enzyme that reduces FAD into $FADH_2$ by consuming NAD(P)H. The earlier studies that had identified the essential genes for pyrrolnitrin did not experience problems with production in the recombinant *E. coli* host despite the fact that no flavin reductase gene had been found within the cluster.[170] During the initial activity assays with PrnA it was determined that the flavin reductase was an unspecific enzyme that could be substituted by other similar enzymes, for instance the flavin reductase SsuE found in *E. coli* strains (this enzyme prefers NADPH over NADH as the redox cofactor).[172] The halogenase enzyme is able to accept any form of reduced FAD regardless of its origin, even chemically reduced flavin (regenerated using organometallic complexes) has been successfully used.[176]

In 2007 another enzyme from *Pseudomonas* was discovered, PrnF, which was demonstrated to be a flavin reductase which prefers NADH as a cofactor to reduce FAD.[177] This ability to mix and match

halogenases with different methods of reducing flavin demonstrates a potential flexibility in applications.

The general catalytic cycle of the flavin-reductase/halogenase coupled reaction is shown in figure 4.8. Reduced flavin is generated by the flavin reductase mediated oxidation of NAD(P)H and is released into solution. This is then taken up by the halogenase enzyme and reacts with molecular oxygen to form a peroxide-linked isoalloxazine ring. Information from the crystal structure of PrnA has shown that the $FADH_2$ and chloride ion are bound at the same site of the enzyme.[178] The chloride attacks the peroxide species to generate hypochlorous acid (HOCl); this is the same proximal chlorinating agent produced in the haloperoxidase class of halogenase enzymes and generates a Cl^+ equivalent that enables the halogenation of electron rich carbon centres.

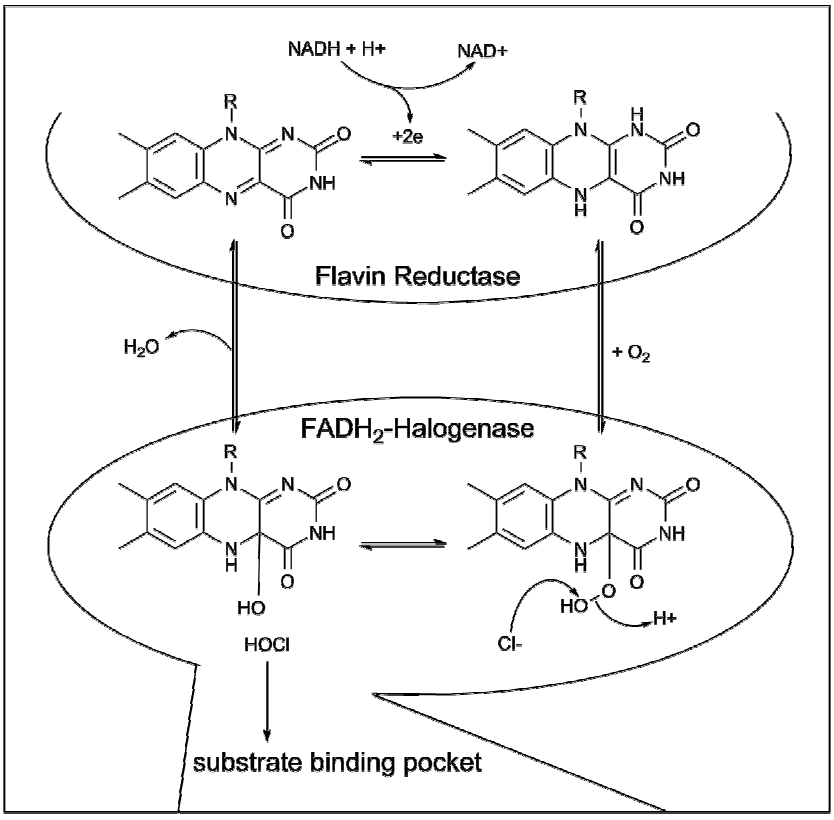


Figure 4.8: General catalytic mechanism for the generation of HOCl chlorinating agent in $FADH_2$-dependent halogenase enzymes. FAD is first reduced to $FADH_2$ by the oxidation of NAD(P)H within the active site of a flavin reductase enzyme. Inside the active site of the halogenase, $FADH_2$ reacts with molecular oxygen to form a peroxide species. Hypochlorous acid is generated following attack of this species by chloride. HOCl is then channelled to the halogenase substrate binding site. Loss of water regenerates the FAD which can again be recycled by the flavin reductase back to the active reduced form.

A very similar process to generate the flavin hydroperoxide (FAD-OOH) occurs in flavin-dependent monooxygenases and there is significant sequence homology between the two sets of enzymes. Both the monooxygenases and the halogenases share a conserved flavin binding module near the N-terminus with the general sequence GxGxxG. Another flavin-dependent halogenase conserved sequence however is not found in the monooxygenases and has the sequence WxWxIP. The two tryptophan residues in this motif are located close to the flavin binding site and have been proposed to block the binding of substrate close to the flavin, therefore preventing the enzyme from acting as a monooxygenase.[178] However exchange of these conserved tryptophan residues, individually or together, for phenylalanine results in no change in halogenation activity. No change in activity is seen even if the first of the tryptophan residues is exchanged for alanine. Only when the second conserved tryptophan is exchanged for alanine is any change observed but all enzyme activity is lost in this case and no monooxygenase activity is seen.[179] The lack of apparent function for these conserved residues is strange considering that they are absolutely conserved in all $FADH_2$-dependent halogenases but the evidence does not seem to support the theory that they function to block monooxygenase activity.

The crystal structure of PrnA shows that the tryptophan binding site is not located adjacent to the flavin binding site but down a 10 Å long tunnel. The hypochlorous acid is channelled down this tunnel to reach the substrate binding pocket. Work performed on the $FADH_2$-dependent halogenase, RebH, supports this view. Yeh and co-workers proved that the flavin redox chemistry is completed before the tryptophan substrate is halogenated and that the formation of the reactive HOCl species can occur in the absence of substrate, therefore showing that the generation of the hypochlorous acid and the chlorination of product are not coupled processes.[180]

The very first proposed mechanism for the FAD-dependent halogenases was suggested by Yeh and co-workers in 2005, prior to the crystal structure of PrnA being obtained, based on the chemistry performed by RebH.[181] Following the formation of the flavin-hydroperoxide (FAD-OOH) the chlorine was suggested to attack this species to generate an FAD-O-Cl intermediate. Halogenation was then achieved by attack on this intermediate by the aromatic π electrons of the tryptophan substrate. The findings from the crystal data and the identification of the formation of HOCl disproved this theory.

Information from the crystal structure showed that a lysine residue (K79), which is conserved in all flavin-dependent halogenases, was located at the end of the hypochlorous acid tunnel. The importance of this lysine was determined by exchanging it for alanine which led to a total loss of activity. Dong *et al.* suggested that this lysine residue formed a hydrogen bond to the HOCl which positioned it in order to react with tryptophan in a regio-selective manner (figure 4.9).[178]

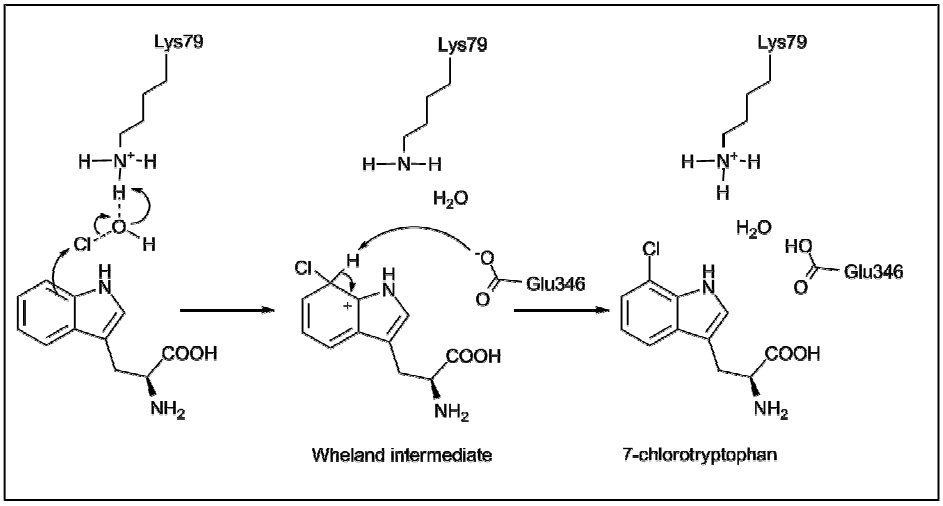


Figure 4.9: Early proposed $FADH_2$-dependent halogenase mechanism. Mechanism based on initial crystal data of PrnA by Dong *et al.* 2005.[178] Conserved lysine-79 forms a hydrogen bond with the HOCl, positioning it to react with tryptophan. Glutamate-346 abstracts the proton from the resulting Wheland intermediate to generate 7-chloro-tryptophan.

This is followed by electrophillic addition of the chlorine to tryptophan, generating a Wheland intermediate which is deprotonated by a nearby glutamate residue (E346), leading to product.

In 2007 Yeh and co-workers characterised and crystallised the similar tryptophan-7-halogenase RebH and presented evidence that questioned the PrnA mechanism put forward by Dong *et al*.[182] HOCl is a potent oxidant and is capable of reacting indiscriminately with many biological molecules including protein side chains; however Yeh and co-workers described that despite purifying out RebH following incubation with $FADH_2$ and NaCl (preventing formation of fresh chlorinating species) the enzyme retained chlorinating activity even 48 hours later when introduced to the L-tryptophan substrate in the absence of the flavin cofactor, suggesting the presence of a long lived chlorinating species. The crystal structure of RebH revealed that the apoenzyme contained a solvent exposed active site and it was viewed doubtful that the HOCl could be maintained for so long under such conditions and would presumably have to be generated and consumed by the enzyme as quickly as possible to minimise side reactions, however as mentioned above, the formation of HOCl was shown to occur even in the absence of bound tryptophan. The crystal structure of RebH revealed that the conserved lysine residue found in all enzymes in the $FADH_2$-dependent halogenase family was located between the flavin and tryptophan binding sites. It was proposed that the hypochlorous acid formed in the flavin binding site reacts with the essential and conserved active site lys79 residue to form a long lived chloramine species, the half life of which (>25 hours at 37°C) matched the half life

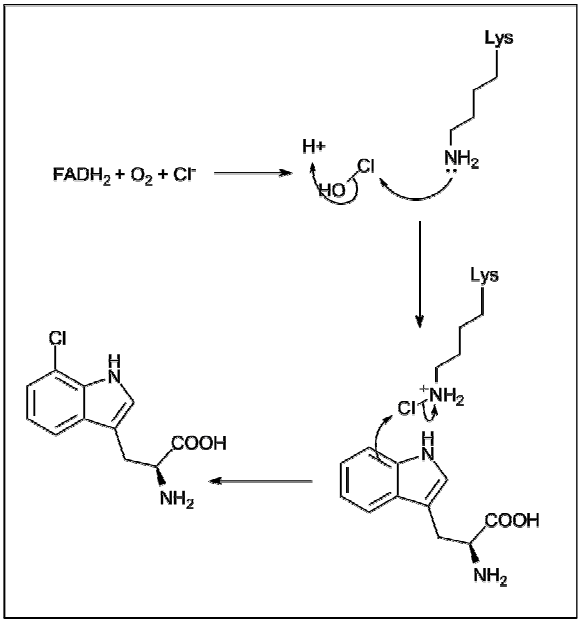


Figure 4.10: The widely adopted proposed $FADH_2$-dependent halogenase mechanism. Reaction of hypochlorous acid with the conserved lysine results in a long lived chloramine species that can act as a chlorinating agent. Evidence for this mechanism was provided by Yeh *et al.*[182] The chlorinating species was formed by reacting $FADH_2$, O_2 and NaCl with RebH. The enzyme and small molecule components were then removed by size exclusion chromatography to prevent the formation of fresh chlorinating agent before RebH was introduced to the L-tryptophan substrate. Chlorination was still observed even after introducing the substrate 48 hours after the removal of the $FADH_2$; indicating the presence of a stable and long lived chlorinating intermediate.

of the chlorinating species detected with RebH (28 hours at 25°C). Therefore a new mechanism was proposed (figure 4.10) that involves the initial formation of a lys-εNH-Cl chloramine species that acts as the proximal chlorinating agent in the same way as the stabilised hypochlorous acid in the previous mechanism. This new mechanism has been widely accepted and is used as the general mechanistic model of the function of all $FADH_2$-dependent halogenases.

Doubts have however been raised in regards to the accuracy of this proposed mechanism. Most of those doubts revolve around the proposed formation of the chloramine. Formation of this species was only identified in the absence of tryptophan and a crystal structure of this important intermediate could not be obtained to prove the presence of this species.[182] Chloramine is a weaker halogenating agent than HOCl and although phenolic or pyrrole rings are susceptible to chlorination by chloramine, it has been shown that tryptophan cannot be halogenated by chloramine in solution.[179] This is a major argument against this proposed mechanism of action.

The flavin-dependent tryptophan halogenase enzymes also contain an additional conserved glutamate residue (E346) that is positioned just across the flavin-access tunnel from the essential K79. E346 is not found in halogenases that utilise different substrates but replacement of this residue in PrnA with glutamine results in a decrease of activity by two orders of magnitude.[178] In their mechanism, Dong *et al.* had assigned this glutamate a stabilisation role, abstracting the proton from the Wheland intermediate (figure 4.9). The absence of this conserved residue in the non-tryptophan flavin halogenases however casts doubt on the validness of this claim. Flecks *et al.* obtained the crystal structure of an E346D mutant protein which showed the aspartate residue occupying a different orientation to the glutamate without altering the position of any other component of the enzyme.[179] This mutant also proved to be inactive and led to suggestions that the formation of a chloramine was not sufficient on its own to catalyse the halogenation of tryptophan. Instead it was suggested that the localisation of the negative change was crucial. Another new

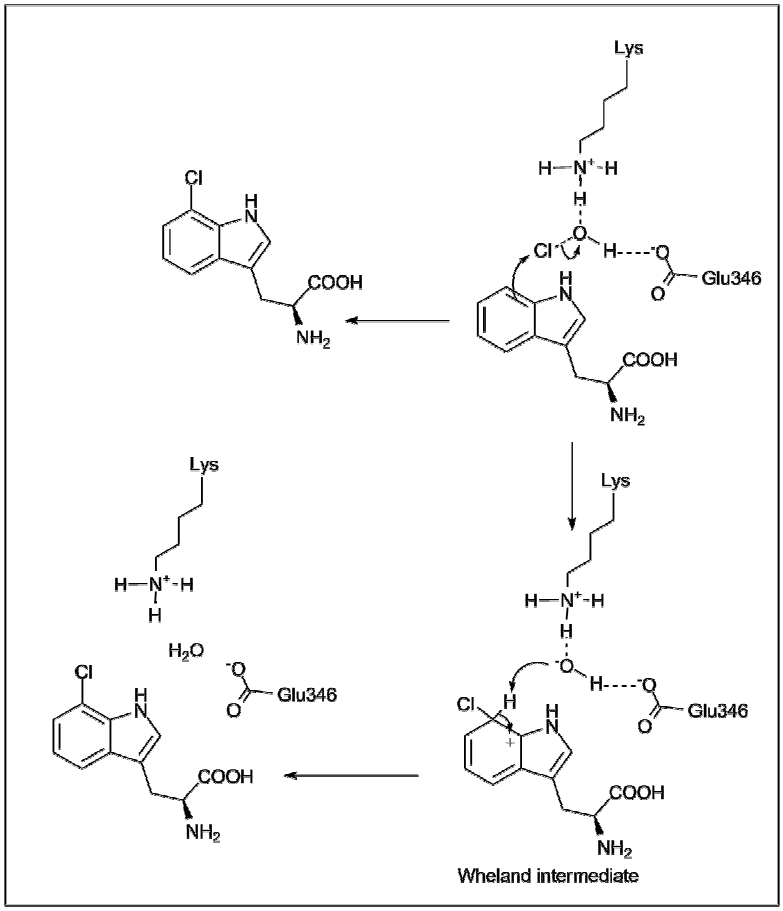


Figure 4.11: Flecks *et al.* proposed mechanism of $FADH_2$-dependent halogenase. The negative charge on the conserved E346 interacts with the hydrogen of the HOCl resulting in an enhancement to its electrophilicity and subsequent electrophilic aromatic substitution reaction with tryptophan.[179]

mechanism was therefore proposed (figure 4.11) that incorporated these findings. In this mechanism the negative charge on the glutamate interacts with the hydrogen of the HOCl which results in enhanced Cl^+ electrophilicity and also helps to localise the chlorine species at the right position for regio-selective incorporation. In the absence of substrate the hypochlorous acid is in equilibrium with the chloramine detected by Yeh *et al.*[182]

The lack of conserved E346 in other flavin-dependent halogenases however suggests that this cannot be a universal mechanism for all the enzymes of this family which is made up of a large number of enzymes accepting different substrates. The tryptophan halogenases make up a small number of the $FADH_2$-dependent family. There are examples of flavin-dependent halogenases that chlorinate phenolic or pyrrole moieties during multienzymatic NRPS assembly lines, for example in the biosynthesis of chlorotetracycline **(15)**[149] or during pyoluteorin **(14)** biosynthesis.[183] Other examples are post-production halogenases such as the enzyme ChlA which dichlorinates the polyketide derived morphogen DIF-1 **(16)** (differentiation-inducing factor 1) following release from the PKS machinery.[184]

Despite the range of different enzymes, the tryptophan halogenases remain the best studied of the family. Investigations on a flavin-dependent halogenase with different substrate specificity may help to answer some of the remaining questions in order to better understand and characterise this broad family.

4.2.3 Enzymatic fluorination

Fluorine is more electronegative than chlorine, bromine or iodine, therefore the formation of an F^+ equivalent is very difficult. Fluorinase enzymes must therefore utilise a different mechanism to the other halogenases.[149] A nucleophilic mechanism has to be involved but F^- has a very high enthalpy of hydration and only under rigorously stringent conditions can the sheath of water be removed and F^- act as a nucleophile. For this reason biological reactions with fluoride are very rare but there are a number of natural products that do incorporate fluorine (figure 4.12),[156] most of which are toxins. For these reasons, fluorinase enzymes show significant differences from the other halogenase enzymes mentioned above.

The first known fluorinase enzyme was identified as a 5'-fluoro-5'-deoxyadenosine synthetase, FlA, and was isolated from a culture of *Streptomyces cattleya*. This enzyme uses *S*-adenosyl-L-methionine (SAM) as a co-substrate in the reversible conversion of SAM to 5'-fluoro-5'deoxyadenosine which is subsequently converted by other enzymes to fluoroacetate **(5)** (figure 4.1) and 4-fluorothreonine **(19)** (figure 4.14) to yield the final fluorinated natural products.[185] Within FlA the fluorine is thought

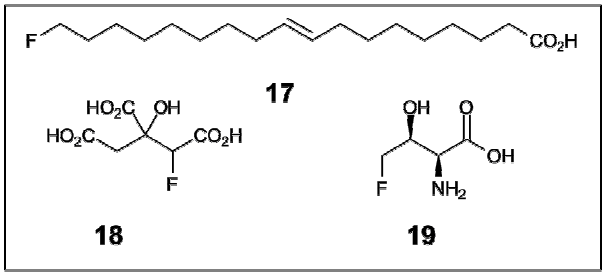


Figure 4.12: Examples of fluorinated natural products. ω-fluorooleic acid **(17)**, **f**luorocitrate **(18)**, 4-fluorothreonine **(19)**.

to be present inside a hydrophobic pocket where desolvation of the fluorine occurs as protein residues in the active site coordinate to water. The crystal structure of FlA, solved by Dong *et al.,* shows that the enzyme is a trimer of dimers, containing three SAM-binding active sites localised at the interfaces between the subunits.[185] Prior to nucleophilic attack on the SAM, the fluoride ion is stabilised by at least three hydrogen bonds to the surface of the protein which stabilises the F^- ion and allows it to act as a nucleophile (see figure 4.13). The crystal data, assisted by quantum and molecular mechanic calculations (QM/MM) suggests that in the transition state the fluoride ion forms two hydrogen bonds to Ser158 and one to the hydroxyl side chain of Thr80 (figure 4.13) within the active site. In addition, the electrostatic interaction between fluorine and the positively charged sulphur of SAM further stabilises the fluoride ion.[186]

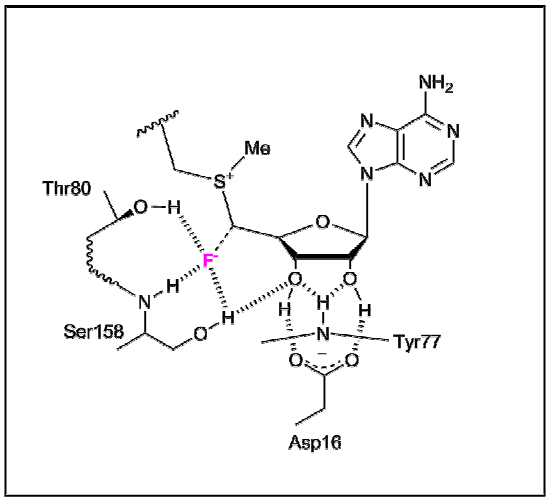


Figure 4.13: Hydrogen bonding to the fluoride ion in the active site of the fluorinase FlA. Prior to nucleophilic attack on the SAM, the fluoride ion is stabilised by at least three hydrogen bonds to the surface of the protein (two hydrogen bonds to Ser158 and one to the hydroxyl side chain of Thr80) which stabilises the fluorine ion in the absence of a hydration shell and allows F- to act as a nucleophile.

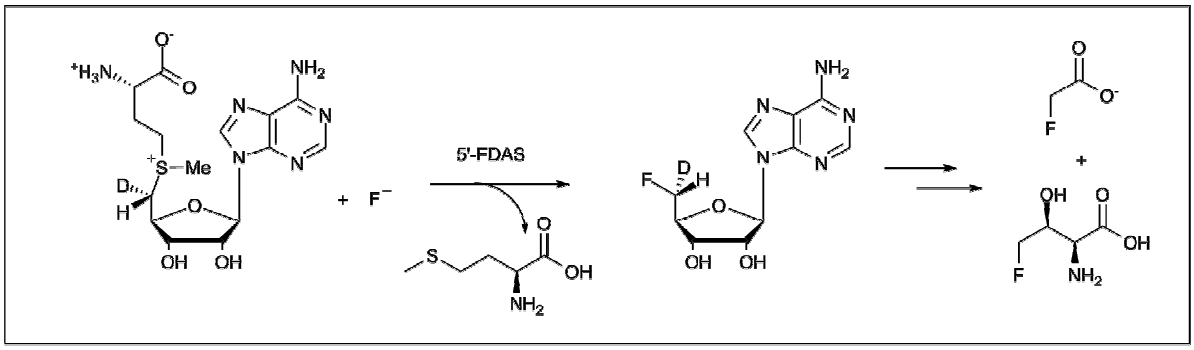


Figure 4.14: Proposed sequence of events leading to conversion of inorganic to organic fluoride. 5'-fluoro-5'-deoxyadenosine synthetase (5'FDAS) mediates the desolvation of fluoride and subsequent S_N2 substitution reaction with SAM that proceeds with inversion of configuration to yield 5'-fluoro-5'deoxyadenosine which is subsequently converted by other enzymes to fluoroacetate and 4-fluorothreonine.

Feeding 5'-*pro-S* deuterium labelled SAM to the enzyme revealed that the overall reaction proceeds with inversion of configuration which is consistent with an S_N2 substitution reaction with the methionine of SAM and the fluoride ion, resulting in the formation of 5'-fluoro-5'-deoxyadenosine (figure 4.14).[187] The hydrogen bond mediated stabilisation of the F^- ion in the active site allows a reduction in the activation energy from 92 kJ mol^{-1} in solution to 53 kJ mol^{-1} within the enzyme, this reduction in activation energy is what allows the enzyme to function.[188]

The discovery of FlA in 2002[189] has driven research into the potential of regio-selective fluorination under relatively mild conditions which would make this class of enzyme very useful in the field of biotechnology. In particular there is currently a lot of interest in the synthesis of 18F-labelled sugars and nucleosides as radiopharmaceuticals for positron emission tomography (PET) imaging (a nuclear medicine imaging technique). The short half life of the 18F isotope means that the synthesis of the required fluorine containing ligands needs to be fast, efficient and easy to achieve in a clinical setting and recently biotransformations using this fluorinase enzyme have been examined for use in this field.[190] Successful biotransformations have been reported of radiolabelled 18F-5'-FDA (5'-fluoro-5'-deoxyadenosine), 18F-5'-FDI (5'-fluoro-5'-deoxyinosine) and 18F-5-fluororibose from 18F-fluoride in radiochemical yields of up to 95% in under two hours.[191]

4.2.4 PrnC – a pyrrole halogenase

Pyrrolnitrin is a tryptophan derived natural product that demonstrates broad-spectrum fungicidal activity and was first isolated from *Pseudomonas pyrrocinia*.[192] As mentioned previously the biosynthetic cluster was discovered in 1997 and four key enzymes were identified to be involved in the biosynthesis (PrnA-D, figure 4.15).[170] Subsequently the cluster has also been identified in a number of other organisms.[193]

L-Tryptophan **(20)** is the initial substrate for the biosynthetic pathway; this undergoes halogenation at the 7 position by the highly characterised PrnA halogenase, following the mechanism described in detail previously, to form 7-chloro-L-tryptophan **(21)**. The second stage of the biosynthesis is mediated by the enzyme PrnB which catalyses an unusual ring rearrangement and elimination of the carboxylate group to form monodechloroaminopyrrolnitrin **(22)**. This enzyme is a member of the heme-b-dioxygenase superfamily and is closely related to indoleamine 2,3-dioxygenase (IDO) although does not show the same functionality despite utilising a similar substrate. The crystal structure of PrnB shows that this difference in activity is due to the orientation of the tryptophan in

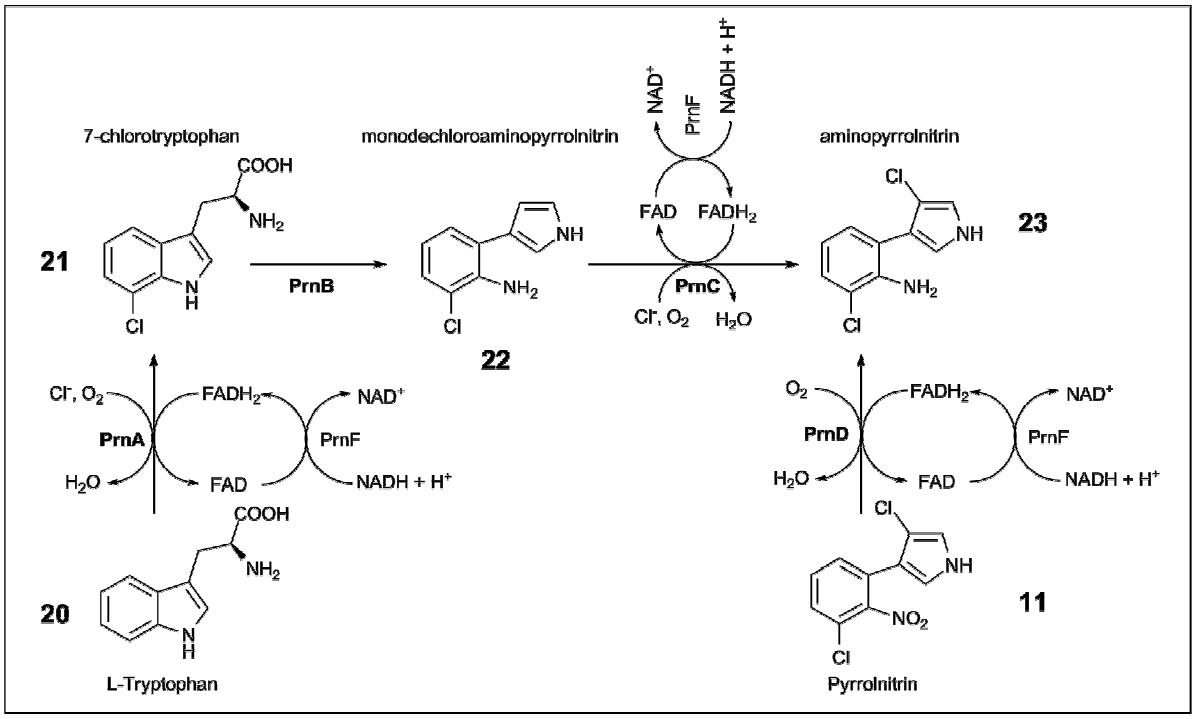


Figure 4.15: Biosynthetic pathway for pyrrolnitrin. Biosynthesis of pyrrolnitrin starts from L-tryptophan and is mediated by four enzymes PrnA-D. The biosynthesis proceeds through chlorination, ring rearrangement, a second chlorination and a final oxidation to reveal the final product.

the active site.[194] In PrnB the tryptophan is orientated to block access of the catalytic iron residues to the sites favoured by traditional dioxygenase enzymes. Following this ring rearrangement (figure 4.15), the resulting pyrrole moiety is then halogenated again at the 3 position by PrnC. The resulting aminopyrrolnitrin is then oxidised by PrnD, to form the final product. PrnD is another unusual enzyme that catalyses an arylamine oxidation reaction in a flavin-dependent manner.[195]

The pyrrolnitrin biosynthetic cluster contains many unusual enzymes that may have great potential as future valuable biocatalysts. Perhaps most attractive of them is the pyrrole halogenase PrnC. As noted before halogenation is a simple change that can result in vast changes in biological activity. Introducing halogens into pyrrole species in a regio-selective fashion and under mild conditions is something that is hard to do with conventional chemical techniques and presents PrnC as an attractive target enzyme for biocatalysis.

Although PrnC is a largely uncharacterised halogenase enzyme the protein sequence contains both the flavin binding motif and the double tryptophan motif that are the hallmarks of $FADH_2$-dependent halogenase enzymes. PrnC has very low sequence identity (14%) to the other flavin-dependent halogenase in the cluster, PrnA. This is perhaps not unexpected due to the large variation in protein sequence seen in these types of enzyme, even between PrnA and RebH which catalyse identical reactions. The protein sequence of PrnC shows very little similarity to any other enzyme, the closest protein homolog is HalA from an *Actinoplanes sp* with 42% homology (which is unsurprising since PrnC probes were used to identify it as a halogenase enzyme).[196] This enzyme is involved in the biosynthesis of pentachloropseudilin **(24)**, a natural product that has a structure very similar to pyrrolnitrin. However it is not known whether it is a specific pyrrole halogenase as the exact substrate of HalA is ambiguous.[196]

Despite the lack of protein homologs there are several known flavin-dependent halogenases that produce chlorinated pyrrole moieties (figure 4.16), such as PltA in pyoluteorin biosynthesis **(14)**, HrmQ during hormaomycin production **(25)** and Pyr29 in pyrrolomycin A biosynthesis **(26)**.[197, 198] In these examples, however, the substrates are produced via PKS/NRPS pathways and are bound to peptidyl carrier proteins during the halogenation. In the case of PltA it has been shown that only the carrier protein bound substrate is accepted and the enzyme will not halogenate free pyrrole-2-carboxylate[183] and it has been suggested that HrmQ has the same restriction.[149]In the biosynthesis of pyoluteorin, proline is bound as a thioester to the PltL carrier protein and is subsequently desaturated to a pyrrolyl-*S*-PltL intermediate which is then chlorinated regio-selectively at C-5 followed by a subsequent chlorination at C-4 by the same enzyme, PltA (see figure 4.17).

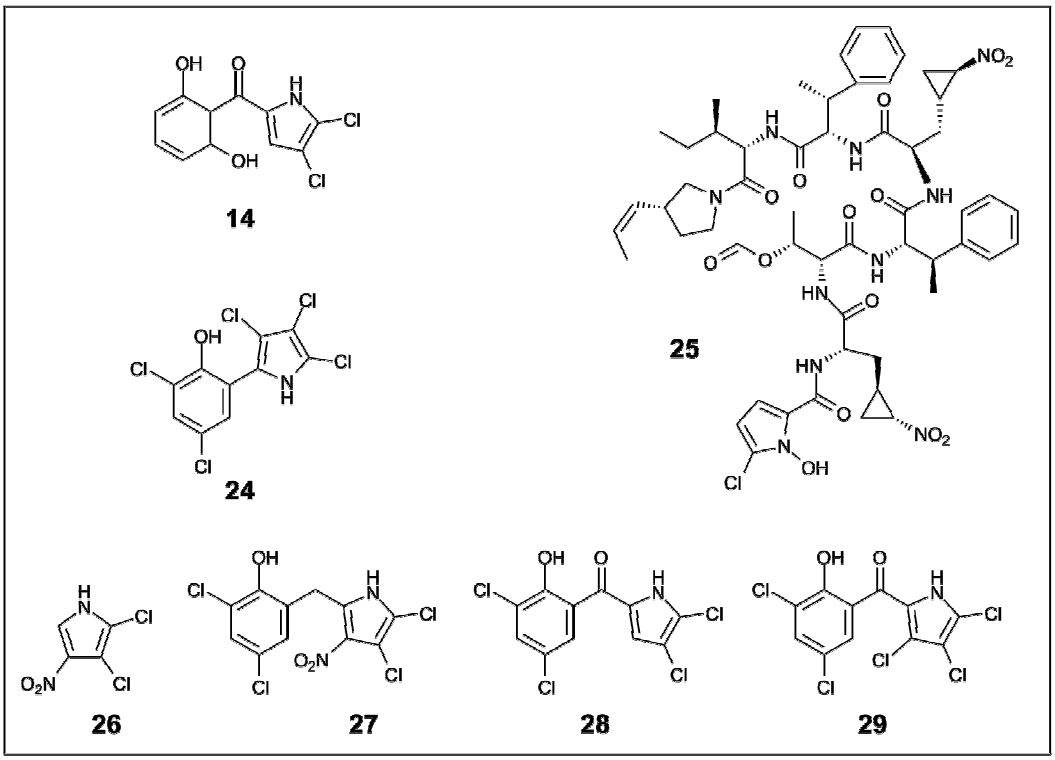


Figure 4.16: Natural products with chlorinated pyrrole moieties halogenated by $FADH_2$-dependenet halogenases. Pyoluteorin **(14)**, pentachloropseudilin **(24)**, hormaomycin **(25)**, pyrrolomycin A-D **(26-29)**.

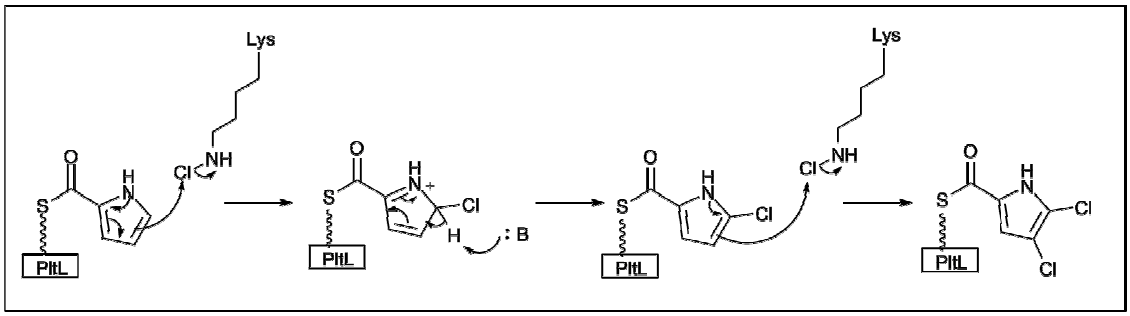


Figure 4.17: Proposed mechanism of PltA meditated chlorination. Mechanism based on chloramine mechanism of flavin-dependent halogenation.[182] Proline is bound as a thioester to the PltL carrier protein. It is then desaturated to a pyrrolyl-*S*-PltL intermediate which is then chlorinated regio-selectively at C-5 by PltA. The same enzyme then catalyses a second chlorination at C-4. Diagram adapted from Dorrestein (2005).[183]

PrnC, in contrast, halogenates a free substrate (monodechloroaminopyrrolnitrin) and is not involved in PKS or NRPS machinery. There are not currently any other confirmed examples of another pyrrole halogenase that acts in this manner. The close homolog HalA may be able to halogenate a free substrate but not enough is currently known about this enzyme to say for certain. Pyrrolomycin A is halogenated by Pyr29 as a carrier protein bound intermediate; however this natural product contains many analogues including pyrrolomycin B, C and D **(27-29)**. These analogues are produced from pyrrolomycin A following the release from the PKS enzymes, alterations include the addition of a phenyl ring and further chlorination steps (figure 4.16). Pyr29 may also be involved in these later chlorinations and may therefore also function on free substrates, although other enzymes may be involved.[198]

At the present time however PrnC is the only enzyme with proven (*in vivo*) activity on an isolated substrate and therefore represents a potentially more flexible candidate for biotechnological applications. Interestingly it is also one of the few examples were a pyrrole moiety is selectively halogenated only once (the other being HrmQ from hormaomycin biosynthesis). Characterisation of this enzyme and determination of substrate specificity would not only be useful in helping to understand the general mechanism of the flavin-dependent halogenases but would also present a fine candidate for the generation of novel halogenated compounds.

4.2.5 Biotechnological potential for $FADH_2$-dependent halogenases

The diversity seen in the halogenated natural products available to nature demonstrates the profound effect halogenation can have on activity. Harnessing these enzymes could provide a route to halogenation at specific sites of substrates that are perhaps not possible with direct chemical halogenations. They may be useful either in the search for better more effective drugs or simply as a tool in chemical synthesis.

The incorporation of these halogenases into the biosynthetic clusters of several natural products has already resulted in the production of numerous unnatural halogenated metabolites. One of the first examples was carried out by Heide *et al.*, where the HrmQ halogenase from hormaomycin biosynthesis **(25)** was used to generate two new chlorinated clorobiocin derivatives. Clorobiocin **(6)** (figure 4.2) normally contains a 5-methylpyrrole moiety but deletion of the relevant methyltransferase and expression of the HrmQ halogenase resulted in the new analogues being produced.[197] More recently the group of Sarah O'Connor demonstrated that the insertion of the gene for the tryptophan-7-halogenase RebH or the tryptophan-5-halogenase PyrH into the medicinal plant *Catharanthus roseus* resulted in the generation of regio-selectively chlorinated alkaloid

products that were derived from tryptophan.[199] In complimentary work Goss *et al.* were able to generate chlorinated analogues of the tryptophan containing pacidamycin antibiotic from *Streptomyces coeruleorubidus* by the insertion of the tryptophan-7-halogenase PrnA into the producing organism. This resulted in pacidamycin analogues that contained a 7-chloro-tryptophan residue.[200] As discussed in chapter 1, the incorporation of chlorine at this position resulted in changes in activity.[88] This previous work however had been conducted by feeding halogenated tryptophan produced by tryptophan synthase biotransformation reactions, the incorporation of the halogenase gene made the production of chlorinated pacidamycin more efficient. In addition to this Suzuki-Miyaura cross-coupling reactions were performed with the pacidamycin analogues under aqueous conditions using a method pioneered in the same group to produce a number of new semi-synthetic analogues with a variety of aryl substituents.[200, 201] This technique could be used to generate libraries of new natural product analogues with interesting pharmacological properties.

The main problem with the flavin-dependent halogenase enzymes however is their strict substrate specificity. PrnA has been show to accept a number of other indole derivatives but the regio-selectivity is lost which limits the enzymes potential.[202] Within pyrrolnitrin biosynthesis it has also been shown that PrnC is unable to accept monodechloroaminopyrrolnitrin as a substrate when it is chlorinated at the 6 rather than the 7-position of the phenyl group.[174]

The origin of the regio-selectivity of the flavin-dependent tryptophan halogenases was determined in 2009 when the structure of PyrH was solved. It was discovered that the binding of the tryptophan substrate in the active site exposed the C5 position to the key catalytic lysine residue.[203] The other reactive positions around the indole ring are all protected by bulky, aromatic amino acid side chains, essentially mimicking protecting group chemistry. This information helped to explain why PrnA lost regio-selectivity when tested with other indole derivatives as orientation in the active site is very important. Using this knowledge site-directed mutagenesis was used to alter the active site of PrnA to remove some of the bulkier residues. This resulted in a mixture of 7- and 5-chloro-tryptophan being produced.[204] The absence of di-chlorinated product was suggested to be due to a lack of space in the active site.

Recently a new fungal $FADH_2$-dependent halogenase from the fungi *Pochonia chlamydosporia* has been identified, Rdc2, which is similar to HrmQ as it is also a post-PKS halogenase.[205] This halogenase has been shown to halogenate various natural products into mono- or di-halogenated derivatives. This highlights the potential of PrnC as a potential new pyrrole halogenation tool and how a greater understanding of its structure and function could lead to a general halogenase capable of regio-selective halogenation of pyrrole moieties.

Chapter 5:

Overexpression and Purification of PrnC

5.1 Promoter systems used to induce protein expression

The tryptophan synthase containing plasmid pSTB7, used extensively in the previous chapters, is constitutively active and therefore continuously expresses the contained genes. Generally however this is not desirable when producing recombinant proteins as unregulated production can affect culture growth rates, decrease the protein yield and can lead to protein aggregation and the formation of inclusion bodies.[206] If the recombinant protein is toxic to the producing organism premature production can also affect the culture viability and lead to loss of the plasmid from significant numbers of the cell population. Therefore it is advantageous to delay protein expression until optimal growth rate is reached and before nutrients begin to become scarce. For this reason most expression systems are regulated under a promoter which must be induced before protein production can begin. Two of the most commonly used are the T7lac promoter and the araBAD promoter.

5.1.1 T7lac promoter

The T7 promoter was first developed by Studier *et al.* and is based on the bacteriophase T7 RNA polymerase.[207] This polymerase is very selective and so active that most of the host cells resources are channelled into target gene expression. Target genes are cloned into plasmids that are under the control of T7 transcription and translation signals (recognised by the T7 RNA polymerase). *E. coli* does not naturally contain this RNA polymerase; so to enable protein expression from target genes under the control of this promoter expression hosts that contain a chromosomal copy of the T7 RNA polymerase gene (such as *E. coli* BL21) must be used. This ensures that the target gene cannot be transcribed by the natural *E.coli* RNA polymerase and so controlling T7 RNA polymerase production affectively controls target gene expression. The T7 polymerase itself is under the control of the *E. coli lac* promoter.[207]

In the T7lac system expression of both the vector and the integrated T7 RNA polymerase gene is provided under *lacUV5* control. This involves a *lac* operator sequence downstream of the T7 promoter and a coding sequence for the *lac* repressor protein LacI (both taken from the *lac* operon which controls the expression of lactose metabolism genes in *E. coli*). In the absence of lactose the repressor protein encoded by *lacI* can bind to the operator sites, effectively preventing the T7 RNA

polymerase from transcribing the target gene. Transcription of the T7 RNA polymerase gene itself by the natural polymerase of the host cell is also inhibited by the *lac* repressor protein.

When lactose (or non-hydrolysable mimic isopropyl β-D-1-thiogalactopyranoside, IPTG) is introduced this binds to and displaces the lac repressor, which in turn leads to the loss of inhibition of T7 RNA polymerase production which allows the binding of the T7 polymerase to the T7 promoter region on the plasmid. This switches on the production of any protein encoded by the genes downstream of the T7 promoter.

In organisms that naturally carry the *lac* operon it is used to ensure that the genes required to metabolise lactose as a carbon source are only activated if necessary i.e. when glucose is not available but lactose is. Therefore as a double control glucose also acts as a repressor of this system, limiting the association between the RNA polymerase and the DNA in high glucose concentrations. This is achieved by binding of the catabolite activator protein (CAP) which assists in the binding of the RNA polymerase to the DNA. To be activated CAP must be first bound to cyclic adenosine monophosphate (cAMP), the abundance of which is inversely proportional to glucose concentration.[208, 209]

This gene expression system is widely used in the pET system (Novagen).

5.1.2 araBAD promotor

The araBAD system of protein expression utilises a different catabolic pathway regulatory system to control target gene expression, this time from the L-arabinose operon controls the breakdown of arabinose into D-xyulose-5-phosphate that can be metabolised by the pentose phosphate pathway. The operator and promoter regions *araO* and *araI* and the *araC* gene that encodes for an activator protein are included in expression plasmids, such as pBAD, that utilise the control of this promoter system.

In the absence of arabinose the activator protein binds to the *araI* and the *araO* regions of the DNA which results in the formation of a 210 base pair loop which prevents the association of the RNA polymerase machinery with the DNA, preventing target gene expression. When arabinose is present, the activator forms a different conformation which allows the RNA polymerase to bind to the promoter region and transcribe the downstream genes. Similar to the T7lac promoter CAP binding is also required; therefore glucose inhibits the expression of downstream genes in the same way.[210] Because DNA polymerase binding is actively inhibited in the absence of arabinose, proteins under the arabinose promoter generally have very low basal expression levels.

5.2 Results and Discussion

5.2.1 Halogenase work prior to start of PhD

Work conducted during an industrially funded summer studentship (unpublished results) focused on the cloning, expression and substrate specificity of four flavin-dependent halogenase enzymes. Three of these were tryptophan halogenases from different natural product biosynthetic clusters; PyrH, Thal and PrnA (tryptophan 5,6 and 7-halogenases respectively). The final enzyme was the third enzyme in the biosynthesis of pyrrolnitrin, PrnC. These four enzymes were cloned into the pET21a(+) (Novagen) vector and expressed in *E. coli* BL21. Following cell lysis all four enzymes were present in the insoluble fraction when grown at 37, 28 or 16°C with PrnC in particular showing almost no solubility at all. Transfer of the three tryptophan halogenase genes into pET28a(+) (Novagen) resulted in an increase in soluble protein produced at 16°C. Of all these enzymes PrnC is perhaps the most interesting from a mechanistic point of view and is the least characterised of all the four enzymes tested. Placing the gene for PrnC into an expression vector other than pET21a(+) may improve the low levels of solubility seen previously and help to produce sufficient quantities of purified enzyme to facilitate characterisation and possible crystallisation.

5.2.2 Production of PrnC under the araBAD promoter

5.2.2.1 Bioinformatics and strategy for improving PrnC expression

In the work conducted prior to the start of this PhD, good levels of soluble protein for the three tryptophan halogenase enzymes had been achieved but no improvement could be made on PrnC production. The protein sequences of the three tryptophan halogenase enzymes PyrH, Thal and PrnA and the pyrrole halogeanse PrnC were compared in an attempt to understand this difference in solubility.

Multiple protein sequence alignment (figure 5.1) showed that the three tryptophan halogenase enzymes shared at least 41% homology with each other with the most similar enzymes being PrnA and Thal (56%). Comparison between these three halogenases and PrnC showed a significantly lower sequence homology. The most similar to PrnC was observed to be Thal but even these enzymes only share 17% sequence similarity.

This is perhaps not a surprising result. The three tryptophan halogenases utilise the same substrate whereas PrnC selectively chlorinates a pyrrole moiety. Despite all the differences, however, there are a few similarities that all three enzymes share with the rest of the flavin-dependent family. There

are two conserved motifs present within the sequence of all the flavin-dependent halogenases. The first is a flavin binding module located near the N-terminus of the protein, this has the consensus sequence GxGxxG.[211] The second conserved motif has the consensus sequence WxWxIP and is located closer to the centre of the protein. This module was suggested to prevent the enzyme from functioning as a monooxygenase by restricting substrate binding close to the FAD[211] but mutagenesis evidence failed to support this hypothesis and currently the function of this conserved region is unknown.[179] Both of these conserved sequences are present in PrnC as well as the three tryptophan halogenase enzymes (these sequences are highlighted in red in the multiple protein sequence alignment in figure 5.1) which is to be expected as all are assigned as $FADH_2$-dependent enzymes.

Alignment of the three tryptophan halogenases highlights the conserved lysine residue (K79 for PrnA) which is necessary for catalysis for all $FADH_2$-dependent halogenases and a glutamate residue (E346 for PrnA) which is only conserved and essential for activity within the $FADH_2$-dependent halogenase enzymes that utilise tryptophan as the substrate. The alignment did not identify these residues in PrnC but the catalytic lysine should be present for the enzyme to mediate halogenation using the same mechanism, but the location of this residue in the sequence of PrnC is unknown. There are 28 lysine residues in PrnC and in the absence of a crystal structure site-directed mutagenesis of each of these sites would be required to ascertain which of them have catalytic activity.

Sequence A	Length	Sequence B	Length	Score
PyrH	511	PrnA	537	41.0
PyrH	511	Thal	531	41.0
PrnA	537	Thal	531	56.0
PyrH	511	PrnC	567	10.0
Thal	531	PrnC	567	17.0
PrnA	537	PrnC	567	14.0

Table 5.1: Protein sequence Homology of Four Flavin-dependent Halogenase Enzymes. PrnA and Thal share the most sequence similarity, while PrnC and PyrH share the least.

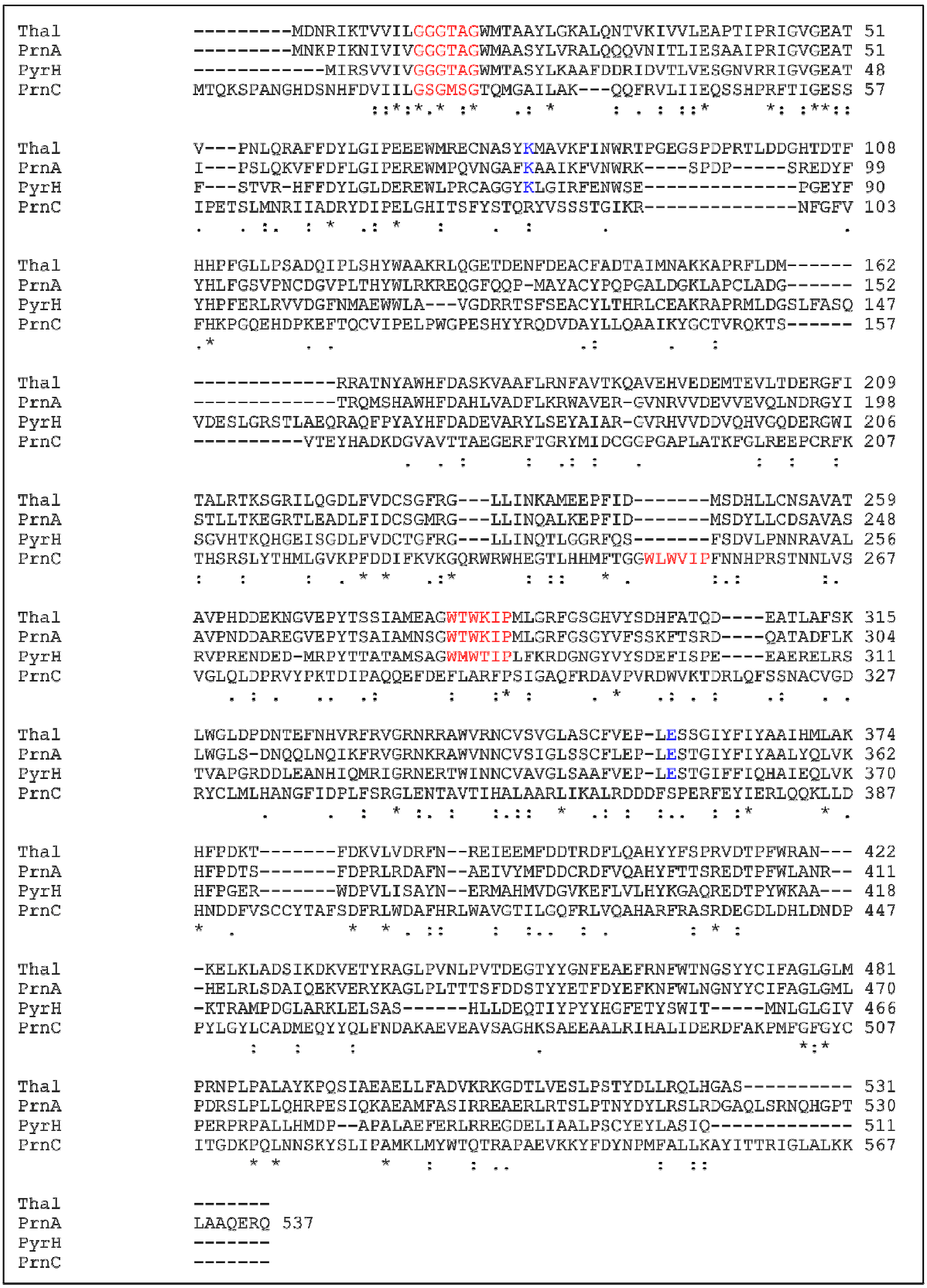


Figure 5.1: Multiple sequence alignment of four flavin-dependent halogenase proteins. Identical residues highlighted with an asterix, conserved residues with two dots and semi-conserved with a single dot. Family conserved motif (WxWxIP) and flavin binding motif (GxGxxG) are highlighted in red. Key catalytic residues present in PrnA (K79 and E346) are highlighted in blue.

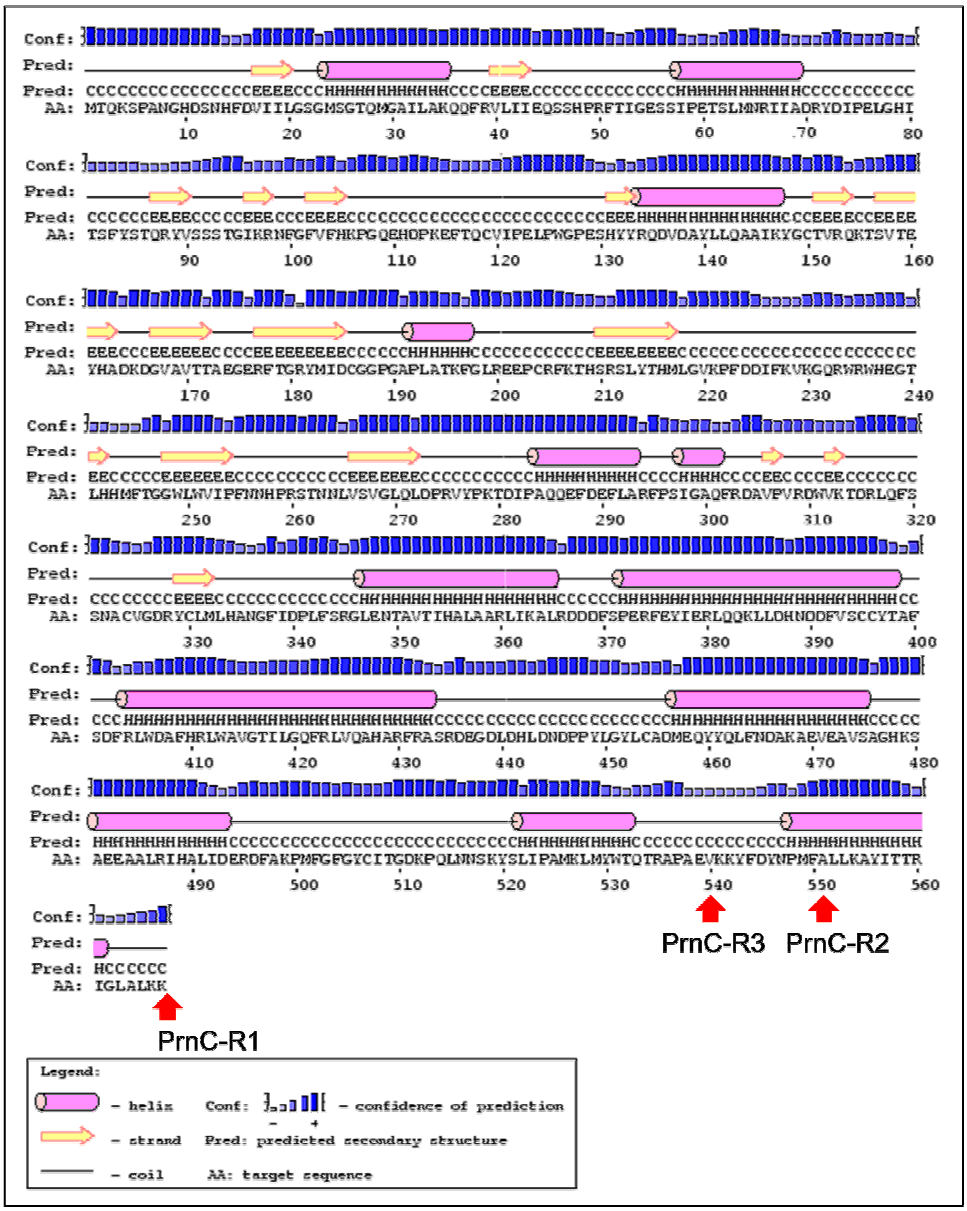


Figure 5.2: Secondary structure prediction of PrnC based on protein sequence. Prediction based on the amino acid composition of PrnC using the online PSIPRED predictor from the Bloomsbury Centre for Bioinformatics from University College London.[212] Truncation positions of produced PrnC variants (R1-R3) are highlighted with arrows.

Due to the lack of homology between PrnC and the other enzymes tested, it is difficult to suggest methods to improve the expression of folded, soluble protein based on the sequence alignment alone. PrnC is a slightly larger enzyme than the others (table 5.1) and contains more positively charged amino acids (lysine and arginine) at the C-terminal end of the protein (figure 5.1). This difference may be enough to alter the solubility of PrnC relative to the other halogenase enzymes tested or help to target PrnC for faster degradation. The possibility that the enzyme could be

truncated to remove some of these positively charged amino acids was explored in an attempt to improve folding; however this could ultimately have an effect on activity.

The use of online secondary structure prediction tools (PSIPRED from UCL)[212, 213] enabled a rough assessment of the degree of structural complexity present at the C-terminus of PrnC. The analysis suggested that the terminal region of the enzyme contained some significant secondary structure (figure 5.2). Ideally truncations should be made in regions that contain no structural features as these are usually involved in activity or conformation. To this end any truncations introduced at the C-terminal end of PrnC may result in loss of activity or further problems with folding.

5.2.2.2 Cloning of PrnC variants and ligation into pBAD vector

The pBAD/HisA plasmid was chosen as the protein expression vector. This is a pBR322 derived plasmid designed for regulated, dose-dependent recombinant protein expression and purification in *E. coli* (see figure 5.3). The vector utilises the araBAD promoter from *E. coli* which is induced in the presence of L-arabinose. Optimum levels of protein expression can be produced by adjusting the concentration of arabinose used. The vector carries an N-terminal hexa-histidine tag which carries some advantages over the C-terminal tag used in the previous work and may help to improve solubility. Proteins are synthesised by the ribosome from the N-terminal end and the folding process often begins co-translationally, therefore installing an affinity tag at the C-terminal end of the protein ensures that only full length proteins are purified. A disadvantage is that because of the way protein folding occurs, C-terminal tags are more likely to become internalised inside the protein,[214] meaning they are unavailable for purification, they can also make the resulting protein more unstable.

To facilitate the cloning of *prnC* into pBAD forward and reverse PCR primers were designed to flank the *prnC* gene and insert unique restriction sites to enable insertion into the vector. Comparison of the restriction sites inside the pBAD multiple cloning site (MCS) with internal restriction sites present within the p*rnC* gene identified BglII as an appropriate restriction site for the forward primer. Most of the other sites in the MCS were also present inside *prnC* so could not be used for the cloning. This problem was circumvented by augmenting the reverse primer with an MfeI (MunI) recognition site. Following digestion MfeI generates sticky ends that are compatible with EcoRI, meaning that although the initial recognition sites are different the resulting overhangs from each enzyme are able to ligate together with the consequential loss of the original recognition site for both enzymes. These compatible ends are illustrated in figure 5.3. A single forward primer was therefore designed, incorporating the BglII restriction site, the primer was designated PrnC-Bgl. Three reverse primers

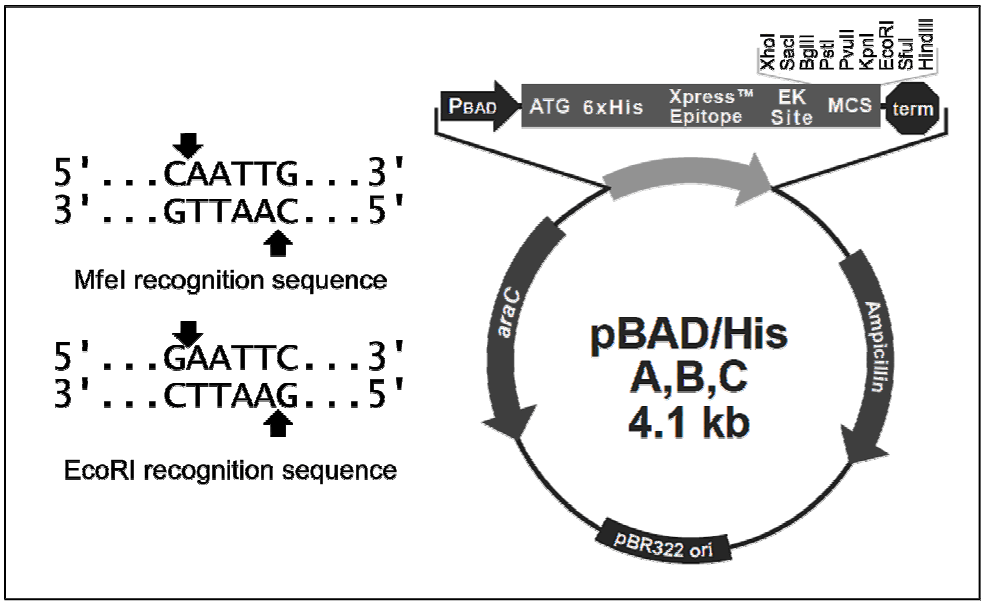


Figure 5.3: Recognition sequences for MfeI and EcoRI restriction enzymes and plasmid map of pBAD/His vector showing cloning region. The two selected restriction enzymes MfeI and EcoRI have slightly different recognition sites; however the DNA products after digestion share the same overhang sequence meaning the products of the two enzymes can be ligated together. The features of pBAD are shown in the plasmid map including araC promoter region, ampicillin resistance and the multiple cloning site.

were designed with the MfeI restriction site, each providing a route to one of three truncated PrnC proteins. The first reverse primer was designed to amplify the entire gene and was designated PrnC-R1. The second reverse primer removed 51 base pairs of DNA (17 amino acids) from the C-terminal end of *prnC* (designated PrnC-R2). The final primer removed 84 base pairs of DNA (28 amino acids) from the end of the gene (designated PrnC-R3). In all three primers the natural stop TAG codon present in the p*rnC* gene was substituted with TAA, as there is a higher incidence of this codon in *E. coli*.

The selected truncation sites are displayed in figure 5.2. The first truncated PrnC analogue (PrnC-R2) was designed to remove the majority of the positively charged lysine and arginine residues at the end of the protein. The potential problem with these truncations is the large number of predicted alpha helices that reside at the C-terminal end of the protein. The PrnC-R2 analogue cuts through one of these predicted helices and therefore may reduce activity or protein folding integrity. The second PrnC variant (PrnC-R3) removes further positively charged residues, further truncating the protein. This truncation entirely removes a predicted helix which may result in the loss of activity and stability. The secondary structure is just a prediction however and may not represent the actual structure accurately.

A preliminary polymerase chain reaction (PCR) amplification was carried out using the forward primer and each of the three reverse primers. A range of annealing temperatures was tested (54.8, 58.4, 64.0 and 68.2°C) and the effect of the addition of 5% dimethyl sulfoxide (DMSO) was also examined. *Pfu* polymerase was used which operates at an elongation rate of 2 min per kb. Therefore as *prnC* is slightly less than 2 kb in size an extension time of 4 minutes was used. The results of the various PCR reactions were analysed on an agarose electrophoresis gel (visualised with ethidium bromide under UV light) and can be seen in figure 5.4 (parts A and B). No PCR product for *prnC-R1* was seen at any temperature lower than 68.2°C. 5% DMSO addition was also seen to be necessary for a band of the correct size to appear on the gel.

In the presence of DMSO, the PCR product of *prnC-R2* was seen to be produced at all temperatures except the lowest (54.8°C). In the absence of DMSO only an annealing temperature of 64°C showed any evidence of product. An annealing temperature of 68.2°C with 5% DMSO showed the least number of non-specific amplifications (figure 5.4B) and therefore was selected as the optimum conditions for this PCR.

None of the conditions tested produced any of the *prnC-R3* variant. In an attempt to improve the amplification of the heavily truncated *prnC-R3* the effect of increasing the DMSO concentration to 10% or the addition of 2.5% formamide was investigated. DMSO is known to improve the amplification of DNA by disrupting base-pairing and reducing inter- and intra- strand re-annealing therefore reducing the incidence of secondary DNA structure such as hairpin loops that can interfere with recognition by the primer.[215] Formamide has also been shown to improve the specificity of PCR reactions.[216] It has been proposed that organic additives like formamide can bind to the major and minor grooves of the DNA, destabilising the double helix,[217] as a consequence of this the melting temperature of DNA is reduced which improves the efficiency of the PCR reaction at lower annealing temperatures. All of the primers were tested with the new conditions in combination with increasing the annealing temperature to 72°C. The resulting PCR products were analysed by electrophoresis and visualised gels can be seen in figure 5.4C.

For the *prnC-R1* and *–R2* reactions adding increased amounts of DMSO or formamide drastically reduced the effectiveness of the PCR reaction and the effect of increasing the annealing temperature to 72°C seemed to reduce the specificity of the reaction relative to the previous optimum conditions. Increased DMSO or formamide also did not show any improvement to the amplification of *prnC-R3*, however keeping the DMSO concentration to 5% but increasing the annealing temperature did show clear bands, for the first time, indicating the production of the

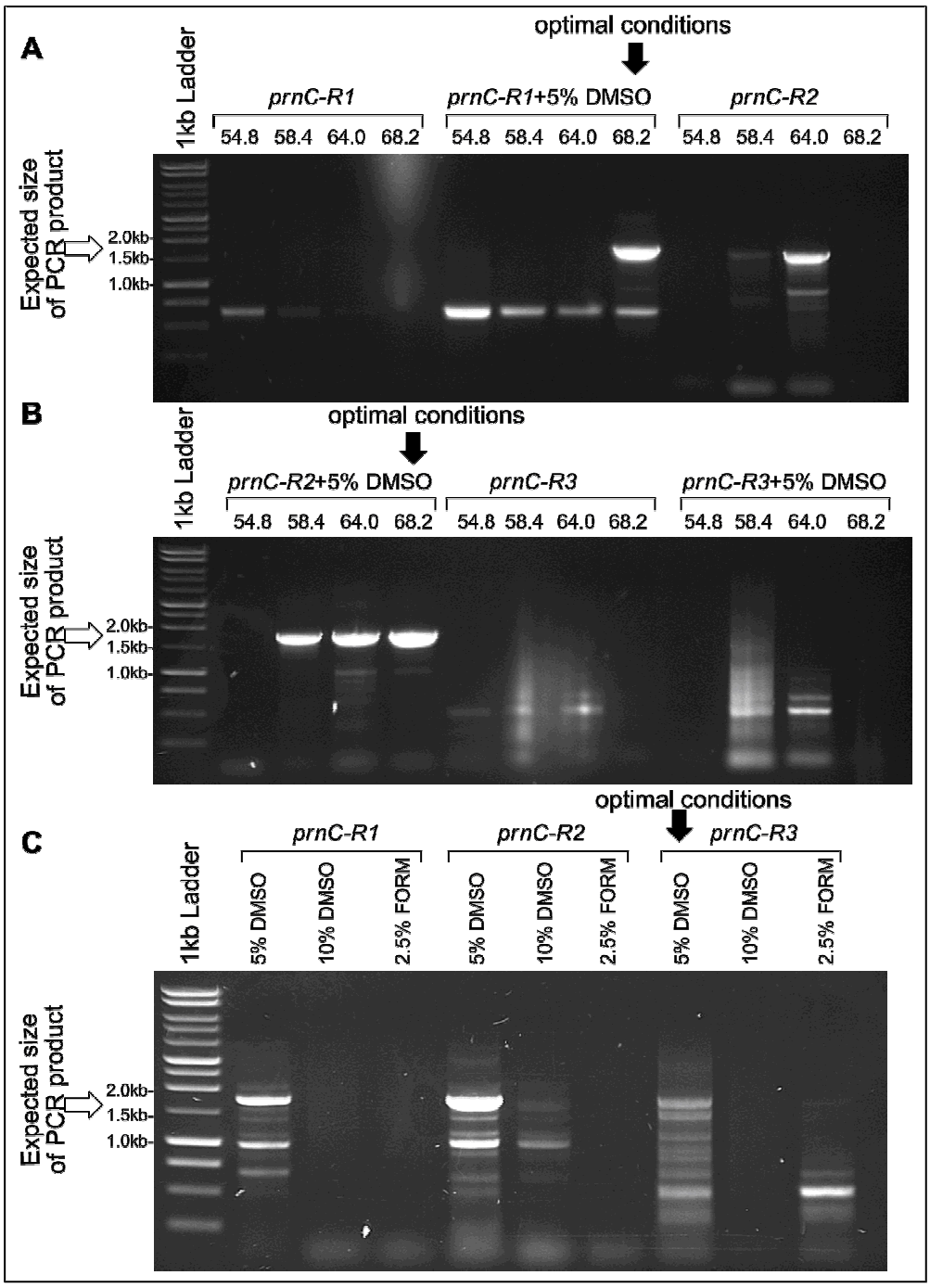


Figure 5.4: DNA gel electrophoresis analysis of *prnC* gene amplification under a variety of different conditions. (A, B) Analysis of PCR reactions carried out with the three different *prnC* primer combinations at four different annealing temperatures and the effect of adding 5% final concentration of DMSO to the reactions. **(C)** PCR products resulting from the increase in DMSO concentration to 10% or the addition of 2.5% formamide (FORM) to PCR reactions carried out with a 72°C annealing temperature. Expected PCR products are *prnC-R1* (1704 bp), *prnC-R2* (1653 bp) or *prnC-R3* (1620 bp) and are marked with a white arrow. 1kb ladder (promega) sizes from top: 10000, 8000, 6000, 5000, 4000, 3000, 2500, 2000, 1500, 1000, 750, 500, 250bp. Selected optimal PCR conditions for scale up are highlighted with a black arrow.

truncated *prnC* variant R3. These conditions for R3 were chosen for scale up as they represented the only successful reaction.

Based on the PCR optimisation results the two longest *prnC* variants (R1 and R2) were scaled up to 100 µl PCR reactions. The less specific R3 reaction was scaled up to 200 µl scale. Due to the high number of low molecular weight amplifications in this case, the extension time for the PCR reaction was increased to 10 minutes. Following this, the PCR products were purified by gel electrophoresis. The PCR product of *prnC-R3* visualised by EtBr produced several bands but clearly showed that only the top band was of the correct size to be the product, this band only was excised with the other two variants. The resulting purification yielded much lower quantities of DNA than the other two (R1 and R2) but was sufficient to continue to the next step.

Following purification the PCR products were digested with BglII and MfeI restriction enzymes and ligated into similarly digested pBAD vector (BglII, EcoRI). The three resulting plasmids were identified as pMW03-05. The plasmids were sequenced (DNA sequencing facility, University of Cambridge) to ensure no mutations had arisen during the PCR and then transformed into chemically competent *E. coli* XLI-Blue cells for long term storage (MW-005, -006 and -007) and into *E. coli* BL21 for protein expression studies (MW-008, -009 and -010).

5.2.2.3 Protein expression of PrnC variants in *E. coli*

With the araBAD promoter system the level of protein production can be controlled by varying the concentration of the L-arabinose inductant. In general terms the closer protein production is to optimum then the better the overall protein solubility will be. A pilot study was first undertaken to determine the optimum concentration of arabinose for PrnC production. A range of 10 ml cultures inoculated with *E. coli* BL21 MW-008 were allowed to reach an optical density of 0.5 (at 600nm) at 37°C before being induced with a range of arabinose concentrations (0.2 – 0.00002% (*m/v*)) and returned to incubation for 4 hours. Subsequent analysis by SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) showed very low levels of protein expression even with 0.2% arabinose.

Larger scale cultures (500 ml) of all three PrnC variants (*E. coli* BL21 MW-008, MW-009 and MW-010) inoculated with ampicillin were grown in the same way and induced with 0.2% L-arabinose. Following induction the cells were transferred to a lower temperature of 16°C for 24 hours to promote better protein folding. The cell lysate was then generated *via* sonication. 10 ml of this cell lysate was subjected to small scale Ni-NTA spin purification to isolate the PrnC enzyme.

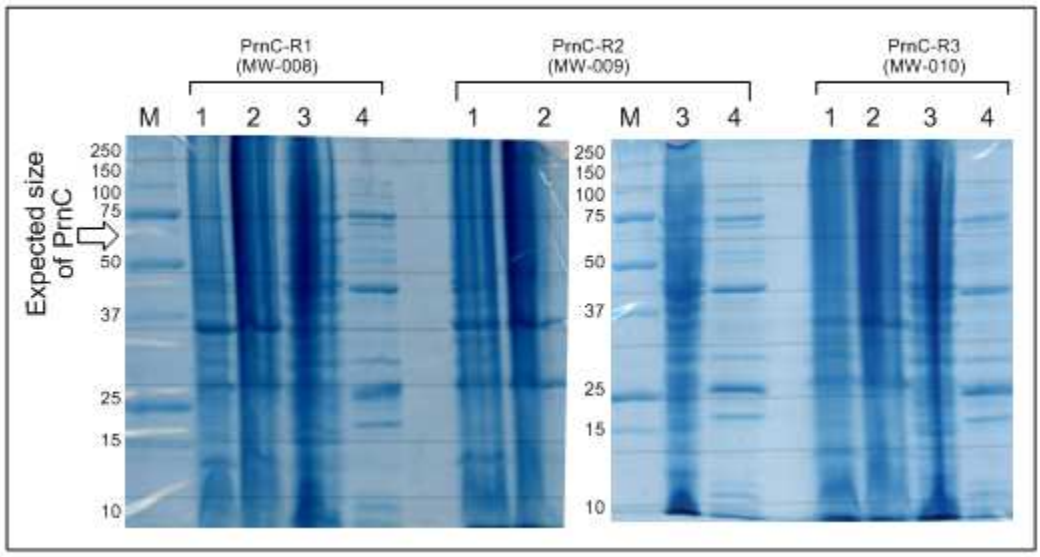


Figure 5.5: SDS-PAGE analysis of the 0.2% (w/v) L-arabinose induced protein expression of PrnC variants R1-3 inside *E. coli* BL21. No clear band corresponding to PrnC was seen in any of the constructs. Lanes are as follows: **(M)** Protein marker (Biorad). Sizes are in KDa. **(1)** Pre-induction sample. **(2)** Protein content of cell debris. **(3)** Protein content of cell free lysate. **(4)** Protein content following Ni-NTA purification. Expected sizes of PrnC: R1 – 64.8 kDa, R2 – 63 kDa and R3 – 61.6 kDa.

The resulting fractions, including a sample of the cell debris left in the pellet following sonication and centrifugation, were analysed by SDS-PAGE (see figure 5.5). The results showed no large, clear bands corresponding to the protein of interest. There did not even seem to be any protein of significance present in the insoluble fraction. A repeat showed the same result.

The insertion of *prnC* into pBAD and the resulting attempts to express the protein in *E. coli* BL21 was not successful. The pBAD manual suggests using a *rec*A, *end*A strain such as *E. coli* TOP10. These strains carry a mutation in the *ara*D gene which makes them capable of transporting arabinose but not metabolising it. This is suggested to be important to protein expression when using the araBAD promoter to ensure levels of the inductant remain constant inside the cell and do not decrease over time.[218]

E. coli DH10B has a genotype that is identical to TOP10 and thus was selected as an alternative host for *prnC*/pBAD expression. For this study just the plasmid containing the full *prnC* gene was (pMW03) transformed into chemically competent DH10B cells to form *E. coli* DH10B MW-021. To verify the insertion of the gene into the *E. coli* DH10B host, plasmid DNA was extracted from the resulting strain and double restriction digested to prove that the *prnC* gene was present. Two bands corresponding to DNA of the expected size were visualised on an ethidium bromide stained electrophoresis gel, strongly implicating the presence of the correct gene.

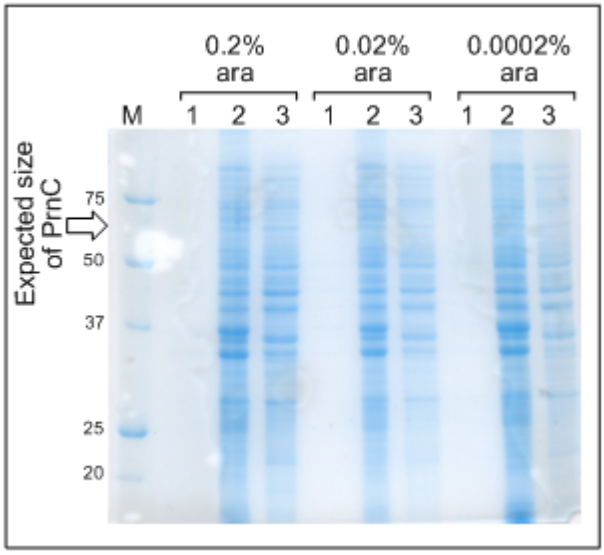


Figure 5.6: SDS-PAGE analysis of PrnC-R1 production in *E. coli* DH10B MW-021 with varying concentrations of L-arabinose. No clear evidence of PrnC production can be seen. Lanes are as follows: **(M)** Protein marker (Biorad). Sizes are in KDa. **(1)** Pre-induction sample. **(2)** Protein content of cell debris. **(3)** Protein content of cell free lysate. Expected sizes of PrnC: R1 – 64.8 kDa.

Three 10 ml scale cultures of MW-021 were grown at 37°C until an OD_{600} of 0.5 was reached. As before a range of L-arabinose concentrations were trialled to try to find the optimum level (0.2-0.0002%). Following induction the cells were cooled to 16°C for 24 hours. The cells were lysed and the lysate was analysed together with samples of the culture before induction and samples of the cell pellet on SDS-PAGE (figure 5.6). However, despite being transferred into a more suitable host, there was still no evidence of over production of PrnC using the pBAD vector. Transferring the gene into a different vector may help to resolve these issues.

5.2.3 Expressing PrnC gene inside pET28a(+)

5.2.3.1 Cloning methods

In my previous work with the tryptophan halogenases, transferring the genes into pET28a(+) resulted in improved levels of soluble protein. Therefore primers were designed to enable the insertion of the *prnC* gene into this vector. The multiple cloning site inside pET28a(+) contains an EcoRI site (see figure 5.7) which that meant that the same reverse primer could be used for cloning as was used for the pBAD construct (PrnC-R1). There is no BglII site contained within the multiple cloning site however so another primer was used that contained an NdeI restriction site (chosen as this was upstream of the EcoRI site in the MCS). This primer was obtained from a co-worker working on a parallel project and was designated PrnC-1.

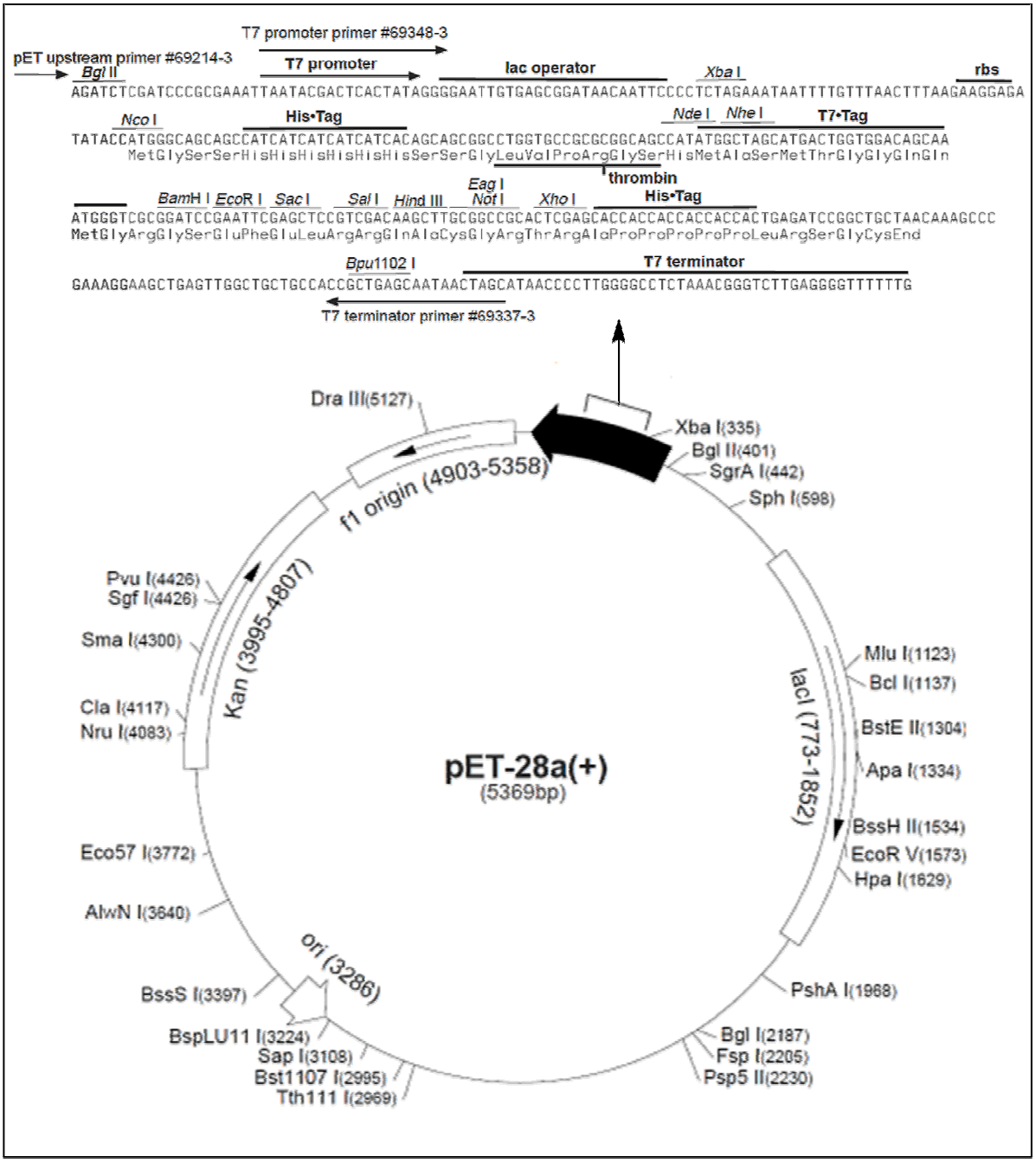


Figure 5.7: Key features of the pET28a(+) expression vector and annotation of the multiple cloning site. pET28a(+) is a cloning and expression vector from Novagen. The multiple cloning site features an N-terminal His6 tag and an optional C-terminal His6 tag (provided outside of the reading frame). The MCS also contains a number of unique restriction sites to enable insertion of recombinant genes. Control of protein production is provided by the T7lac promoter.

As with the pBAD construct, a trial PCR was first performed to determine the optimum levels of gene amplification using this new combination of primers. The trial demonstrated that the optimum conditions to produce the maximum amount of *prnC* with the highest specificity was a 60°C annealing temperature with the addition of 5% DMSO (see figure 5.8).

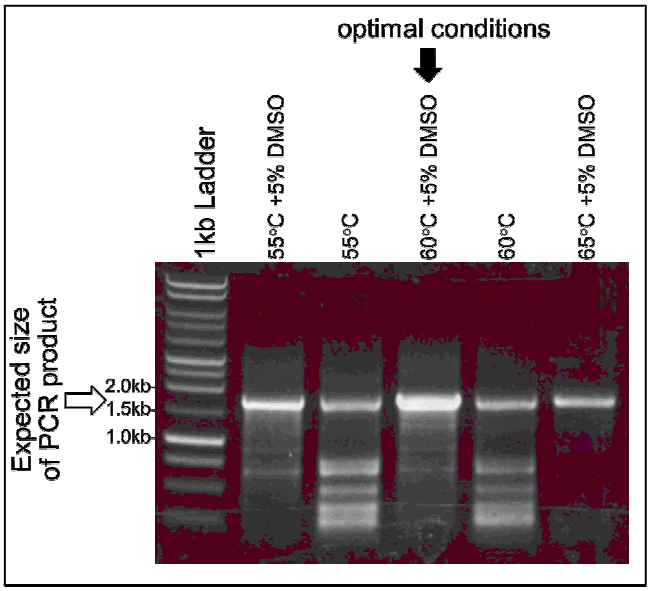


Figure 5.8: DNA gel electrophoresis analysis of *prnC-P1* gene PCR amplification under a variety of different conditions. Analysis of PCR reactions carried out with the PrnC-1 and PrnC-R1 primers at different annealing temperatures and the effect of adding 5% final concentration of DMSO to the reactions. Expected PCR products are PrnC-P1 (1704 bp). The expected size is marked with a white arrow. 1kb ladder (promega) sizes from top: 10000, 8000, 6000, 5000, 4000, 3000, 2500, 2000, 1500, 1000, 750, 500, 250bp. Selected optimal PCR conditions are highlighted with a black arrow.

These conditions were used for a 100µl scale up and the band corresponding to the correct size was excised and gel purified. This new PCR product was designated *prnC-P1*, in order to differentiate it from the previous *prnC* PCR product. *prnC-P1* was treated with the restriction enzymes NdeI and MfeI and was subsequently ligated into pET28a(+)that had been digested with NdeI and EcoRI. This formed the plasmid pMW06. For protein expression the plasmid was transformed into *E. coli* BL21 to form strain MW-012. Confirmation of the inserted gene was obtained by isolating the plasmid from the strain and double digesting it with NdeI and PvuII. The second restriction recognition site used to insert the gene (MfeI) was destroyed during ligation into the EcoRI site of the plasmid, therefore PvuII was used as an alternative. This enzyme does not cut pET28a(+) but does cut the *prnC* gene 1181 base pairs from the start. Therefore if *prnC* is present inside the vector a double digest should generate a fragment approximately 1181 base pairs in size and a remaining vector fragment of 5892 base pairs. These expected fragments were present following the restriction digest, strongly confirming the presence of the insert inside pMW06. The plasmid DNA was also sequenced to confirm the presence of the insert, ensuring that the gene was in the correct frame for translation and that no errors had been introduced during amplification.

5.2.3.2 Protein expression, purification and initial activity assays with *E. coli* BL21 MW-012

5.2.3.2.1 Protein expression and purification of PrnC and PrnF

A 500 ml main culture of *E. coli* BL21 MW-012 was grown at 37°C until an OD_{600} of 0.6 was reached. Protein expression was then induced with the addition of 0.1 mM final concentration of IPTG, the culture was then grown at 16°C for an additional 24 hours. The cell free extract was generated and subjected to Ni-NTA purification to check for the production of the his-tagged halogenase enzyme. The samples were analysed by SDS-PAGE (see figure 5.9). Following the purification steps, a very faint band of the appropriate size for PrnC-P1 could be seen in the elution fraction (figure 5.9, A7). There seemed to be significantly more protein in the cell lysate than was present in the cell debris pellet (figure 5.9, A2 and 3 respectively), which was significantly better than the expression levels shown with the pBAD vector. It can be seen from figure 5.9 (A4) however that despite fair levels of protein present in the cell lysate, the protein of the required molecular weight was being washed

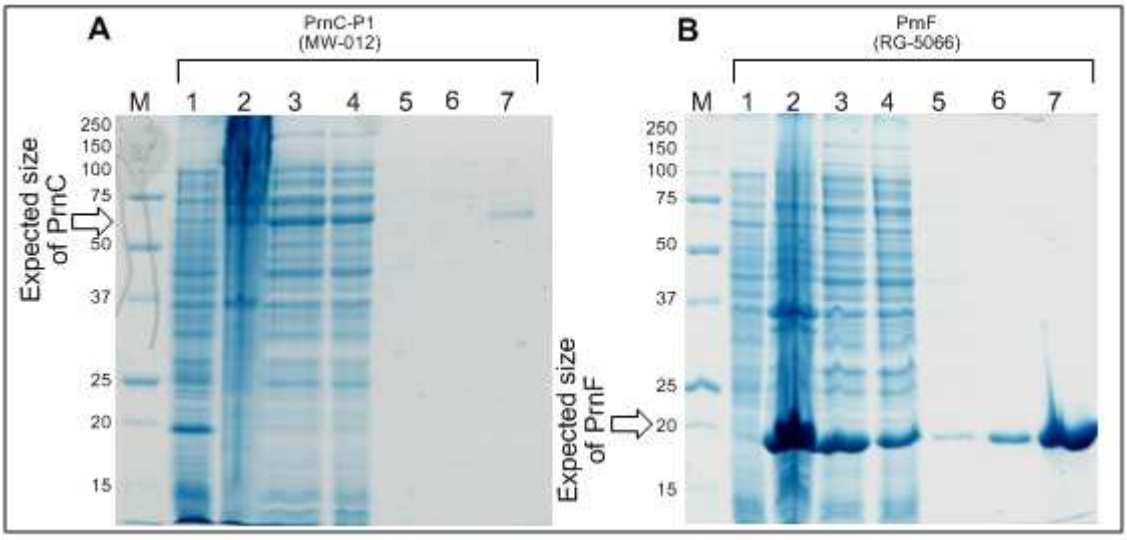


Figure 5.9: SDS-PAGE analysis of PrnC-P1 production in *E. coli* BL21 MW-012 and of PrnF production in *E. coli* BL21 RG-5066. (A) Ni-NTA purification of PrnC from culture of MW-012 as followed by SDS-PAGE. **(B)** Ni-NTA purification of flavin reductase PrnF from culture of RG-5066 as followed by SDS-PAGE analysis. Lanes are as follows: **(M)** Protein marker (Biorad). Sizes are in KDa. **(1)** Pre-induction sample. **(2)** Protein content of cell debris. **(3)** Protein content of cell free lysate. **(4)** Proteins not bound to Ni-NTA resin. **(5)** Proteins washed off resin with Ni-NTA lysis buffer. **(6)** Proteins washed off resin with Ni-NTA wash buffer. **(7)** Elution from Ni-NTA resin following wash with Ni-NTA elution buffer. Expected sizes: PrnC – 64.8 kDa, PrnF – 21 kDa (indicated by white arrows).

straight off the Ni-NTA resin and that the observed levels of resin binding were very low. This perhaps suggests that the poly histidine tag attached to the protein is poorly available and may be internalised.

The elution fraction from the Ni-NTA column was concentrated tenfold. The protein was then buffer exchanged from the imidazole containing elution buffer into a HEPES protein storage buffer containing glycerol which enabled long term storage by flash freezing with liquid nitrogen. The amount of protein present in the eluent was assessed by means of a Bradford type assay which was prepared by first constructing a standard curve of concentration versus absorbance based on samples of known concentration of bovine serum albumin (BSA) in the range of 0.2-1.0 mg ml^{-1}.

A sample of the elutant containing purified PrnC-P1 was diluted two times to bring the colour change of the Bradford reagent into the linear range of the BSA standards. Recording the absorbance of the standards allowed a standard curve to be drawn; comparison of the PrnC-P1 absorbance to this standard curve allowed the estimation of the protein concentration of the eluant to be 1.2 mg ml^{-1} which corresponds to a protein concentration of 17.2 µM. This concentration is very low but was considered sufficient to attempt some initial activity assays on the enzyme.

As a flavin-dependent halogenase, PrnC requires a second flavin-reductase enzyme in order to function. Therefore a previously prepared strain (*E. coli* BL21 RG-5066) which had been transformed with a vector (the pET21a(+) based pSG24) containing the gene for PrnF was used to inoculate 500 ml of LB and growth, protein expression and Ni-NTA purification were performed under the same conditions used for MW-012. The purification of PrnF can be seen in figure 5.9, B. Unlike PrnC, the flavin reductase was expressed at very high levels and was largely soluble. There was no problem with binding to the Ni-NTA resin and a large amount of protein was eluted as the final step of the purification (figure 5.9, B7). Following identical volume reduction and buffer exchange steps, the protein concentration of PrnF was also assessed using the Bradford type assay. In this case the protein sample had to be diluted twenty times to bring the colour change into the range of the BSA standards. The absorbance measurements at 595 nm allowed the concentration of PrnF to be estimated as 14.7 mg ml^{-1}, which meant an overall concentration of 700 µM (significantly higher than PrnC).

5.2.3.2.2 Initial assay of PrnC activity

At this time the natural substrate for PrnC (monodechloroaminopyrrolnitrin) was not available so a selection of other available pyrrole derivatives were trialled as possible substrates for the enzyme. The compounds used were a selection of pyrroles that were present in the laboratory. The candidate

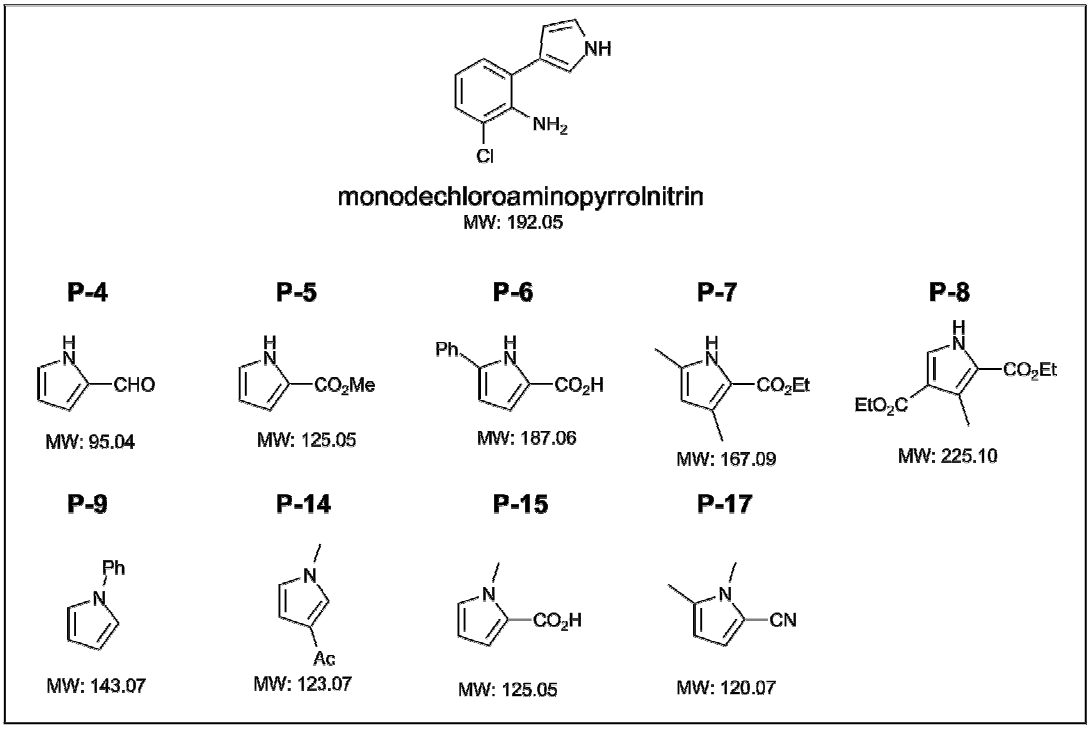


Figure 5.10: Structure of the natural substrate for PrnC and the structures of the pyrrole derivatives that were tested for activity with PrnC. The natural substrate for PrnC is the pyrrole moiety in monodechloroaminopyrrolnitrin. The structures below this represent the nine different pyrrole derivatives that were fed to PrnC to test for activity. The molecular weights are included.

pyrroles were examined for how easy they were to detect with the HPLC and LCMS as this would be the method used to determine product formation. Nine pyrrole related compounds were found to be amenable for this process as they produced clear peaks on the HPLC and had masses that were detectable by the LCMS; the structures are shown in figure 5.10 together with the structure of the natural substrate for the enzyme.

Preliminary PrnC assays were constructed using the purified PrnC and PrnF enzymes produced above. As the concentration of the flavin reductase was very high it was diluted twenty times before use. The Preliminary assays contained 25 µl of the purified PrnC enzyme together with 3 µl of the diluted, purified flavin reductase PrnF. To these, the essential cofactors NADH and FAD were added to final concentrations of 1 mM and 10 µM respectively. The pyrrole derivatives were tested individually and added at a final concentration of 1 mM. The assays were performed in 15 mM Tris-HCl buffer at pH 7.5 with the addition of 5% DMSO to improve the solubility of the substrates. An assay was set up for each of the nine pyrroles selected as potential substrates (compounds shown in figure 5.10) with appropriate negative controls and were performed at 30°C for 2 hours. Following

this period of incubation all the reactions were quenched by the addition of an equal volume of formic acid (10%) and centrifuged to remove the precipitated protein.

The samples were analysed by HPLC to see if any formed product could be detected. No additional peaks appeared in the chromatogram, indicating that no halogenated pyrrole was being produced. Analysis with LCMS also failed to detect evidence of chlorinated product (Appendix 2 contains the HPLC traces from this experiment).

The negative results in the assays can be explained by several factors. The most obvious is that the natural substrate was not tested and that the pyrrole derivatives available are not able to function as a substrate for PrnC. This would be consistent with the general report of high substrate specificity seen with the members of the $FADH_2$-dependent family of halogenase enzymes.[202] Alternatively the concentration of the enzyme may have been too low for effective assaying and was certainly too low to enable the generation of a crystal structure. The overall levels of production of soluble protein was low but acceptable, the main issue arose from the lack of binding to the purification resin. If this binding could be improved then higher concentrations of pure protein may be obtainable.

5.2.3.3 Steps to improve binding of PrnC to Ni-NTA resin

5.2.3.3.1 Altering quantity of Ni-NTA resin and adjusting buffer composition

The initial approach taken to try to improve the binding efficiency of PrnC-P1 to the Ni-NTA purification resin was to simply adjust the composition of the Ni-NTA buffers and to vary the amount of resin added to the cell lysate. The ionic strength of the buffer was lowered by reducing the NaCl content from 300 mM to 150 mM. Normally higher ionic strengths are required to reduce the amount of non-specific ionic bonding to the resin by other proteins[136] but in this case it was lowered to see if it would have an effect on the binding of PrnC. The amount of Ni-NTA resin used was also increased fourfold. The two different buffer systems with different NaCl concentrations were tested with both quantities of nickel resin. Neither adjustment showed any effect on the binding of PrnC to the resin. It is likely that all or part of the polyhistidine tag is buried inside the protein or otherwise unavailable for binding, therefore more significant changes to the protein were trialled.

5.2.3.3.2 Carrying out the Ni-NTA binding step in the absence of imidazole

The Qiagen Ni-NTA purification handbook suggests that in the case of poor target protein binding, imidazole concentration in the binding step can be reduced.[136] The Ni-NTA lysis buffer used

previously contained 10 mM imidazole, this is already at the lower end of the recommended level but if PrnC-P1 is binding at really low affinity then removing it altogether may have some impact on binding.

A 500 ml culture of *E. coli* BL21 MW-012 was grown and protein production induced in the usual way. The culture was then split in half and the cells collected. One half was re-suspended and sonicated in standard Ni-NTA lysis buffer, the other was re-suspended and sonicated in buffer containing no imidazole. The lysate of each culture was cleared and added to 0.5 ml of Ni-NTA resin that had been pre-equilibrated into the appropriate buffer (with or without imidazole). After incubation the resins were applied to columns and the flow-through collected. Both sets of resin were then washed with aliquots of the appropriate lysis buffer before being treated according to the standard Ni-NTA purification protocol with the usual wash and elution buffers containing imidazole. The purifications were followed by SDS-PAGE analysis and can be seen in figure 5.11.

The standard Ni-NTA purification showed a similar result as before; the majority of the protein was either in the insoluble form or was not adhering to the nickel resin. Only a very faint band of the appropriate size could be seen in the eluant (figure 5.11, lane A8). Disappointingly, removing the

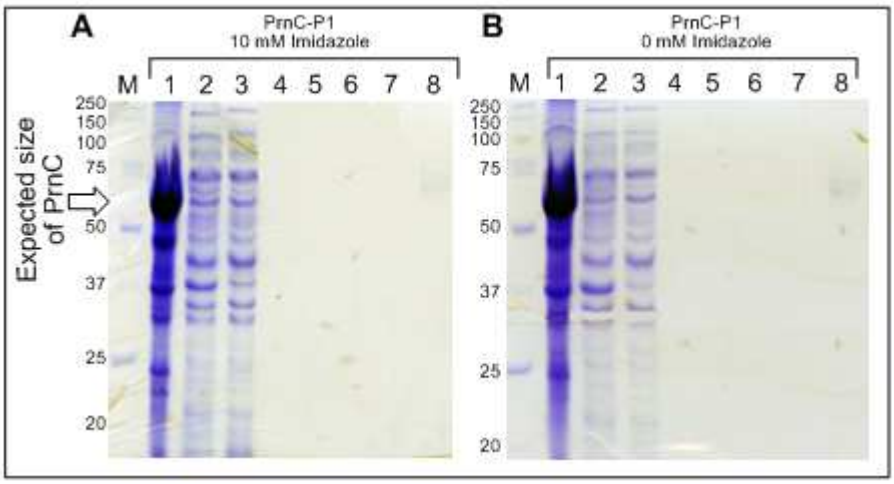


Figure 5.11: SDS-PAGE analysis showing the effect of removing imidazole from the binding step of the Ni-NTA purification of PrnC-P1. Lanes are as follows: **(M)** Protein marker (Biorad). Sizes are in KDa. **(1)** Protein content of cell debris. **(2)** Protein content of cell free lysate. **(3)** Proteins showing no adherence to Ni-NTA resin. **(4-5)** Proteins removed from resin following Ni-NTA lysis buffer wash. **(6-7)** Proteins removed from resin following Ni-NTA wash buffer wash. **(8)** Proteins eluted off Ni-NTA resin with elution buffer. Expected size of PrnC – 64.8 kDa. **(A)** Standard purification of PrnC-P1 in the presence of 10 mM imidazole showed very minimal levels of purified PrnC, the band at the

correct size can barely be seen following coomassie straining. **(B)** Removal of imidazole during the resin binding step demonstrated no significant improvement.

Imidazole from the initial binding step did not seem to improve matters. As can be seen in figure 5.11B, the majority of the protein was still being produced in an insoluble form. Although perhaps the elution buffer contains slightly more protein than the adjacent purification containing imidazole; this is certainly not conclusive. Due to this lack of improvement imidazole was returned to the binding buffer for all subsequent PrnC purifications.

5.2.3.3.3 Quikchange mutagenesis on pMW06 to increase the length of His6 tag

In order to improve the binding efficiency of PrnC-P1 to the Ni-NTA resin, site directed mutagenesis was performed on pMW06 to insert two additional histidine residues into the affinity tag to make a His8 tag. This was designed to increase the overall length of the tag and perhaps help to compensate if some of it is internalised by the protein. For analysis this had the effect of increasing the protein weight slightly to 67.24 kDa. The site directed mutagenesis was based on the Quikchange protocol from Stratagene.[219]

Mutagenic primers were designed to amplify the entire plasmid but with the addition of two extra histidine residues. The primer design guidelines from the protocol were followed to produce two primers that contained the desired mutation in the middle, with 10-15 base pairs of correct sequence on both sides, and that annealed to the same sequence on opposite DNA strands of the plasmid. Primers were required to be between 25 and 45 base pairs long with a melting temperature of less than or equal to 78°C (the Tm was calculated using a formula provided with the protocol – Tm = 81.5 + 0.41 (%GC) – 675/N, where N is equal to the primer length not including the inserted bases). The two primers designed were designated pET28a-8xHis-F1 and pET28a-8xHis-R1.

A purified sample of pMW06 was used as the template for a PCR reaction using the two mutagenic primers. The PCR reaction was run using the conditions outlined in the Strategene protocol which suggested 18 cycles of reaction with 55°C annealing and a 68°C extension temperature.[219] The extension time was set at 15 minutes with a final extension of 30 minutes. A range of template concentrations were initially chosen: 10 – 500 ng. A series of no primer controls were also set up in parallel.

Following the PCR reaction the enzyme DpnI was added to each tube, this enzyme digests the methylated template DNA and increases the chance of transformation occurring with a mutated

vector. The no primer control reactions were used to judge the efficiency of this DpnI digest. 5 µl of each of the PCR products were transformed into aliquots of *E. coli* DH10B chemically competent cells using the standard transformation protocols and plated onto LB-agar plates. Following incubation at 37°C overnight all the plates contained large numbers of colonies, including the control plates. Carrying out a second DpnI digest on the PCR product prior to a repeat transformation into *E. coli* DH10B showed no improvement.

The PCR reaction was repeated with just two concentrations of plasmid DNA: 50 and 500 ng. This time the DpnI digest was carried out using fresh enzyme. Following transformation the control plate for the 500 ng reaction contained approximately 200 colonies, the actual plate contained at least three times as many. The 50 ng plates showed approximately equal numbers of colonies on both the reaction and control. A second round of DpnI digestion on the PCR product prior to a repeat transformation reduced the number of colonies on the control plate significantly.

Four colonies were picked from the 500 ng plate and used to inoculate 10 ml of LB; the plasmids were purified from the resulting cultures and sent for forward sequencing to identify whether any of them carried the mutated gene. Of the four colonies two contained the desired mutation. Full sequencing of the first colony revealed no other mutations in the gene so was carried forward for protein expression studies with the designation pMW12. For protein expression the plasmid was transformed into *E. coli* BL21 and given the strain designation MW-016.

500 ml cultures of this strain were grown up alongside cultures of the original MW-012 strain containing the unaltered pMW06 plasmid. Protein expression was induced with IPTG and growth switched to 16°C for 24 hours. The cell lysates were then incubated with Ni-NTA resin for purification of the PrnC. The side by side comparison of the two plasmids can be seen in figure 5.12.

The purification of the normal PrnC construct (pMW06, MW-012) showed consistent results with previous experiments; the majority of the PrnC can be seen in the insoluble fraction with most of the PrnC in the lysate and not being retained by the nickel resin. The mutated construct with the longer histidine tag shows similar results. The elution lane however (figure 5.12B lane 4) does show a slightly higher amount of purified enzyme, indicating that the longer histidine tag may be improving the binding.

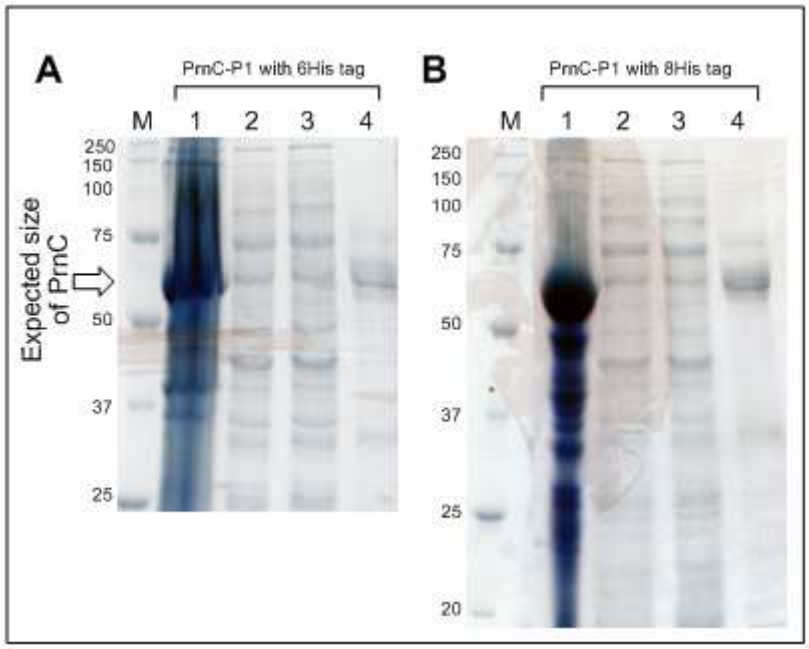


Figure 5.12: Side by side comparison of MW-012 and MW-016 production and Ni-NTA purification of PrnC. Lanes are as follows: **(M)** Protein marker (Biorad). Sizes are in KDa. **(1)** Protein content of cell debris. **(2)** Protein content of cell free lysate. **(3)** Proteins showing no adherence to Ni-NTA resin. **(4)** Proteins eluted off Ni-NTA resin with elution buffer. Expected size of PrnC – 64.8 kDa. **(A)** Purification of *E. coli* MW-012 containing PrnC-P1 with a His6 tag. Consistent with previous work the purification of PrnC-P1 showed mostly insoluble protein with a small amount present in the purified elution fraction. **(B)** Purification of *E. coli* MW-016 containing mutated plasmid pMW12; PrnC with a His8 tag. Purification of the mutated PrnC-P1 with the longer histidine tag also showed the majority of protein in the cell pellet. Compared to the unmutated PrnC, however, there does seem to be a larger amount of protein present in the elution.

5.2.3.4 Other methods to improve protein solubility

5.2.3.4.1 Auto-induction of *E. coli* BL21 MW-012

Protein production under the T7 promoter is switched on in the presence of allolactose (or the non-hydrolysable mimic IPTG) but is inhibited in the presence of glucose. In standard protein induction cells are grown to mid log phase and then induced with IPTG. The regulation of the T7lac operon by glucose can be used to slowly induce protein expression without the sudden on switch that is used conventionally. The underlying concept is that turning on protein production gradually leads to less protein aggregation and the formation of fewer inclusion bodies,[206] therefore producing more soluble protein. This is achieved by growing the culture in a medium that contains both glucose and lactose. At the start of the culture glucose is in abundance so protein expression is repressed but as

it is metabolised by the cells the repression begins to decrease and gene expression begins. As more and more of the glucose is used up the activation of the T7lac promoter increases, resulting in a very smooth transition from 'off' to 'on'.

A culture of *E. coli* BL21 MW-012 (containing the pMW06 PrnC-P1 plasmid) was initially grown up in 2 ml of glucose rich medium (ZYP-0.8G supplemented with 50 µg ml^{-1} of kanamycin) until the culture reached an OD_{600} of approximately 6. The presence of high levels of glucose (0.8%) should inhibit PrnC protein production at this point. This culture was used as a starter culture and 400 ml of ZYP-5052 media was inoculated with 200 µl of this starter. This main media contained glucose (0.07%), glycerol (0.7%) and lactose (0.27%) as potential carbon sources. The main culture was incubated for 15 hours at 24°C during which time the glucose and glycerol would be consumed preferentially over the lactose and production of PrnC would slowly be induced.

The cell free lysate was generated and subjected to Ni-NTA purification as described previously. The purification process was followed by SDS-PAGE (can be seen in figure 5.13).

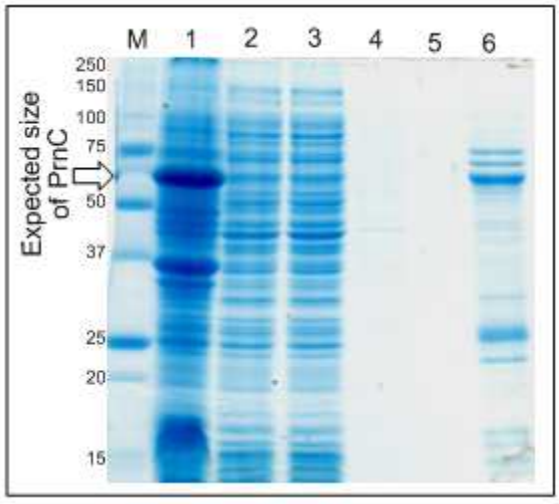


Figure 5.13: SDS-PAGE analysis of PrnC-P1 production and Ni-NTA purification in *E. coli* BL21 MW-012 under auto-induction conditions. Lanes are as follows: **(M)** Protein marker (Biorad). Sizes are in kDa. **(1)** Protein content of cell debris. **(2)** Protein content of cell free lysate. **(3)** Proteins showing no adherence to Ni-NTA resin. **(4)** Proteins removed from resin following Ni-NTA lysis buffer wash. **(5)** Proteins removed from resin following Ni-NTA wash buffer wash. **(6)** Proteins eluted off Ni-NTA resin with elution buffer. Expected size of PrnC – 64.8 kDa. The cell debris in lane 1 shows significant levels of insoluble protein compared to the cell lysate, lane 2. The elutant in lane 6 shows many contaminating proteins but a more intense band at the appropriate size range suggests PrnC is being purified.

Despite the theoretical advantages of using auto-induction to produce larger amounts of soluble protein, in this case the opposite seems to have occurred. Although the overall production of PrnC seemed to be good it was almost exclusively present in the insoluble cell debris fraction (figure 5.13 lane 1). Some of the protein was purified using the Ni-NTA resin but the levels are low enough that the required protein seems to make up only about half of the protein present in the elutant, with the rest composed of other contaminant proteins.

Auto-induction did not seem to offer any particular improvements over the standard method of inducing protein expression used earlier.

5.2.3.4.2 Co-expression of PrnC with chaperone proteins

The mutated plasmid pMW12 containing the lengthened histidine tag had demonstrated improved binding properties with the Ni-NTA resin compared to the non-mutated pMW06. Despite the improved performance the majority of the protein was still insoluble and localised inside the cell pellet after sonication. Numerous methods to improve this had been largely unsuccessful so a new approach was trialled; the inclusion of chaperone proteins.

As a protein begins to fold a number of partially folded intermediate states are formed along the way. These intermediates have a tendency to aggregate with each other which can lead to the formation of inclusion bodies; this can be exacerbated when the protein is being overexpressed (see Fink 1999 for review)[206]. The role of molecular chaperones is to prevent this protein aggregation. There are several families of molecular chaperones with many also being known as heat shock proteins because protein aggregation has a higher propensity when protein is produced at higher temperatures due to the increased rate of protein production. One family of chaperone proteins are known as HSP70 chaperones and they function by sequestering partially folded proteins and thereby preventing aggregation with other non-native proteins but they do not actively become involved in the folding process. An example of this kind of chaperone is DnaK from *E. coli*. Another family are the HSP60 chaperones which contain a large central cavity where partially or unfolded proteins can be accommodated, isolating the protein to enable proper folding. The *E. coli* chaperone GroEL is an example of this family. Members of both of these families require cochaperones (also known as chaperonins) for activity; GroEL requires a protein known as GroES, while DnaK has two co-proteins known as DnaJ and grpE.[206] Protein aggregation may be a reason why PrnC is predominantly insoluble even at low temperatures; overexpression of molecular chaperones may assist in the proper folding of the enzyme.

Three plasmids were purchased from Takara Bio Inc[220] that contained different molecular chaperones. Plasmid pGro7, which contains the groEL-groES chaperone system;[221] pKJE7 which contains the dnaK-dnaJ-grpE chaperone system[221] and the third plasmid; pTf16 which contains the gene for trigger factor (tig), a comparatively more recently identified chaperone that has been shown to improve protein folding in *E. coli*.[222] All three plasmids contain chloramphenicol resistance and the araB promoter which means that chaperone over production is induced by increasing L-arabinose concentration. The three plasmids were all transformed into *E. coli* BL21 cells (RG-5310, RG-5311 and RG-5313 respectively), cultures of these strains were themselves made chemically competent and each transformed with the *prnC*-containing pMW12 vector to form strains *E. coli* BL21 MW-022, MW-023 and MW-024. Colonies containing both plasmids were selected with kanamycin (for pMW12 selection) and chloramphenicol resistance. Colony PCR was used to check that the colonies contained the *prnC* gene.

Starter cultures of all three strains were used to inoculate 10 ml of LB media supplemented with both antibiotics and with arabinose to induce the expression of the chaperone protein at 37°C. For each chaperone three different arabinose concentrations were trialled; 0.05, 0.2 and 0.4%. IPTG was added (0.1 mM) to induce the production of PrnC once the OD_{600} of the cultures reached 0.5. Incubation temperature was lowered to 16°C and growth continued for 24 hours. The cell free extracts of the cultures were generated and visualised with SDS-PAGE (figure 5.14).

Analysis of the groEL/groES strain (figure 5.14A) showed a thick band, at around 60 kDa, appearing in the cell pellet and the cell lysate. This band increased in size as the arabinose concentration increased. A similar band was also present in the pre-induction sample. This all suggests that the band in question is GroEL which has a mass of 60 kDa. This presents a problem for identifying any improvement with PrnC, with a mass of 67.24 kDa the two proteins would be hard to separate with SDS-PAGE. The chaperonin GroES has a low molecular weight (10 kDa) and would not be seen on this gel. Ni-NTA purification may help to resolve this issue.

The dnaK system, the second chaperone trialled, has a similar problem as the chaperone is 70 kDa in size. The two chaperonins (dnaJ and grpE) are 40 kDa and 22 kDa so do not obscure the PrnC signal. The SDS-PAGE gel of this co-expression (figure 5.14B) showed two prominent protein bands between the 50 and the 75 kDa marker proteins of the ladder. These two bands could be PrnC and the dnaK chaperone. As PrnC has a lower molecular weight it should be represented by the lower band which is more prominent in the cell lysate than the top band. This would seem to suggest that the chaperone is having the desired effect. However the lower band is present in the pre-induction sample (figure 5.14B lanes 1) but this may be to do with the T7 promoter 'leaking' expression in the

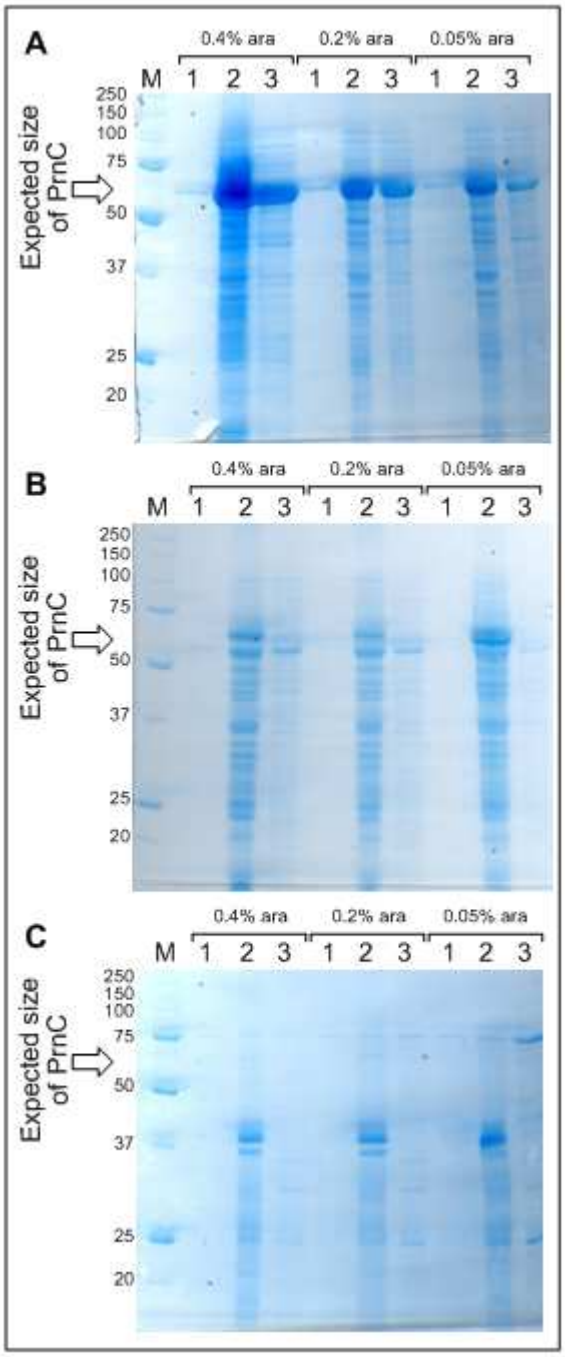


Figure 5.14: SDS-PAGE analysis of PrnC and chaperone protein co-production. The levels of PrnC and chaperone expression and degree of solubility of PrnC are shown. Chaperones were induced at a range of L-arabinose concentrations; 0.05, 0.2 and 0.4%. Expected size of PrnC – 67.24 kDa. Lanes are as follows: **(M)** Protein marker (Biorad). Size in kDa. **(1)** Pre-IPTG Induction sample. **(2)** Proteins present in insoluble pellet fraction. **(3)** Proteins present in soluble cell lysate. **(A)** Co-expression of PrnC with the pGro7 plasmid carrying the groEL/groES chaperone system. Expected size of groEL – 60 kDa, groES – 10 kDa. **(B)** Co-expression of PrnC with pKJE7 plasmid carrying the dnaK-dnaJ-grpE chaperones. Expected size of dnaK – 70 kDa, dnaJ – 40 kDa, grpE – 22 kDa. **(C)** Co-expression of PrnC with thepTF16 vector carrying the trigger factor chaperone. Expected size of tig – 56 kDa.

absence of the inducer. This is reinforced by there being no evidence of any of the two chaperonins being present in the pre-induction sample. Interestingly the top band shows higher levels of expression with the lower concentration of arabinose, so this value might represent the optimum expression level of these genes. As before the overall results are inconclusive as to whether the chaperone is enhancing PrnC production, purification of the affinity tagged PrnC may help to illuminate matters.

The final chaperone tested with PrnC was the trigger factor (tig). This chaperone has a molecular weight of 56 kDa so should not run at the same position as PrnC. This chaperone could not be seen to be overexpressed at the correct size. A prominent band was visible in the region of 40 kDa, but this is too low to be the chaperone. Another band can be visualised close to the 75 kDa marker. This band is faint in all of the cell lysate fractions and can be seen in the pre-induction as well. The size of this protein is too large to represent PrnC. No obvious evidence of PrnC overexpression could be seen.

The experiment was repeated using the same range of L-arabinose concentration to induce production of the chaperone but this time the resulting cell lysates were resolved using Ni-NTA small scale spin purification with the aim of pulling out the PrnC from the chaperones of the same size. All three chaperone containing PrnC strains were purified using this method. The insoluble, soluble and purified fractions were analysed on SDS-PAGE as before (figure 5.15).

Figure 5.15A shows the results of the purification of the pGro chaperone and PrnC. The cell pellet and the cell lysate showed the same large band in the 50-75 kDa range that had been seen in the earlier expression trials. Following the Ni-NTA purification this large band was isolated. The size of the band was bigger than anticipated and raises the possibility that the chaperone may be co-purifying with the halogenase enzyme. The two proteins are so close in size that it is not really possible to tell them apart using this method of analysis. The gel gives a slight impression that there may be two bands in the eluting fraction but this could also be a gel artefact and is not conclusive proof that the chaperone is being dragged through the purification.

The strain carrying the pKJE7 plasmid with the dnaK chaperone carried two bands in the cell lysate in the 50-75 kDa region. This is consistent with the previous results where the higher band was identified as belonging to the chaperone (70 kDa) while the lower band was consistent with the size of the mutated PrnC (67.24 kDa). However following the Ni-NTA purification, the top band was purified in preference over the lower one (figure 5.15B lane 3) for all the arabinose concentrations. This could indicate that the top band is PrnC and not the lower one which would be inconsistent

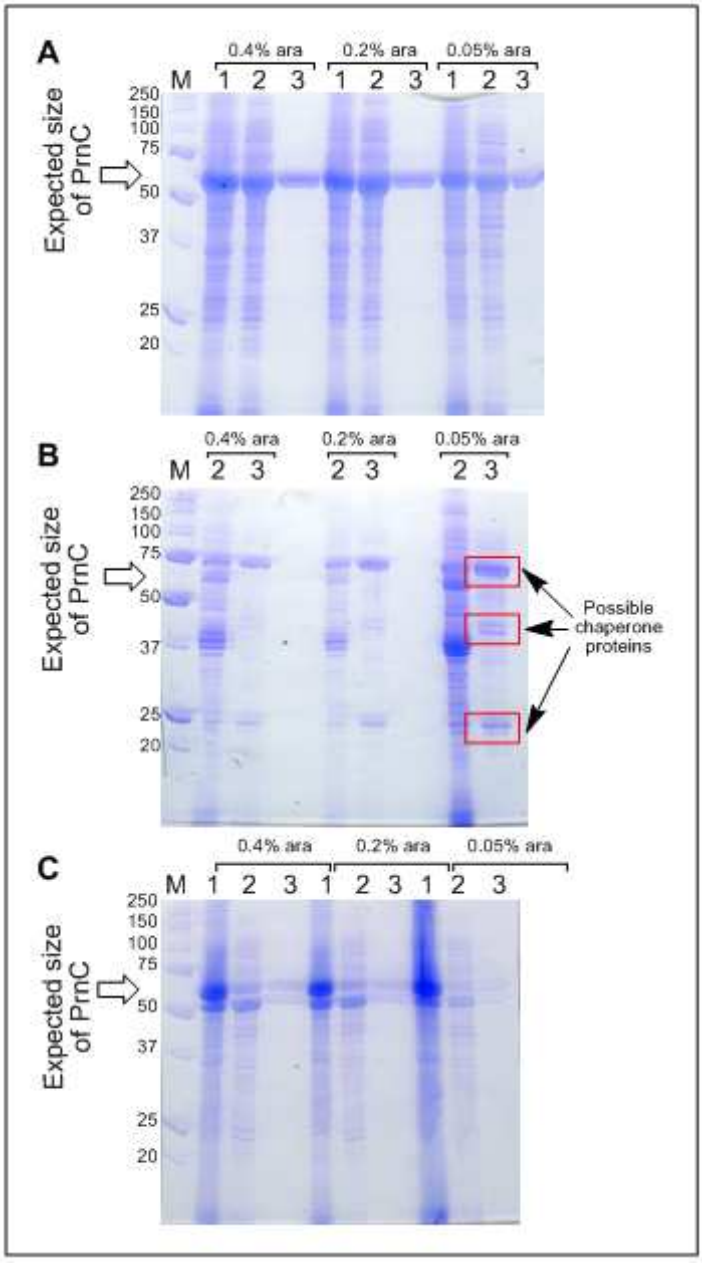


Figure 5.15: SDS-PAGE analysis of PrnC and chaperone protein co-expression and small scale Ni-NTA spin purification. Illustrates levels of PrnC and chaperone expression and degree of solubility of PrnC. Chaperones induced at a range of L-arabinose concentrations; 0.05, 0.2 and 0.4%. Expected size of PrnC – 67.24 kDa. Lanes are as follows: **(M)** Protein marker (Biorad). Size in kDa. **(1)** Proteins present in insoluble pellet fraction. **(2)** Proteins present in soluble cell lysate. **(3)** Proteins purified using Ni-NTA resin. **(A)** Co-expression of PrnC with the pGro7 plasmid carrying the groEL/groES chaperone system. Expected size of groEL – 60 kDa, groES – 10 kDa. **(B)** Co-expression of PrnC with pKJE7 plasmid carrying the dnaK-dnaJ-grpE chaperones. Expected size of dnaK – 70 kDa, dnaJ – 40 kDa, grpE – 22 kDa. Possible locations of chaperone and chaperonins in the Ni-NTA eluent can be seen in lane 3 marked with red squares. **(C)** Co-expression of PrnC with thepTF16 vector carrying the trigger factor chaperone. Expected size of tig – 56 kDa.

with the calculated molecular weight. The previous experiment had shown that the lower band was present before IPTG induction of the PrnC plasmid which again provides evidence that it is actually the upper band that contains PrnC. However there is also evidence that in the Ni-NTA purified layer the two chaperonins that work alongside dnaK are also present. Bands on the SDS-PAGE that are likely to represent these co-chaperones are highlighted with red boxes in one of the elution lanes of figure 5.15B. These two prominent bands are in the 37-50 kDa and the 20 kDa range respectively so it becomes highly likely that these two bands represent dnaJ (40 kDa) and grpE (22 kDa). If these two chaperonins are being purified then it is logical to assume that dnaK is also, once again shifting the identity of the purified protein back to the chaperone and not the halogenase enzyme. Either way the uncertainty that this provides makes this chaperone system less desirable as an expression strain.

The repeat of the pTf16 (trigger factor) containing strain (MW-024) also produced multiple bands in the purified fraction. Figure 5.15C shows two bands in the cell debris, the lysate and the Ni-NTA elution. The upper band is presumably the larger PrnC (67.24 kDa) while the lower band is assigned as the tig chaperone (56 kDa). Both bands could be seen purifying together in more or less equal quantities on the nickel resin. Overall levels of purification were also seen to be very low.

In all three cases it appears that some chaperone protein is co-purifying with the PrnC enzyme. In the case of trigger factor this is very clear. GroEL is too close in size to PrnC to be sure if both are being purified together and the same is true for PrnC and dnaK. The use of protein sequencing techniques or MALDI analysis could help to resolve this question in the future.

To test whether the chaperones themselves are capable of binding to the Ni-NTA resin in the absence of PrnC, a further two 10 ml cultures were grown, cell lysates generated and Ni-NTA purified in the same way as before. The cultures included the mutated PrnC on its own (pMW12, *E. coli* BL21 MW-016) and the groEL/groES chaperones on their own (pGro7, *E. coli* BL21 RG-5310).

Figure 5.16 shows that in the absence of chaperone the PrnC construct does not bind very well to the Ni-NTA resin and is mostly found in the insoluble cell pellet. Equally the chaperone, although it does express very well in the cell lysate, does not show much affinity to the resin in its isolated state. This very clearly demonstrates that it cannot be the chaperone alone that is generating the large purified band in figure 5.15. However the evidence still points to the fact that the chaperone is co-purifying and contaminating the pure protein, although the exact proportions are not known.

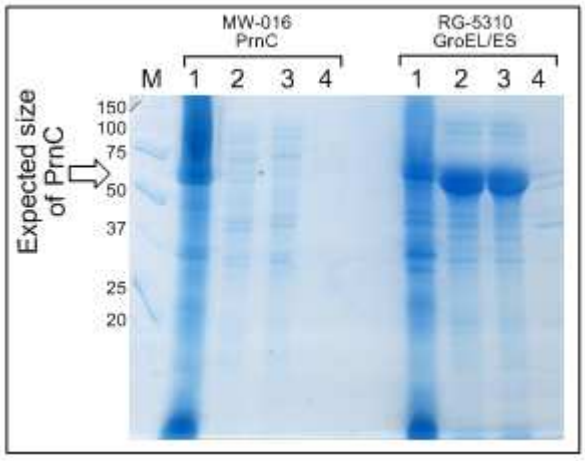


Figure 5.16: SDS-PAGE analysis of individual Ni-NTA purification of PrnC and GroEL/ES molecular chaperones. PrnC and GroEL/ES proteins were purified with Ni-NTA resin. Lanes are as follows: **(M)** Protein marker (Biorad). **(1)** Proteins present in insoluble pellet fraction. **(2)** Proteins present in soluble cell lysate. **(3)** Proteins showing no adherence to Ni-NTA resin. **(4)** Proteins purified using Ni-NTA resin. Both PrnC and the chaperone demonstrated poor binding to the resin.

5.2.3.4.3 Placing PrnC into the ligation independent vector pNYCOMPS-LIC-FH10T-ccdB

While attempting to improve the solubility of PrnC within pET28a(+), another vector became available for use in the laboratory. This was a pNYCOMPS-LIC-FH10T-ccdB vector (Arizona State University Biodesign Institute, see figure 5.17 for map). There are a number of features of this plasmid that differ from those found in pET28a(+). This vector contains a Flag epitope (conserved sequence DYKDDDDK) which is hydrophilic and so is likely to reside on the surface of the fusion protein and not be buried internally. The flag tag is also fused to a 10-histidine tag to enable nickel purification. The biggest difference between this vector and the ones tried previously is the lack of a multiple cloning site; insertion of a recombinant gene is instead achieved by ligation independent cloning methodology.[223]

The vector contains two SnaBI restriction sites (one at position 5314 and another at 6841) the process of inserting a new gene into the vector begins by digesting the vector at these points. SnaBI generates blunt ends after digestion. The lost section of the vector contains two selectable markers. The first is a gene that encodes for chloramphenicol resistance and the second is the lethal gene *ccdB* (a DNA gyrase inhibitor).[224] The idea behind this is that any competent *E. coli* cells taking up un-altered vector would die due to translation of the lethal gene. If used in *E. coli* strains that have a specific mutation in the gyrase enzyme, the chloramphenicol resistance can be used as a negative marker.

After digestion with SnaBI and subsequent agarose gel purification to isolate the vector from the removed section, the plasmid was treated with T4 DNA polymerase. This enzyme possesses 3' to 5' exonuclease activity. This is normally a DNA repair mechanism which involves the enzyme excising the DNA from the 3' end and replacing the removed DNA bases with fresh ones. However if the replacement bases are not available then a 5' overhang will be generated at the end of the DNA strand. The LIC vector is designed to contain only A, T and C bases on the vector side of the SnaBI restriction sites for the first 15 base pairs at which point a G base is positioned. If the T4 polymerase is only provided with dGTP then the first 15 base pairs will be excised until the enzyme comes across

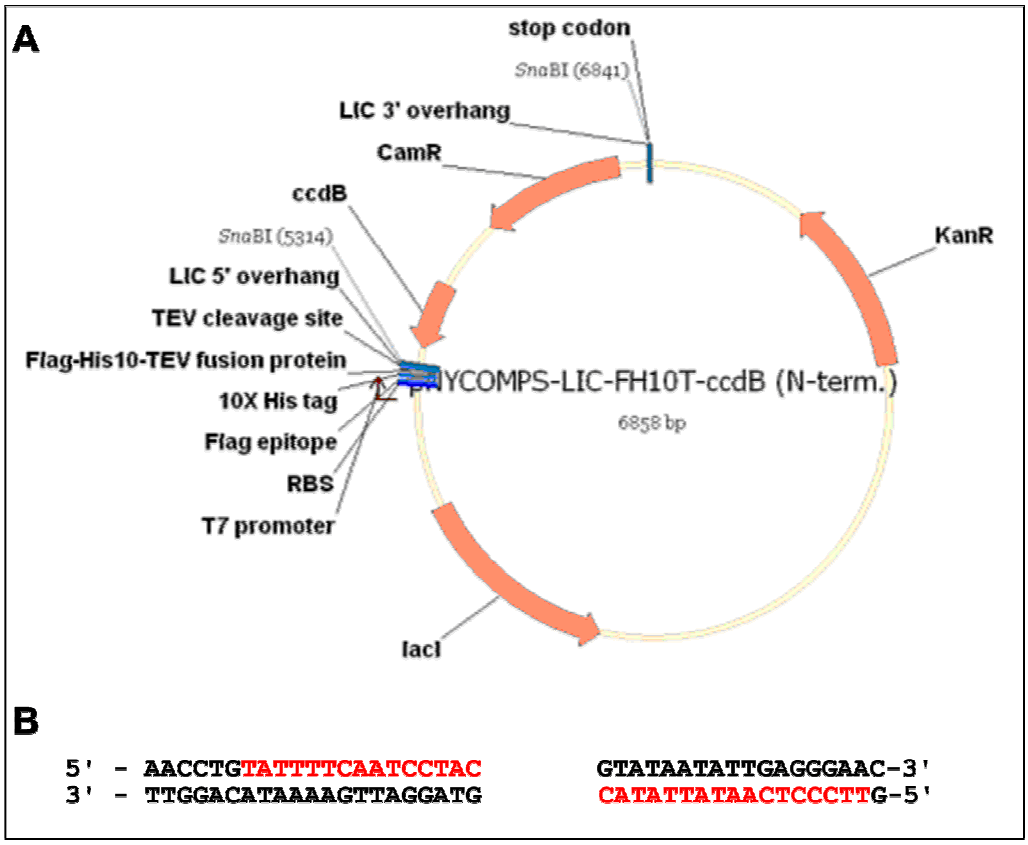


Figure 5.17: Vector map of pNYCOMPS. (A) The pNYCOMPS vector (Arizona State University Biodesign Institute) contains two SnaBI restriction sites (one at 5314 and the other at 6841) which allows the removal of this region, which contains a gene for chloramphenicol resistance and another carrying the *ccdB* lethal gene. Formation of sticky overhangs at either end of the created gap allows the insertion of a recombinant gene carrying complementary DNA overhangs. Protein production is controlled by the T7 promoter and selection of non-lethal recombinant plasmids is achieved with the inclusion of kanamycin resistance gene *KanR*. Recombinant proteins contain both a Flag epitope and a 10-histidine tag. **(B)** The ends of the vector after SnaBI digest are shown. The red bases are removed when DNA is treated with T4 DNA polymerase, which leads to the sticky overhangs shown in black.

the G base at which point it will be able to replace the excised G with a fresh one and dissociate. This occurs at both ends of the DNA and creates 5' overhangs as shown in figure 5.17. These created overhangs can be used to insert a recombinant gene that is engineered to contain complementary overhanging DNA.

Fresh PCR primers were designed for *prnC* to insert a specific DNA sequence at the start and the end of the PCR product. This insert was identical to the base pair sequence removed from the plasmid by the T4 DNA polymerase (figure 5.17, B). In this way an identical digest with the polymerase on the PCR product will result in sticky ends that are complementary to the plasmid. The difference is that the T4 DNA polymerase will be digesting the complementary strand so the reaction will need to have dCTP added instead of dGTP to ensure the dissociation of the enzyme. The forward and reverse primers were designated PrnC_LIC-F1 and PrnC_LIC-R1a. To ensure that the T4 DNA polymerase dissociates before the start of the *prnC* gene an extra codon, GGA (encoding for glycine), was added before the start of *prnC*. This starts with a G and therefore the complimentary strand will contain a C codon that will stall the polymerase. The same was performed with the reverse primer with the insertion of a GCC codon (alanine) which will make the polymerase dissociate in the presence of dCTP.

The previously established PCR conditions for *prnC-P1* were used with the new primers to amplify the *prnC* gene with the addition of the extra sequence at both ends. The resulting PCR product was identified as *prnC-LIC* and was produced with a high level of specificity (see figure 5.18). This product

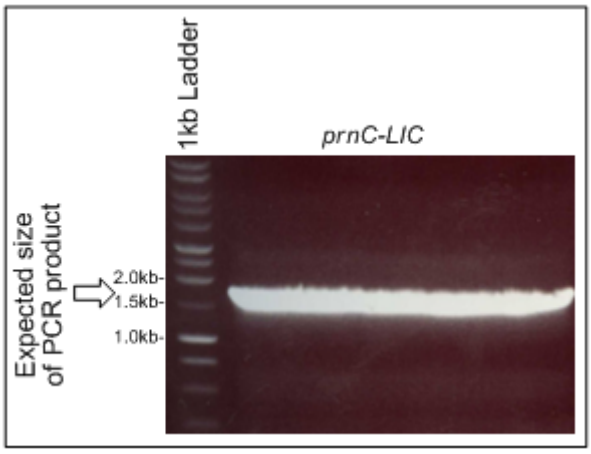


Figure 5.18: DNA gel electrophoresis purification of *prnC-LIC* gene amplification performed using the pre-established PrnC PCR conditions. Expected size of PCR product *prnC-LIC* is 1739 bp and is marked with a white arrow. 1kb ladder (promega) sizes from top: 10000, 8000, 6000, 5000, 4000, 3000, 2500, 2000, 1500, 1000, 750, 500, 250bp. PCR reaction showed very pure band at the appropriate size.

was gel purified and both the purified pNYCOMPS vector and the insert were treated with T4 DNA polymerase in the presence of a single dNTP (dGTP for the vector and dCTP for the insert).

The reactions were incubated for 30 minutes at 22°C before EDTA was added to a final concentration of 10 mM and the reaction heated to 75°C for 20 minutes to inactivate the polymerase enzyme. Following this, 6 µl of the sticky ended vector was mixed with 3 µl of the sticky ended insert without further purification. A no-insert control was also performed. The resulting mixtures were incubated at room temperature for 10 minutes before 3 µl of EDTA was added to the reactions which were rapidly heated to 75°C and then slowly cooled back to room temperature. 2 µl was then used to transform chemically competent *E. coli* DH10B cells.

The following day a few large colonies were visible on the control plate and a similar number of large colonies were observed on the reaction plate. In addition to these a number of smaller colonies could also be seen on the reaction plate. Ten large and nine small colonies were picked from the reaction plate and colony PCR used to determine which, if any, contained *prnC-LIC*. After the colony PCR only one of the ten large colonies was shown to contain the *prnC-LIC* gene (figure 5.19A). This was expected as the larger colonies were also seen on the no insert control plate. The one large colony that did show evidence of *prnC-LIC* was located very near a smaller colony so cross contamination during colony picking may have caused this anomaly. Conversely every single small colony picked was positive for *prnC* (figure 5.19B), one of these colonies was picked and plasmid DNA recovered, sequencing of which confirmed the presence of the *prnC* insert without any mutations. This plasmid was designated pMW22 and the *E. coli* DH10B strain carrying this was designated as MW-037. The plasmid was transformed into *E. coli* BL21 for protein expression studies (forming strain MW-038).

Starter cultures of the LIC strain MW-038 were used to inoculate 500 ml cultures of LB (+ 50 µg ml^{-1} of kanamycin). Culture growth and protein production under the pNYCOMPS T7 promoter was carried out using the same methods as with the pET28a(+) based constructs. The resulting cell free extracts were incubated with 1 ml of Ni-NTA resin and purified in the same way as before. The purification process was visualised by SDS-PAGE (figure 5.20). The majority of the PrnC was still located in the insoluble fraction with very little seen in the cell lysate (figure 5.20 lane 1 and 2). Following purification and subsequent concentration the elution did seem to contain much more protein than previous strains tested. The only concern was the presence of what looked like two bands at the appropriate size. However this shadowing was also observed occurring with the ladder so is probably a gel artefact rather than a real result. Re-boiling the samples and running the analysis again continued to show this double band effect. This result also highlights the earlier work with the

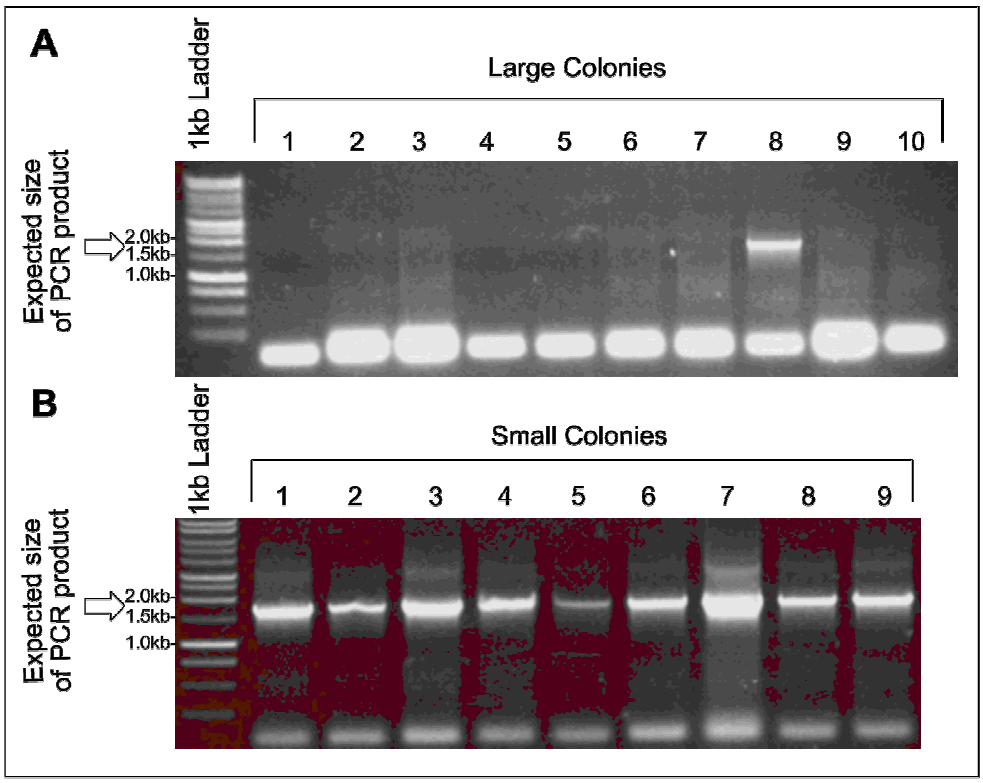


Figure 5.19: Colony PCR of large and small colonies present following ligation independent cloning with *prnC* and pNYCOMPS. The majority of the large colonies **(A)** showed no evidence of *prnC* insert. The smaller colonies **(B)** all clearly contained the new gene. The single positive result in lane A8 may have been the result of cross-contamination during the colony picking process. Expected size of PCR product *prnC-LIC* is 1739 bp and is marked with a white arrow. 1kb ladder sizes from top: 10000, 8000, 6000, 5000, 4000, 3000, 2500, 2000, 1500, 1000, 750, 500, 250bp.

chaperone proteins where a similar double banding effect suggested the co-purification of the chaperone proteins with PrnC. As before MALDI analysis would be useful to determine the exact composition of this second band.

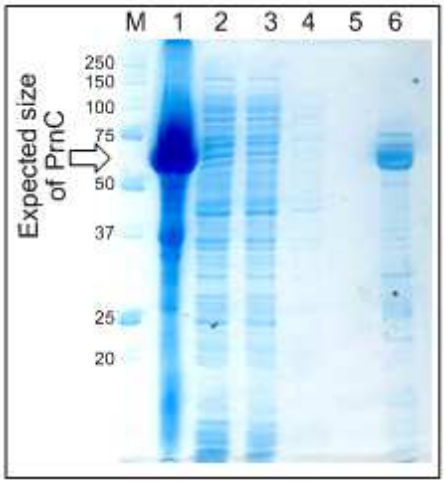


Figure 5.20: SDS-PAGE analysis of the purification of PrnC-LIC from pNYCOMPS vector. There is a large amount of PrnC in the insoluble fraction. The Ni-NTA elution of the nickel resin however shows a band of increased size compared to previous constructs. Lanes are as follows: **(M)** Protein marker (Biorad). Size in kDa. **(1)** Proteins present in insoluble pellet fraction. **(2)** Proteins present in soluble cell lysate. **(3)** Proteins showing no adherence to Ni-NTA resin. **(4)** Proteins washed off resin following Ni-NTA lysis buffer wash. **(5)** Proteins washed off resin following Ni-NTA wash buffer wash. **(6)** Proteins purified using Ni-NTA resin. Expected PrnC size – 67.24 kDa (shown with white arrow).

5.2.4 Selection of best constructs and large scale production and purification

Many different constructs had been trialled in the attempt to optimise the production of pure PrnC enzyme for crystal trials and enzyme characterisation. Of all the different methods used two were selected as the best candidates: MW-022 (pMW12 + pGro7 chaperones) (section 5.2.3.4.2) and MW-038 (Ligation independent) (section 5.2.3.4.3). These two were selected as they produced the most protein after Ni-NTA purification.

Both selected constructs were grown in 1 litre of LB culture. MW-022 was supplemented with 0.2% L-arabinose to induce the production of the groEL/ES chaperone. Growth was performed at 37°C until the density of the cultures reached the appropriate OD_{600} (0.5 for MW-022 and 0.8 for MW-038) and then protein production was induced with IPTG and incubation switched to 16°C for a

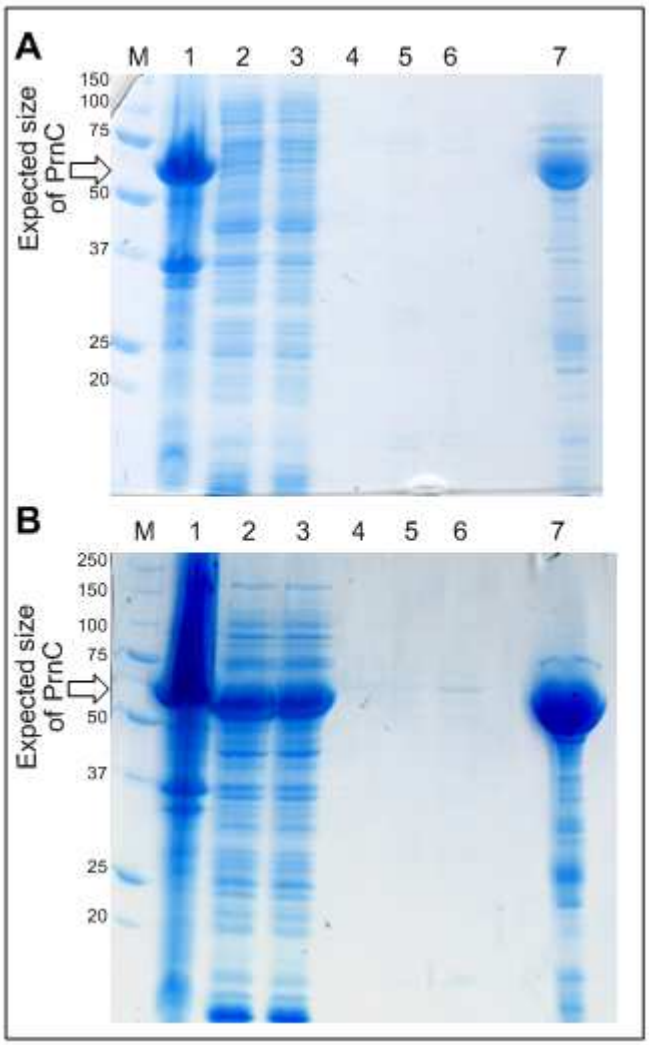


Figure 5.21: SDS-PAGE analysis of PrnC purification of the two chosen constructs. Lanes are as follows: **(M)** Protein marker (Biorad). Size in kDa. **(1)** Proteins present in insoluble pellet fraction. **(2)** Proteins present in soluble cell lysate. **(3)** Proteins showing no adherence to Ni-NTA resin. **(4)** Proteins washed off resin following Ni-NTA lysis buffer wash. **(5-6)** Proteins washed off resin following Ni-NTA wash buffer wash. **(7)** Proteins purified using Ni-NTA resin following concentration to 1 ml. Expected PrnC size – 67.24 kDa. **(A)** Purification of MW-038. The elution lane shows a good amount of PrnC present although there are a number of impurities. The same double banding feature can be seen that was observed in earlier experiments with this construct. **(B)** Purification of MW-022. A good amount of protein is present in the eluting fraction at the appropriate size range however as previously noted this could be due to the presence of chaperone protein groEL (size 70 kDa). A fair amount of non-specific proteins can also be seen.

further 22 hours. The cell extracts were generated *via* sonication and the Ni-NTA purification carried out using 2 ml of nickel resin. The resin was washed in the usual way, then binding proteins eluted from the resin using buffer containing a high concentration of imidazole (300 mM). The eluting fractions were combined and concentrated down to 1 ml. The buffer was then exchanged into HEPES protein storage buffer (1000x dilution of original elution buffer) and aliquoted into 100 µl fractions which were flash frozen and stored at -80°C until required.

The purification of both constructs was followed by SDS-PAGE analysis and the results can be seen in figure 5.21. Both constructs resulted in good amount of protein of the appropriate size, although the identity of the protein with MW-022 is still in doubt as some molecular chaperone may also be present. The LIC construct MW-038 shows the same double banding shadow that was observed in earlier studies again the identity of this is unknown. Both constructs also show additional protein contamination. This level of purity would not be sufficient for crystal structure studies but should be adequate for initial assessment of the activity of the enzyme and its substrate specificity; with a future aim of purifying the protein to homogeneity by FPLC (fast protein liquid chromatography).

A Bradford-type assay was used to assess the protein concentration within the eluting fraction of both constructs. As before a standard curve of BSA concentration versus absorbance at 595 nm facilitated the estimation of the protein concentration of MW-022 (PrnC + pGro7 chaperones) to be 10.6 mg per millilitre (158 µM) and the concentration of MW-038 (PrnC-LIC) to be 3.86 mg per millilitre (57.53 µM).

5.3 Conclusions and future work

Numerous attempts were made to optimise the production of soluble PrnC protein. Some improvement was made by increasing the length of the his-tag from six to eight histidine residues, which suggested that the affinity tag was not fully exposed on the surface of the protein. Further improvements to protein folding were made by co-expression of PrnC with the groEL /groES chaperone protein, although this made attempts at the purification of PrnC more complicated. Analysis of the purified protein with MALDI would be needed to verify this purification and to determine if any chaperone protein was being co-purified and to what extent.

The production of PrnC within the ligation independent plasmid pNYCOMPS-LIC also improved the solubility of produced PrnC, possibly due to the presence of the FLAG epitope. Purification may also have been aided by the presence of an affinity tag consisting of ten histidine residues. This construct also exhibited a double banding effect however and the composition of this pure protein should also be examined with MALDI.

Both constructs show a significant improvement in protein yield when compared to levels present at the beginning of the project and sufficient quantities can now be purified to begin preliminary characterisation of PrnC. Further purification steps would however be needed to enable progress towards a crystal structure.

Chapter 6:

Testing the Activity of Purified PrnC

6.1 Development of the assay

The activity of PrnC has not yet been demonstrated, with the previous attempt at assessing activity with purified protein having failed (chapter 5.2.3.2.2). We therefore chose to use another flavin dependent halogenase from the same biosynthetic cluster in order to benchmark the assay before investigating the activity of the purified PrnC enzyme obtained at the end of the previous chapter. The activity of tryptophan-7-halogenase, PrnA, has been demonstrated by the van Pee laboratory[172] and was also assessed as part of an Astrazeneca funded summer project that took place prior to the start of this PhD. The same assay used to investigate PrnA could also be applied to PrnC. Therefore the original PrnA assays were repeated, alongside PrnC, as a positive control to ensure all the assay components were functioning correctly, including the second flavin reductase enzyme that is required to maintain levels of $FADH_2$ in solution.

6.1.1 Producing PrnF and PrnA

Two strains previously prepared for the industrially funded summer studentship containing the tryptophan halogenase PrnA (RG-5076) and the flavin reductase PrnF (RG-5066) were selected for use as a positive control. PrnA had been amplified by Dr Sabine Grüschow from genomic DNA of *Pseudomonas fluorescens Pf-5* by PCR. During the summer studentship this gene was cloned into pET28a(+) with an N-terminal hexahistidine tag (for purification) to form plasmid pSG28. PrnF was prepared in a similar way but was cloned into pET21a(+) with a C-terminal histidine tag to form plasmid pSG24. These plasmids were both individually transformed into competent *E. coli* BL21 cells to form strains RG-5076 and RG-5066 respectively.

Both of these strains were grown in 500 ml of LB. Once the OD_{600} of the cultures had reached 0.8 the production of the proteins was induced by IPTG and growth cooled to 16°C for a further 20 hours. The cell lysate of each was generated by sonication and the proteins purified on Ni-NTA resin. Buffer exchange and storage was performed in the same way as previously used with PrnC. Visualisation of this purification by SDS-PAGE revealed that PrnA was produced at a fair level and at reasonable purity required for these assays. PrnF produced at very good levels and in very pure form (see figure 6.1).

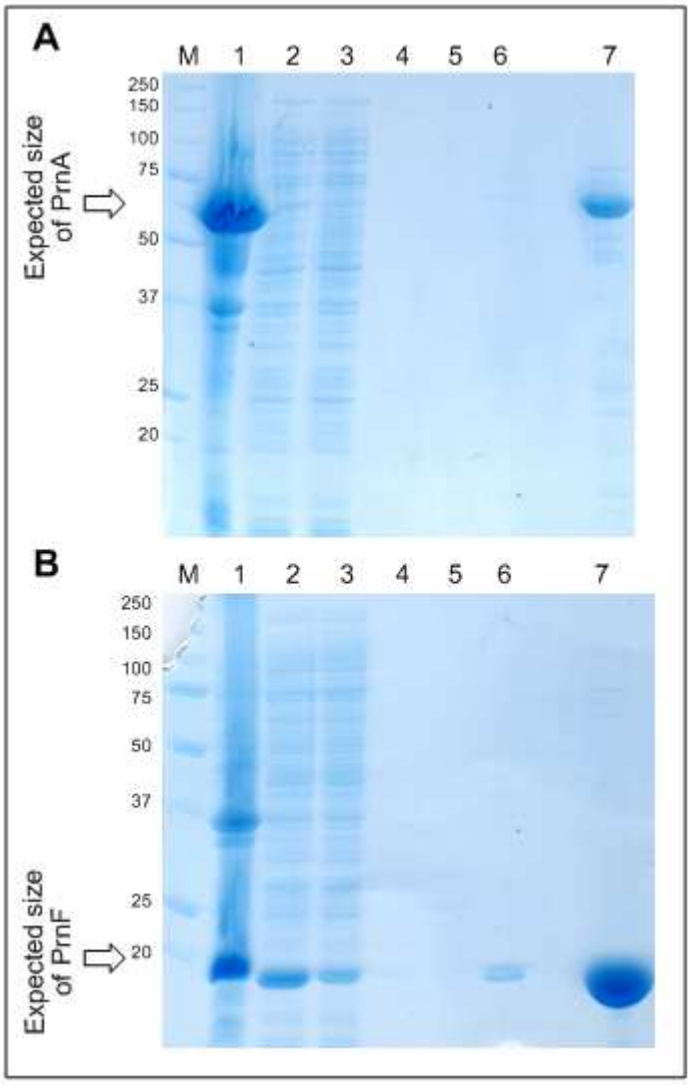


Figure 6.1: SDS-PAGE analysis of PrnA and PrnF purification. Lanes are as follows: (**M**) Protein marker (Biorad). Sizes in kDa. (**1**) Proteins present in insoluble pellet fraction. (**2**) Proteins present in soluble cell lysate. (**3**) Proteins showing no adherence to Ni-NTA resin. (**4**-**5**) Proteins washed off resin following Ni-NTA lysis buffer wash. (**6**) Proteins washed off resin following Ni-NTA wash buffer wash. (**7**) Proteins purified using Ni-NTA resin following concentration down to 1 ml. Expected PrnA size – 63.2 kDa and PrnF – 21 kDa is shown with a white arrow. **(A)** Purification of RG-5076. The elution lane shows a good amount of PrnA present. **(B)** Purification of RG-5066. A very good amount of PrnF protein is present in the eluting fraction.

As with the PrnC constructs, a Bradford style assay was used to determine the protein concentration of both proteins. PrnA was estimated to contain 2.2 mg of protein per millilitre (34.81 µM) and PrnF concentration estimated to be 11.15 mg per millilitre (530.95 µM). As seen with the flavin reductase produced for the preliminary PrnC assays in the last chapter, the concentration of PrnF was deemed too high to be directly added to the assays so was diluted ten times before use.

6.1.2 Testing PrnF activity

The flavin reductase component is essential to the activity of the halogenase enzymes in this family. This enzyme continually recycles the flavin cofactor back into the reduced form ($FADH_2$) by oxidising NADH. Before setting up the halogenase assays, the activity of PrnF was assessed. The conversion of NADH from the reduced form to the oxidised form can be monitored by following the UV absorbance at 340 nm. Reduced NADH has a distinctive UV absorption spectra with two peaks; one at 260 and another at 340 nm (see figure 6.2). When it is oxidised the peak at 340 nm disappears. Therefore by monitoring the flavin reductase reaction at 340 nm it is possible to observe the oxidation of NADH in real time as the flavin substrate is being reduced by PrnF.

Simple assays were set up in 1 ml cuvettes containing NADH (200 µM), FAD (30 µM) and NaCl (50 mM) in 20 mM Tris-HCl buffer pH 7.5. The absorbance of the resulting cuvette was monitored for a set amount of time (initially for 30 minutes but later reduced to 20). Half way through this time period, 10 µl of the PrnF enzyme (diluted 10x in protein storage buffer, final concentration 53 µM) prepared above was added and any change in absorbance monitored at 340 nm.

In the absence of enzyme (blank protein buffer added to assay) the absorbance was not seen to decrease over the 30 minutes of the reaction (figure 6.3A).When the enzyme was added (10x diluted

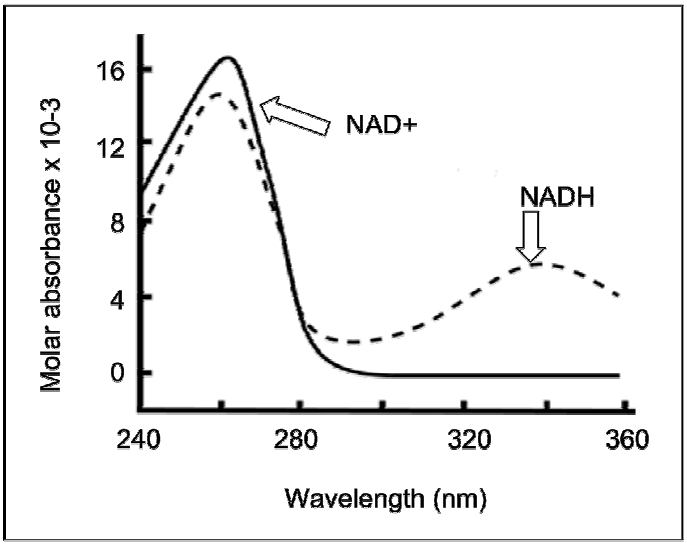


Figure 6.2: UV spectrum of NAD+ and NADH at 260 and 340 nm. The NADH spectrum exhibits two peak absorbance values at about 260 nm and at 340 nm which is shown in the diagram with a dotted line. Oxidation to NAD+ results in the loss of absorption at 340 nm which is represented in the diagram with a solid line. This can be used to distinguish between the reduced and oxidised forms of NAD.

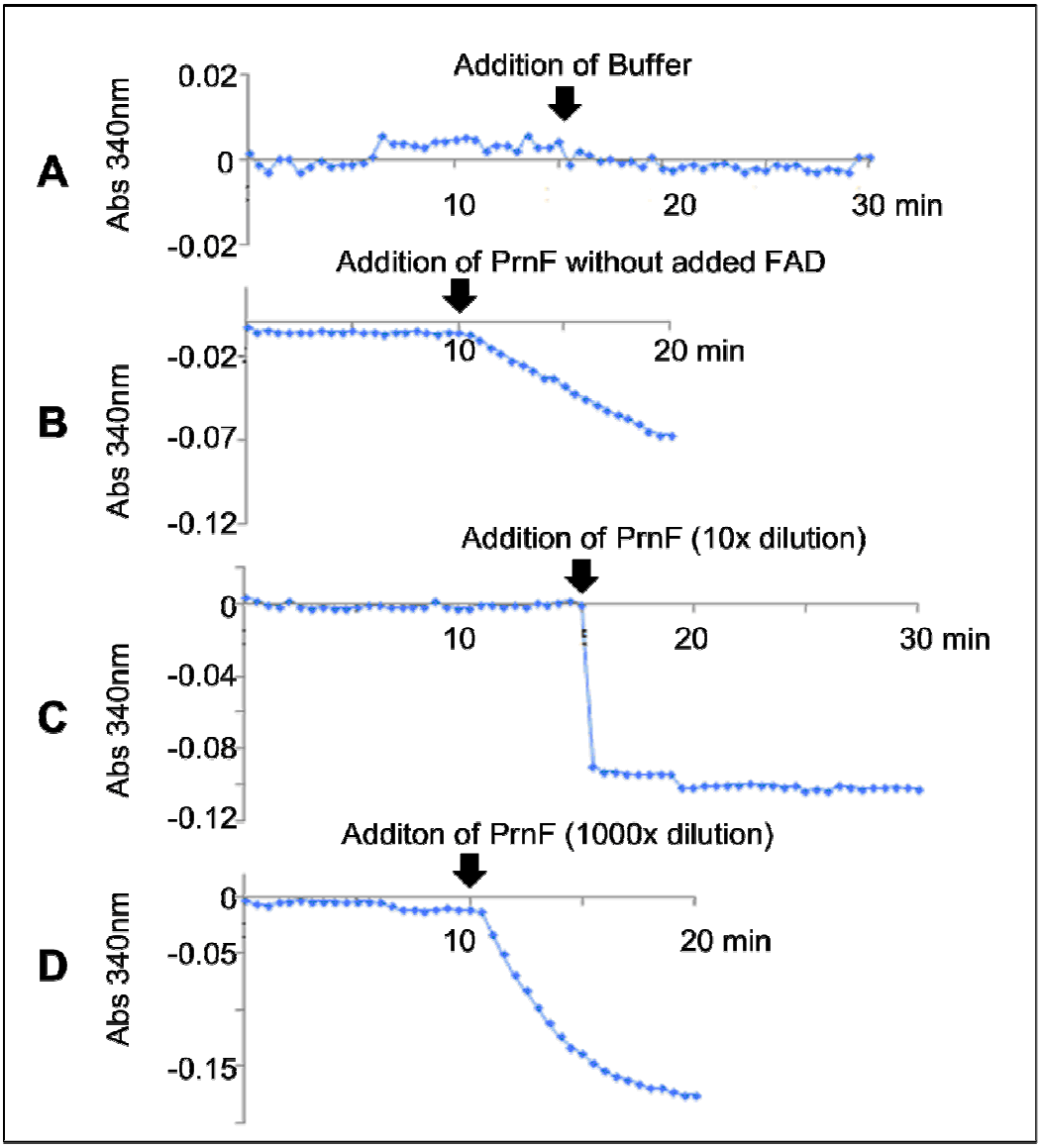


Figure 6.3: Consumption of NADH over time by PrnF, present at different concentrations, as monitored by the absorbance change at 340 nm. (A) Absorbance change of the assay solution over 30 minutes. Buffer was added after 15 minutes as a negative control. No absorbance change was detected. **(B)** Absorbance of assay solution in the absence of FAD cofactor at 340 nm over 20 minutes. Reaction was supplemented with PrnF after 10 minutes. A slow decrease in absorbance was observed in the absence of added FAD, possibly due to bound FAD present within the purified enzyme. **(C)** Absorbance of assay mixture over 20 minutes. A 10 fold diluted PrnF solution was added following 10 minutes. A rapid loss of absorbance at 340nm corresponding to oxidation of NADH by the enzyme was seen. **(D)** Absorbance of assay mixture over 20 minutes. A 1000 fold diluted PrnF solution was added following 10 minutes. A slower loss of absorbance at 340nm corresponding to oxidation of NADH by the enzyme was seen compared to the 10 fold diluted PrnF. Enzyme solution always added half way through run as indicated by position of black arrow.

from stock) to a second reaction the absorbance at 340 nm rapidly decreased suggesting that the NADH was being very quickly oxidised to NAD+ (figure 6.3C). Decreasing the enzyme concentration (1000x dilution, 5.3 μM) slowed this rate of oxidation down to an observable level (figure 6.3D). This reduction in rate with lowering of enzyme concentration very clearly demonstrates that this is an enzyme dependent process, therefore showing that the prepared and purified PrnF was performing its function and could be used for the halogenase assays. Interestingly when the reaction was allowed to proceed in the absence of FAD substrate and the less diluted enzyme (10x), the absorbance at 340nm slowly decreased (figure 6.3B). This is likely to be due to FAD being bound in the active site of the purified enzyme (the purified protein contains a distinctive yellow appearance). With the activity of PrnF demonstrated, the activity of PrnA was assessed as a positive control.

6.1.3 Testing PrnA activity

A simple assay was designed to assess the activity of PrnA based on previous work within our lab as well as procedures reported in the literature.[202] PrnA enzyme prepared earlier (50 μl, final concentration 17.5 μM) was added to an assay mixture containing FAD (10 μM), NADH (5 mM), NaCl (100 mM), L-tryptophan (1 mM) and PrnF (4 μl 10x dilute) in a final volume of 100 μl Tris-HCl, pH 7.5. Tryptophan is the substrate for PrnA, while the NaCl acts as a source of chloride ions. As a negative control experiment, one reaction was supplemented with the same volume of protein buffer instead of the PrnA containing buffer solution.

The reactions were incubated at 30°C for approximately 18 hours before the reaction was quenched with the addition of equal volumes of formic acid. Following centrifugation to collect the precipitated protein, the samples were analysed for tryptophan production using the same HPLC method developed for the detection of halo-tryptophan in chapter 2 (method LCgrad_MeOH). The HPLC traces can be seen in figure 6.4. In the absence of halogenase enzyme there was a single peak at 7 minutes that corresponded to the retention time of the tryptophan substrate (figure 6.4A). The large initial peak appearing from 2 minutes is the signal of NADH. When PrnA was added to the enzyme mixture an additional peak appeared on the HPLC at the later retention time of 9 minutes (figure 6.4B). The retention time and PDA spectrum of the new peak was consistent with 7-chloro-tryptophan standards analysed on HPLC at the same time and the shows the formation of product with the enzyme. Comparison of the peak areas of the starting material and product enabled the overall conversion to be estimated to be 11.5%. This yield however was much lower than previous work conducted with this enzyme. Further studies highlighted that the stock of NADH was the cause of this lower than expected yield. The percentage of oxidised/reduced NADH in a solution of known

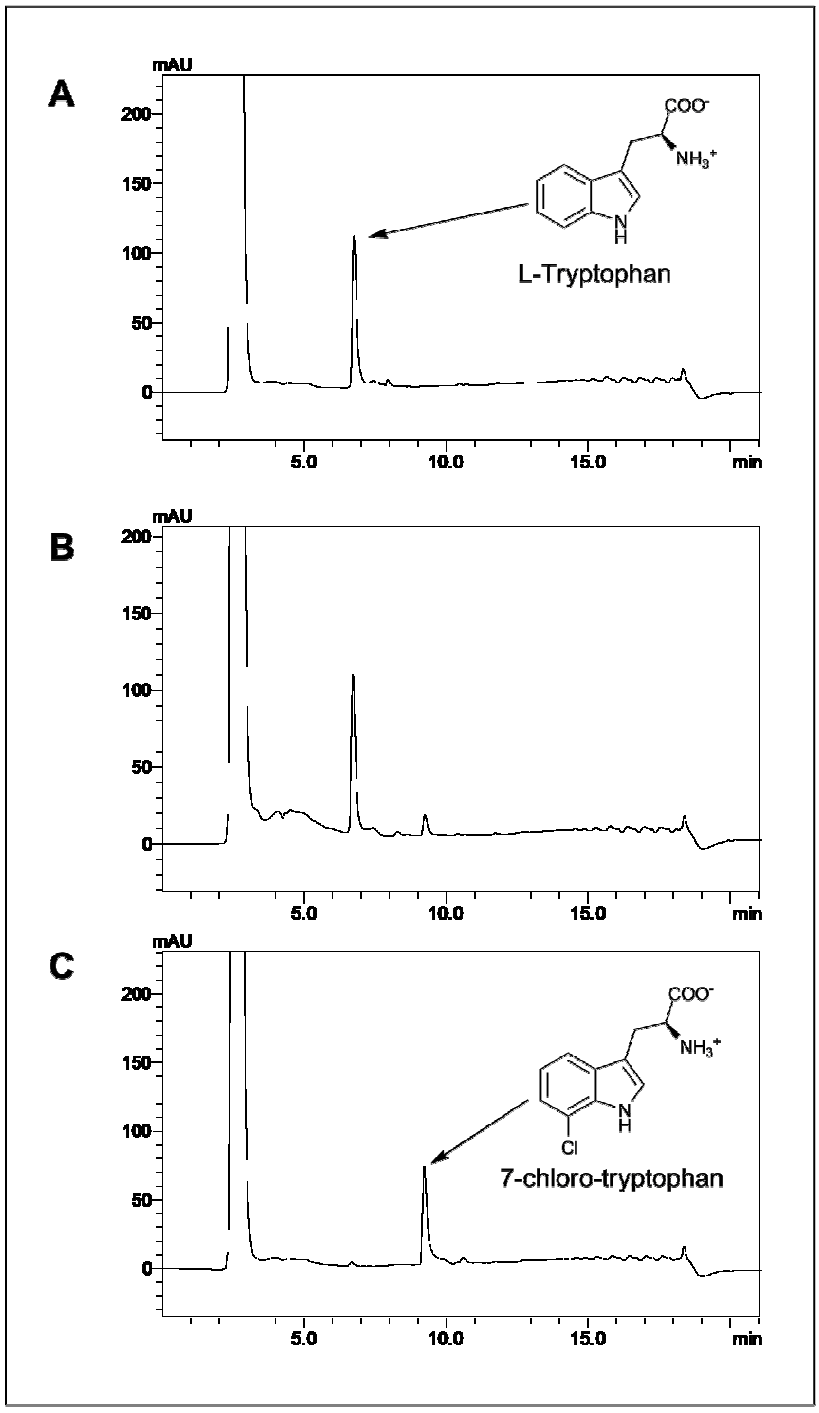


Figure 6.4: HPLC chromatograms of the activity of PrnA. (A) Chromatogram of the no PrnA control. In the absence of halogenase enzyme a single peak corresponding to un-halogenated tryptophan can be seen with the retention time of approximately 7 minutes. **(B)** The chromatogram of the assay in the presence of PrnA shows the un-halogenated tryptophan peak at 7 minutes. A new peak has formed with a later retention time of 9 minutes. This corresponds to the retention time of halogenated product. **(C)** The same assay conducted with fresh NADH demonstrates enhanced performance over the degraded NADH stock. The tryptophan peak at 7 minutes has vanished and is replaced with the 9 minute peak corresponding to the chlorinated product. In all three chromatograms there is a very large peak at 3 minutes corresponding to NADH.

concentration can be estimated by measuring the absorbance of the solution at 340nm and using the beer-lambert law (A = εcl). This law states that the absorbance is equal to the product of the extinction coefficient of the compound (ε), the concentration of the sample (c) and the path length of the cuvette (l). Using this law, the extinction coefficient of NADH at 340 nm (6220 M^{-1} cm^{-1}), the known concentration of the sample and the path length of 1 cm it was calculated that only 20% of the NAD stock was in the reduced form. This meant that the old stock of NADH had oxidised and was limiting the halogenase reaction. A fresh solution was made and the same calculation estimated a greater than 94% proportion of NAD in reduced form. This new solution was used in a set of repeat experiments.

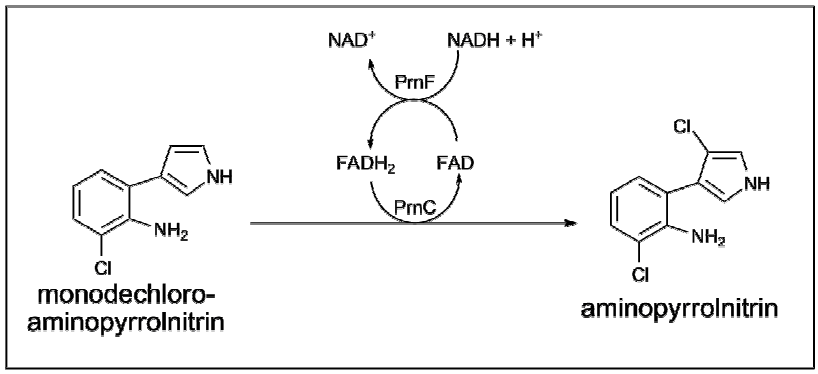
This effect of replacing the NADH with a freshly purchased batch was significant. As can be seen in figure 6.4C, after 18 hours of reaction the 7 minute tryptophan peak completely disappeared to be replaced with a 9 minute peak corresponding to the product with an estimated yield of greater than 95%. It is unknown whether the same factor may have been limiting the results of the preliminary PrnC assays prepared earlier with the pyrrole substrates. Using fresh NADH may have improved the findings.

This reaction was successful and provided a positive control, and an assay that could be utilised to assess in parallel other flavin dependent halogenase enzymes. The assay was now ready to be applied to PrnC, an enzyme for which no activity had previously been demonstrated *in vitro*.

6.2 PrnC assays

6.2.1 Assays and HPLC analysis

The previous preliminary assays conducted with purified PrnC did not result in any product formation. The exact reason for this was unknown but could well be due to the fact the none of the substrates tested were the actual monodechloroaminopyrrolnitrin substrate that this enzyme is postulated to act upon (scheme 6.1).[170] The substrate specificity of this family of enzymes is known to be strict, *in vivo* studies by Seibold *et al.* had previously suggested that PrnC does not accept monodechloroaminopyrrolnitrin chlorinated at the 6 position rather than at the 7, suggesting the substrate specificity of PrnC is very high.[174] However the addition of a bulky chlorine substituent may prevent this modified substrate from fitting inside the substrate binding pocket so other, less bulky substrates may fare better.

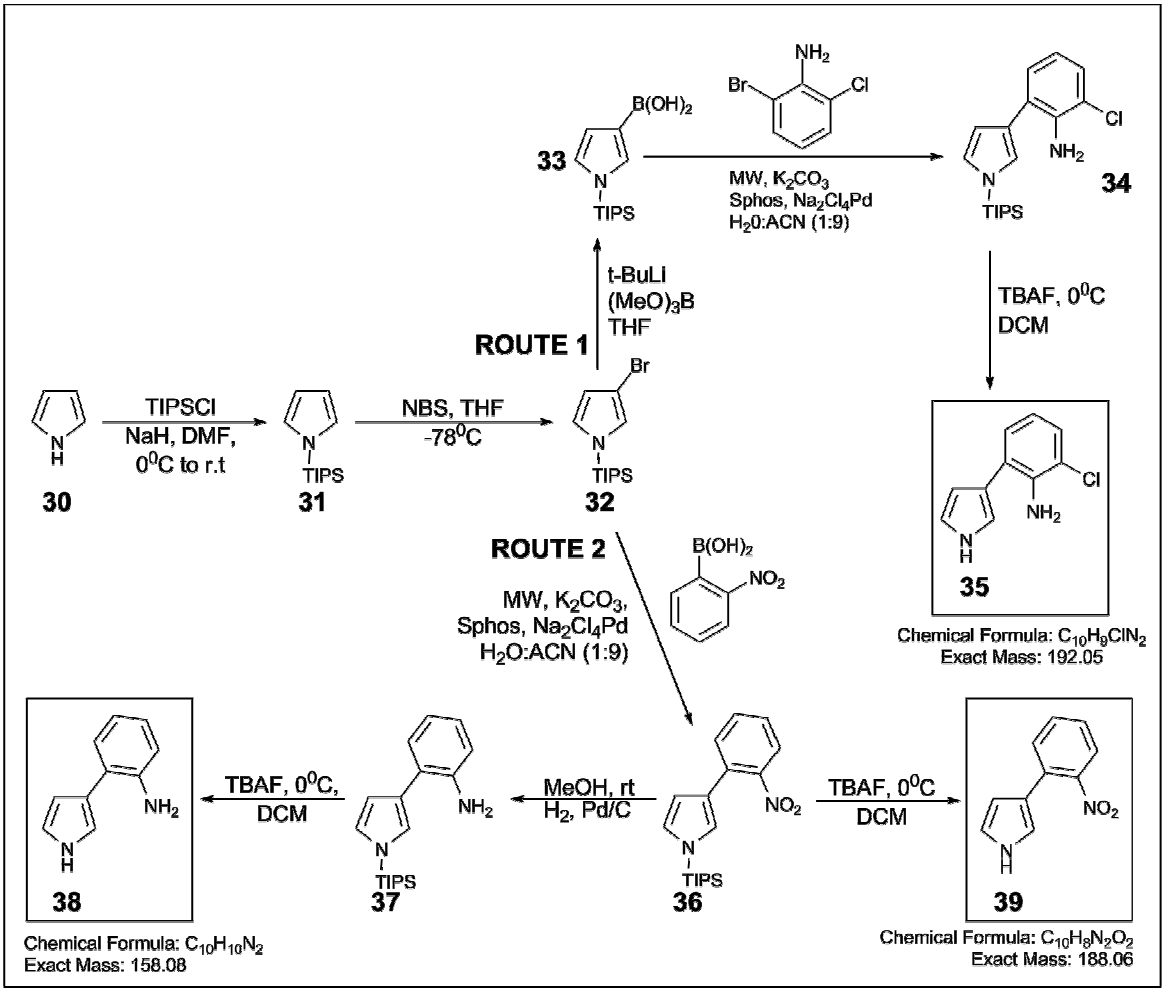


Scheme 6.1: Reaction catalysed by PrnC and PrnF. PrnC selectively chlorinates the pyrrole moiety of monodechloroaminopyrrolnitrin (the product of previous enzymes in the biosynthetic pathway PrnA and PrnB). PrnC utilises $FADH_2$ to generate the Cl^+ equivalent needed for the electrophilic aromatic substitution reaction. The reduced form of this FAD cofactor is continually re-formed by the partner enzyme PrnF which oxidises NADH.

To properly attempt the characterisation of PrnC the natural substrate had to be chemically synthesised. The synthesis was kindly attempted by synthetic chemists working within the laboratory: Dr Abhijeet Deb Roy and Tony Abou Fayyad. Full details of their synthesis and full compound characterisation can be found in appendix 3. The approach was loosely based on and adapted from Morrison *et al.* (2009).[225] Two different synthetic routes were taken (scheme 6.2). Route 1 enabled the generation of the natural substrate monodechloroaminopyrrolnitrin (4-(2-amino-3-chloro-phenyl)-pyrrole) which is shown as compound **35** in scheme 6.2. Route 2 resulted in two slightly modified substrates, the first (compound **38**) being the natural substrate minus the chlorine on the phenyl ring (4-(2-amino-phenyl)-pyrrole) and the second (compound **39**) being the nitro form of compound **38** (4-(2-nitro-phenyl)-pyrrole). The two non-natural substrate analogues were easier and faster to synthesise so were prepared as initial testing substrates while the more complicated natural substrate was being synthesised.

Both routes utilised the same chemistry to produce mono-brominated *N*-TIPS (triisopropylsilyl) protected pyrrole (**32**). This is where the two synthetic routes diverged. The route to the two substrate analogues involved a Suzuki cross-coupling reaction between this pyrrole and 2-nitro-phenylboronic acid allowing the generation of **36** (4-(2-nitro-phenyl)-*N*-TIPS-pyrrole) which was deprotected to yield the second substrate analogue **39**. To yield the first substrate analogue, **38**, compound **36** was hydrogenated and then deprotected.

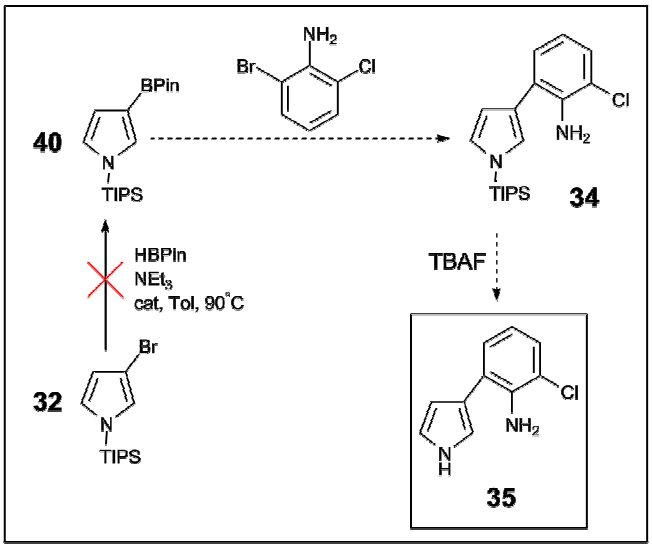
The first attempted route to access the monodechloroaminopyrrolnitrin substrate based on the synthesis described by Morrison *et al.* can be seen in scheme 6.3. This synthesis involved the formation of *N*-TIPS-pyrrole-3-pinacolboronate ester from **32. C**oupling of **32** to 2-bromo-6-chloro-



Scheme 6.2: Synthetic scheme for the preparation of possible PrnC substrate and analogues. The starting material for the synthesis was pyrrole, **30**, which was protected and brominated to yield **32.** At this point the synthesis diverged. **Route 1** allowed the generation of the PrnC substrate monodechloroaminopyrrolnitrin by the generation of the *N*-TIPS-pyrrole-3-boronic acid **33**. Suzuki cross-coupling with 2-bromo-6-chloroaniline and subsequent deprotection of **34** yielded the product **35.** The second route coupled **32** with 2-nitrophenylboronic acid to yield 4-(2-nitrophenyl)-pyrrole **36.** Deprotection yielded product **39** while hydrogenation followed by deprotection yielded product **38.** Reaction conditions are shown. Full experimental details can be found in appendix 3.

aniline and subsequent deprotection to yield the product, **35**. However after several attempts the *N*-TIPS-pyrrole-3-pinacolboronate ester could not be successfully synthesised. Therefore a modified route took the protected, brominated pyrrole, **32**, and converted it into the boronic acid, **33** instead (scheme 6.2). Suzuki cross-coupling with 2-bromo-6-chloroaniline resulted in the *N*-TIPS protected monodechloroaminopyrrolnitrin, **34**. Subsequent deprotection resulted in the natural substrate for PrnC (cmpd **35**).

The first compound to be successfully synthesised was compound **39** (4-(2-nitro-phenyl) pyrrole) and



Scheme 6.3: First attempted synthetic strategy to obtain the natural PrnC substrate 35. The starting material for the synthesis was *N*-TIPS-pyrrole, **32**. The proposed route involved the formation of the *N*-TIPS-pyrrole-3-pinacolboronate ester **40**, which would be coupled with 2-bromo-6-chloroaniline to yield **34**. Deprotection of **34** would yield monodechloroaminopyrrolnitrin **35**. The formation of the pinacolboronate ester, however, failed after several attempts.

therefore this was the first compound tested for activity with PrnC. The identity of the compound was confirmed by ^{1}H and ^{13}C NMR and high resolution mass spectrometry at the EPSRC National Mass Spectrometry Service at Swansea identified the correct mass of *m/z* 188.

Both variants of PrnC purified earlier (The PrnC-LIC and the mutated PrnC grown in the presence of groEL/ES molecular chaperones) were tested for activity against this substrate. Of the three substrates studied this one is the least like the natural one. The amine group at position 2 of the phenyl ring has been replaced with a nitro group relative to the natural substrate, the chlorine at position 3 is also missing. The chlorine normally present at this position is sterically bulky and may play a role in orientating the substrate correctly in the active site, therefore if any chlorination occurs with this substrate it may appear stereospecifically altered.

Before any assays were performed a 0.5 mM solution of this substrate was analysed on the HPLC using the standard tryptophan method (LCgrad_MeOH) to assess whether or not a new method would have to be designed for these compounds. The resulting chromatogram (figure 6.5) showed a very clear peak corresponding to the compound at 12 minutes. The purity of the compound and the detection by the PDA detector were good enough that trial assays were performed with PrnC. The only problem was that the compound was seen to elute in 90% methanol at the start of the wash phase. Any chlorinated product would be expected to elute later due to the increased lipophilicity.

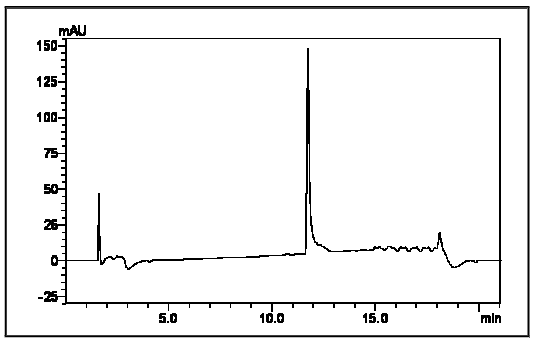


Figure 6.5: Reverse phase HPLC chromatogram of 0.5 mM 4-(2-nitro-phenyl) pyrrole (compound 39). A very clear peak of compound **39** was seen by the HPLC. The early peak at 1.5 minutes was also present in the blank. LC method: LCgrad_MeOH.

Therefore a new solvent gradient with a slower increase in methanol concentration was applied. The new method raised the percentage of methanol: water from 10% to 95% over 20.5 minutes (the previous method increased the same over just 12.5 minutes). This new method was called LCgrad_Pyrrole. Using this new improved gradient the compound eluted at 16.03 minutes in 75% methanol.

The initial test assays were designed to be very similar to the successful PrnA assays used previously. 50 µl of PrnC solution (final enzyme concentration of 28.5 µM of LIC or 80 µM of the pGro construct) was added to a final volume of 100 µl containing PrnF (2 µM), FAD (10 µM), NAD (5 mM), NaCl (100 mM) and compound **39** (1 mM) in 20 mM Tris-HCl buffer, pH 7.5. A control reaction containing just protein storage buffer instead of PrnC solution was also performed. The reactions were incubated at 30°C for four hours. The assays were analysed by HPLC to check for any signs of product formation.

The enzyme blank looked much the same as the earlier standard of the compound with a large new peak present at 2-3 minutes which is consistent with the presence of NADH (**A**, figure 6.6). The addition of PrnC however results in the formation of new peaks not present in the blank or the standard. The PrnC construct co-expressed with chaperone proteins shows the formation of a small peak at 16.8 minutes, adjacent to the peak representing the starting material (**B**, figure 6.6). There is also the possibility of another smaller peak at around 18 minutes. However this peak is very small and may just be baseline.

The PrnC-LIC construct shows very similar results. Two new peaks have appeared adjacent to the starting material, one at 16.8 minutes and another at around 18 minutes (**C**, figure 6.6). The first additional peak (16.8 minutes) was assumed to belong to mono-halogenated product while it was supposed that the possible lack of regio-selectivity due to the loss of chlorine from the substrate,

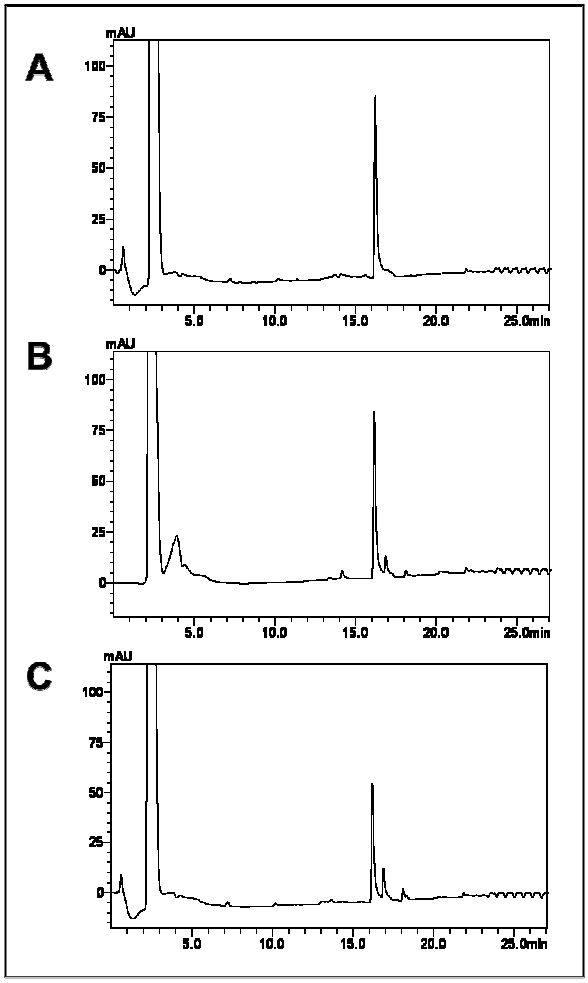


Figure 6.6: Reverse phase HPLC traces of initial trials on the halogenation of compound 39 by PrnC. **(A)** This assay was run in the absence of halogenase enzyme and shows a clear starting material peak at 16.2. **(B)** Assay performed with PrnC grown in the presence of chaperones shows the starting material at 16.14 minutes. Additional peaks can be seen that are not present in the enzyme control. There is a new small peak at 16.8 and an even smaller one at 18 min. **(C)** The reaction performed by the PrnC-LIC protein again shows the starting material at 16.18 min. Multiple new peaks not present in the standard or the control can be seen at 16.8 and at 18 min. The large initial peak at 2-3 minutes is NADH. LC method: LCgrad_Pyrrole

may enable the enzyme to halogenate the substrate more than once and therefore result in the formation of multiple products.

Just as with the tryptophan halogenase assays an estimation of the conversion was made based on the relative peak areas of the different peaks. Based on the first peak it was estimated that 15% of the substrate had transformed into the second HPLC peak by the chaperone pGro7 containing PrnC enzyme. The PrnC-LIC enzyme was estimated to have converted 24% of the starting material into a

product. Even though the LIC construct was technically lower in protein concentration it appeared to perform better than the protein grown in the presence of molecular chaperones, this may be evidence to suggest that some of the chaperone proteins had been co-purified with PrnC and therefore not all the protein present was catalytically active. This assumed higher purity of the ligation independent cloning protein was sufficient to restrict later assays to this construct. The identity of the new peaks was not known and therefore it was not possible to prove that product had been formed. LCMS analysis was needed to confirm activity.

The extended gradient does not seem to be necessary as the possible halogenated compounds retention times are not shifted dramatically, therefore to enable higher throughput analysis the shorter LCgrad_MeOH method was reclaimed for future analysis.

With the completion of the synthesis of the other two substrates they were also tested with the PrnC-LIC enzyme. 1 mM solutions of the natural substrate **35** (monodechloroaminopyrrolnitrin) and the second analogue **38** (4-(2-aminophenyl) pyrrole) were first analysed with the HPLC to check retention times.

The HPLC trace of the natural substrate (**A**, figure 6.7) did not show the expected peak. The only major peak present (apart from cluster of peaks present in the blank at 2-3 minutes) showed at 5 minutes. The retention time of the substrate analogue **39** had previously been shown to be 12 minutes using the same HPLC method (LCgrad_MeOH), it was expected that the presence of the additional chlorine atom in the natural substrate would increase the hydrophobicity of the natural substrate comparatively. As such the early appearance of the peak at 5 minutes seems incongruous. The intensity of the peak is also extremely low compared to the previous substrate analogue at the same concentration, suggesting a lower than expected concentration and therefore a significant issue with purity. These two factors made it doubtful that the synthesised natural substrate was of the correct structure. Analysis of the ^{1}H and ^{13}C NMR data for compound **35** and high resolution mass spectrometry analysis (appendix 3) failed to identify the correct product and therefore the attempt to synthesise the natural substrate was deemed a failure. The attempt to characterise PrnC was continued with the two non-natural substrates.

The HPLC trace of the substrate analogue **38** was much more in line with expectations (**B**, figure 6.7), showing a clear peak with approximately the same intensity and retention time as the substrate analogue **39**. The presence of a few smaller peaks in the same region of the HPLC (7-9 minutes and 13-14 minutes) suggest possible impurities in the synthesised compound and potentially even

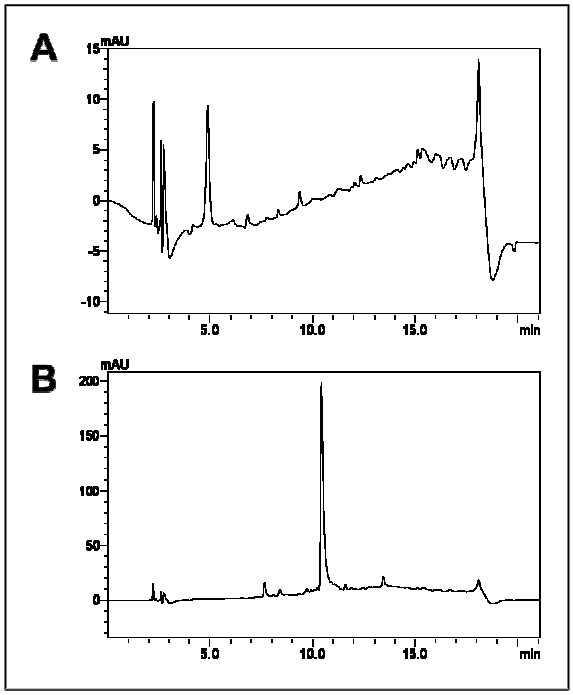


Figure 6.7: Reverse phase HPLC chromatogram of 1 mM 4-(2-amino-phenyl) pyrrole (compound 38) and monodechloroaminopyrrolnitrin (compound 35). (A) Standard sample of the synthesised natural substrate (monodechloroaminipyrrolnitrin) failed to display a clear peak. The major peak for this compound at 5 minutes seems too low in retention time and intensity to be the correct compound. **(B)** Standard sample of the second substrate analogue, **38,** demonstrates a clear peak at 10-11 minutes. Some residual smaller peaks may indicate impurities in the sample. LC method: LCgrad_MeOH

evidence of product decomposition. ^{1}H and ^{13}C NMR analysis of the synthesised starting material and high resolution mass spectrometry confirmed the structure of product **38**.

With the natural substrate discounted, the two remaining compounds **38** and **39** were added to a set of assays identical to the conditions used for the initial testing of **39** above but with reduced NADH concentration to enable eventual effective LCMS analysis of the reactions.

Compound **39** (4-(2-nitro-phenyl) pyrrole) had already shown a possible positive result with the enzyme. Repeating this experiment produced the same result: HPLC traces of the reaction mixture showed three peaks in the range of 11-13 minutes where the starting material had previously been seen to elute using this HPLC method (**B**, figure 6.8). These multiple peaks could correspond to starting material and halogenated products. The samples were set aside for mass spectrometry analysis.

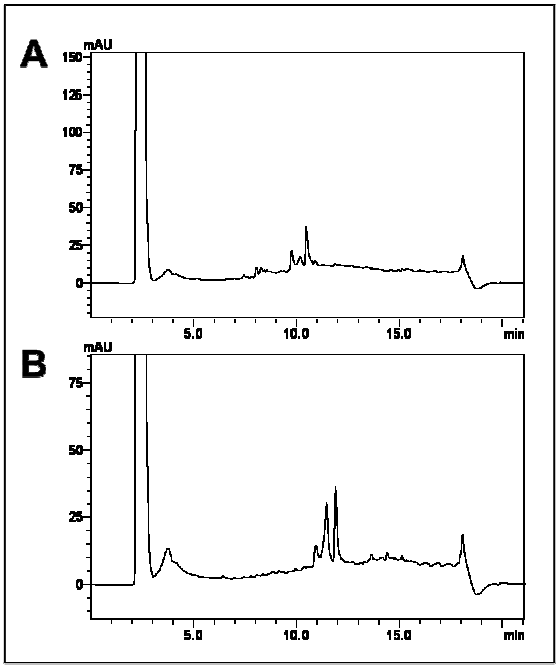


Figure 6.8: Reverse phase HPLC traces of halogenation assays with PrnC substrate analogues; 38 and 39, by PrnC-LIC. (A) The reaction performed by the PrnC-LIC protein with compound **38** (4-(2-amino-phenyl) pyrrole) shows the same peak at 11 minutes that was present in the previous control reactions. Other multiple peaks are present in the 10 minute region of the starting material. **(B)** Reaction performed with compound **39** (4-(2-nitro-phenyl) pyrrole) also shows the peak at 11 min visible in the control experiments. Two additional peaks can be seen in the 11-13 minute region where the starting material elutes. Between 2-3 minutes the broad NADH peak can be seen again. LC method: LCgrad_MeOH.

The reaction mixture with the aminated substrate analogue **38** (4-(2-amino-phenyl) pyrrole) also resulted in multiple peaks when analysed by the HPLC (**A**, figure 6.8). Samples of the starting material run on the HPLC had a retention time of between 10 and 11 minutes using this HPLC method. In this region there were three unknown peaks in the reaction mixture analysed by the HPLC (**A**, figure 6.8). It was unclear what these peaks represented; starting material, halogenated products or decomposition products. The fact that the substrate analogues, **38** and **39,** are missing a chlorine substituent on the phenyl ring relative to the natural substrate may make them more prone to flexible regio-selective halogenation with possible multiple products which may explain the appearance of multiple peaks. As with the other assays tested mass spectrometric analysis would be useful in answering the outstanding questions.

6.2.2 Analysis of the substrates and reaction mixtures with LCMS

1 mM standards of the two substrate analogues were analysed by LC-MS using the method LC-Xbridge-Pac using an electrospray ionisation source (ESI) in positive ion mode (figure 6.9).

Unfortunately none of the two compounds showed the correct mass when analysed with the LCMS; the 4-(2-amino-phenyl)-pyrrole, **38**, gave a mass of *m/z* 205 and the 4-(2-nitro-phenyl)-pyrrole **39** gave three masses of *m/z* 313, 335 and 350. This may be due to insufficient ionisation conditions on the instrument at UEA as the high resolution mass spectrometry data from Swansea had previously verified the masses of the substrates. Despite the incorrect mass results for the starting materials the earlier HPLC analysis had identified new peak formation when these compounds were exposed

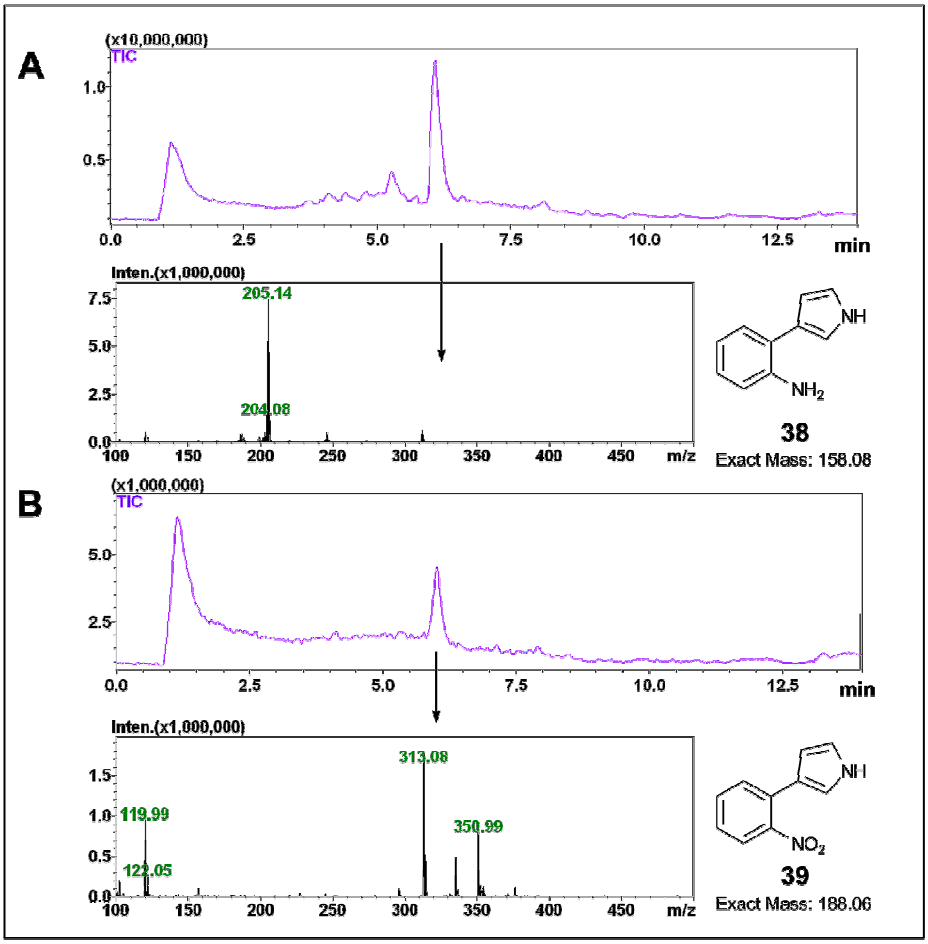


Figure 6.9: LCMS analysis of PrnC substrate analogues 38 and 39. Analysis of substrate compounds showing the total ion chromatogram and the mass spectrum of the major peak. Expected structure and mass shown. **(A)** Aminated analogue **38. (B)** Nitro analogue **39**. LCMS method: LC-Xbridge-Pac.

to PrnC, therefore the same assays were also analysed by LCMS to determine the identity of these new peaks.

The total ion chromatograms and the PDA data from the LCMS analysis of the assays are shown in Figure 6.10. Ions that corresponded to the theoretical molecular weight of the starting material were

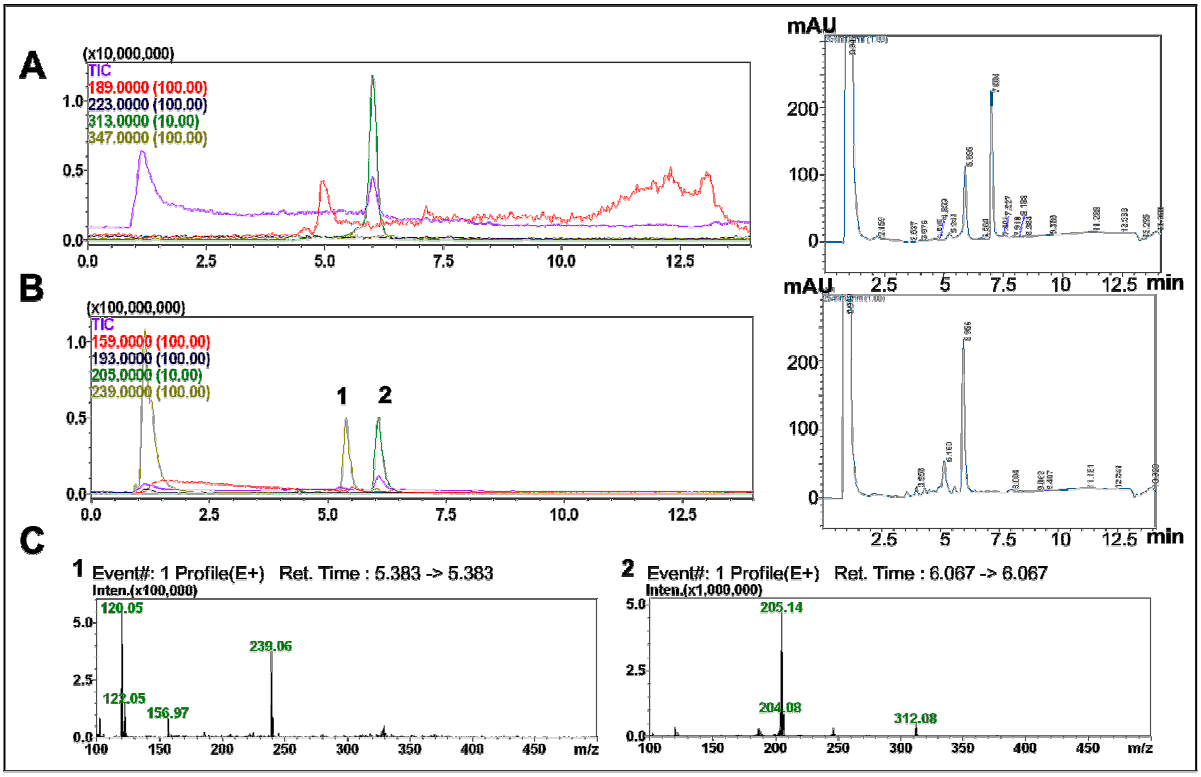


Figure 6.10: LCMS analysis of PrnC assays with substrate analogues 38 and 39. Analysis of substrate compounds showing the total ion chromatogram and the PDA spectrum over the 15 minute LCMS analysis. Abundance of ions of the expected *m/z* of the starting material are shown in red, the abundance of ions representing chlorinated product of that starting material are shown in blue. Abundance of ions of the *m/z* of the actual starting material identified in the previous LCMS analysis of the starting material is shown in green, abundance of ions representing chlorination of these detected ions is shown in olive. **(A)** Enzyme assay mixture containing substrate analogue **39**. MS data shows no evidence of halogenated product but the PDA trace clearly shows the formation of a new PDA peak at a later retention time than the starting material (7.5 min). **(B)** Enzyme assay mixture containing substrate analogue **38**. PDA trace shows a large peak corresponding to the starting material and a smaller peak that corresponds to a mass that is M+34 of the detected starting material but the earlier retention time is counter-intuitive. **(C)** Individual mass spectra of the two individual peaks, **1** and **2**, found in chromatogram **B**.

targeted as well as the actual mass of the starting material that had been detected in the earlier LCMS analysis of the pure synthesised compounds. Ions representing monochlorination of both potential starting masses were also targeted. The HPLC traces of the two substrate analogues (**38** and **39**) both demonstrated the formation of new peaks that may represent reaction products.

The LCMS analysis of a standard sample of compound **39** (4-(2-nitro-phenyl)-pyrrole) showed a single ion peak at around 5.8 minutes. This peak was composed of ions of *m/z* 313 which matches the mass of the detected starting material. No evidence of the chlorination of this precursor ion (*m/z* 347) could be found (figure 6.10A). A small peak of *m/z* 189 was detected which could represent the expected mass of the synthesised compound but this was very small and did not match any peaks found on the PDA analysis and for these reasons it is unlikely that this peak represents the starting material. The PDA chromatogram displayed three major peaks at 1, 5.9 and 7 minutes. The peak at 1 min can be explained by the presence of the NADH cofactor while the other two presumably represented the substrate and a possible halogenated or breakdown product. The peak at 5.9 minutes is represented by the large 313 ion peak in the mass spectrum, so presumably represents starting material. The second peak has a later retention time which would be consistent with the addition of a chlorine atom and so could potentially represent product, however no associated ion peak was present on the mass chromatogram and no mass of *m/z* 347 (M+Cl) could be seen. If this second peak does represent the product then a lack of associated mass could be explained by poor ionisation conditions. Another ionisation method such as APCI (atmospheric pressure chemical ionisation) may help to detect the compound responsible for this second PDA peak.

Analysis of the assay with the alternative substrate analogue, compound **38** (4-(2-amino-phenyl)-pyrrole), also showed two additional peaks in the PDA analysis at 5.1 and 5.9 minutes in addition to the NADH peak (figure 6.10B). The second of these peaks was also present in the total ion chromatogram (figure 6.10B) at a retention time of 6 minutes. This peak corresponded to a mass of *m/z* 205 (figure 6.10 C2) which is the same as the detected mass for the starting material. The 5.1 minute peak on the PDA trace was represented by a mass of 239 (figure 6.10B). This mass is +34 of 205 and could therefore correspond to the chlorination product of the ion peak at 6 minutes. However this peak is very small and occurs at an earlier retention time than the starting material making it extremely unlikely that this is the correct product. Targeted analysis at *m/z* 158 and 193 (which represent the correct molecular mass of compound **38** and its potential monochlorinated product) showed no corresponding ions.

Due to the uncertainty of the LCMS data collected the standard solutions of all three compounds and samples of the three reaction assays were sent to the John Innes Centre (Norwich) Metabolomics facility for higher resolution LC-MS analysis.

6.2.3 John Innes Centre (JIC) Metabolomics LCMS Analysis

LCMS analysis at the John Innes Centre (JIC) was performed by Dr Lionel Hill. The standard samples of the compounds were analysed at 200 µM and the reaction mixtures were diluted fourfold prior to analysis. Samples were run on a Surveyor hplc attached to a DecaXPplus ion trap MS. Separation was on a 150x2mm 4µ PolarRP column using a gradient of methanol versus 0.1% formic acid in water, using the method LC_JIC_Method1 at 0.3 ml min^{-1} (see materials and methods for details). Detection was by UV (200-600nm) and positive mode electrospray MS.

LCMS analysis of the aminated analogue, **38**, demonstrated a major peak at around 13 minutes on both the UV and MS base peak chromatograms (**A** and **B**, figure 6.11). This peak is associated with a mass of *m/z* 205 (**C**, figure 6.11), which is identical to the mass detected for this compound with the LCMS instruments at UEA. The extra mass of +46 (relative to the hydrogen adduct of the expected mass *m/z* 159), is the mass of formic acid, which both the UEA and JIC chromatography was run with. Therefore it was proposed that this mass could be accounted for by the incorporation of formate with the product without condensation, although this was deemed extremely unlikely. Re-running the LCMS in the absence of formic acid retained this 205 mass, strongly indicating that this result is genuine. MS2 analysis of this peak failed to show the appearance of any new mass peaks (**C**, figure 6.11). There was an additional smaller peak on the UV and base peak chromatogram at 18.1 min that corresponded to *m/z* 405, which does not correspond to any expected masses. A further peak also appeared on the MS base chromatogram at 19.2 minutes but this was associated with *m/z* of 367 and so also did not correspond to any expected mass value. The isotope pattern on the parent ion (**D**, figure 6.11) suggests the presence of boron in this compound. Boronic acids were used during the synthesis of these compounds (figure 6.11) and perhaps this isotope pattern indicates the presence of contamination from this process. None of the starting materials or reagents in the synthesis have a mass of 204 however so the presence of boron cannot currently be explained without further investigation.

The LCMS analysis of the assay mixture detected two possible small UV peaks at the retention time of the starting material (approximately 12-13 minutes) (**A**, figure 6.12) but no obvious related peaks were visible on the MS base chromatogram. Small peaks in this time range were also visible in the

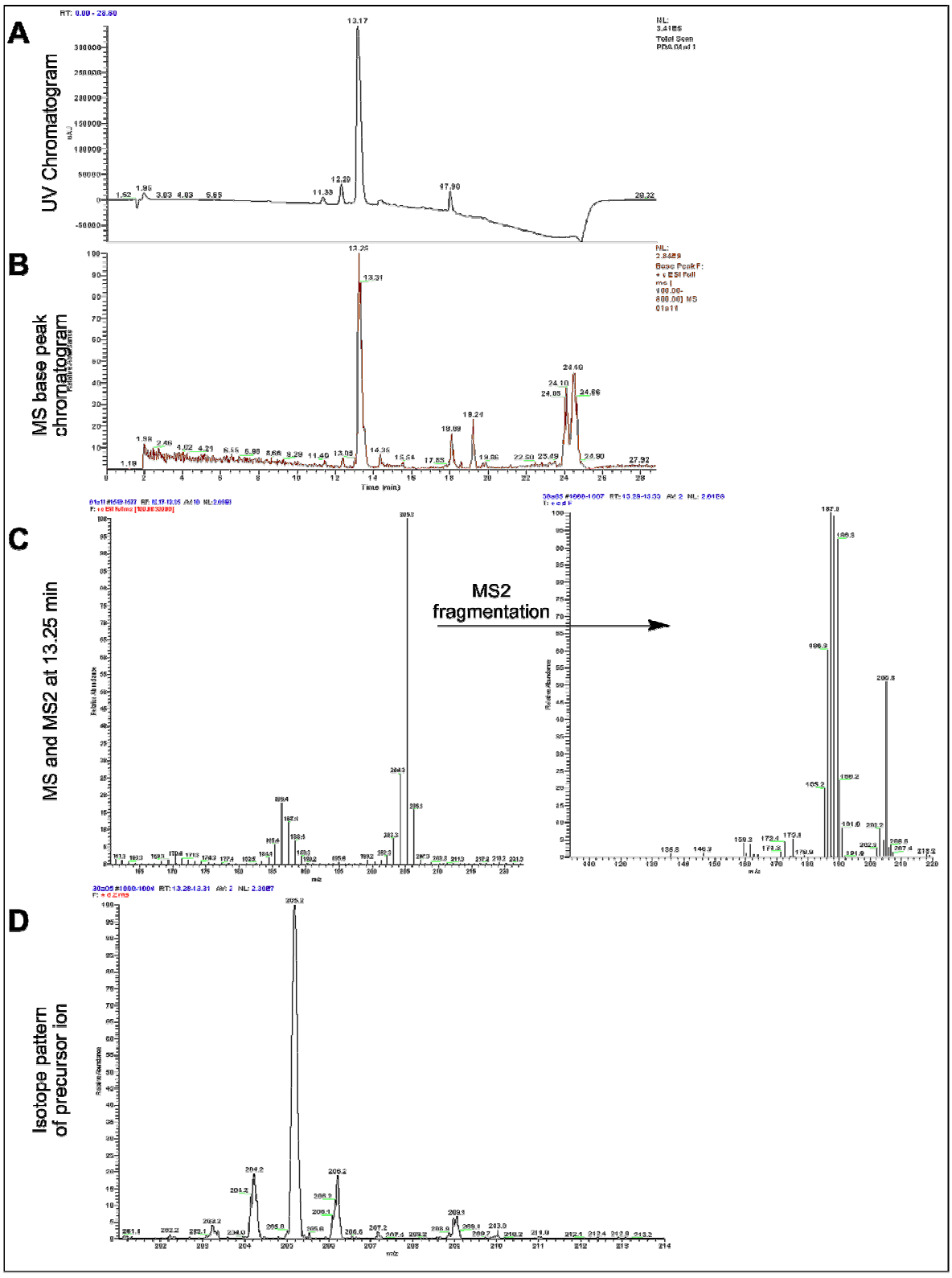


Figure 6.11: JIC Mass spectrometry analysis of compound 38 200 µM standard. (A) UV chromatogram. **(B)** MS base peak chromatogram shows clear single peak at 13.25 minutes. **(C)** MS and MS2 analysis of 13.25 peak from the base chromatogram assigns a *m/z* value of 205 for this compound. MS2 fragmentation did not identify any new peaks **(D)** Isotope pattern of the 205 precursor ion found at 13.25 minutes suggests presence of boron.

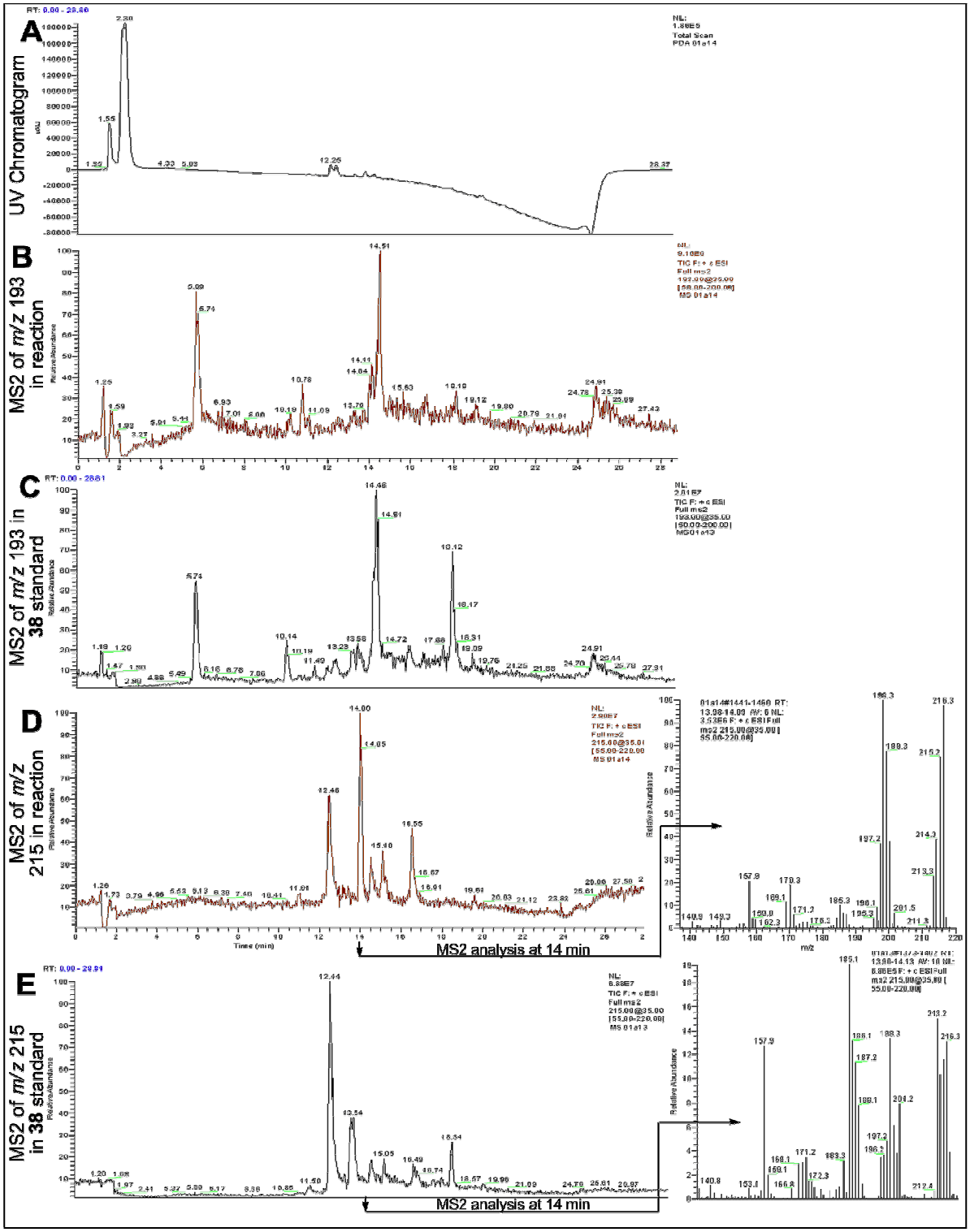


Figure 6.12: JIC Mass spectrometry analysis of compound 38 PrnC reaction mixture. (A) UV chromatogram. No clear starting material can be seen. **(B)** Abundance of *m/z* 193 in the reaction mix (M+H of the predicted product). **(C)** Abundance of *m/z* 193 in the starting material standard. **(D)** Abundance of *m/z* 215 in the reaction mix (M+Na of the predicted product) and fragmentation MS2 analysis of the peak at 14 minutes. **(E)** Abundance of *m/z* 215 in the starting material standard and fragmentation MS2 analysis of the peak at 14 minutes.

standard sample of the starting material so can probably be discounted as relating to either starting material or product. Targeted MS2 analysis showed peaks for masses corresponding to the hydrogen and sodium adducts of the possible product (*m/z* of 193 and 215 respectively) (**B** and **D**, figure 6.12). These peaks however were mostly discounted as insignificant as a similar targeted MS2 for the same masses in the starting material standard sample also show these ion peaks at the same retention times (**C**, **E**, figure 6.12). It is possible that the 14 minute peak in the chromatogram targeting the sodium adduct of the product (**D**, figure 6.12) represents a new product as the intensity is much higher in the reaction mix than the starting material (**E**, figure 6.12), but the MS2 spectrum at 14 minutes in both the reaction mix and the starting material show the same masses (**D**, **E**, figure 6.12), which suggests no new product formation.

The analysis of the standard sample of the final substrate analogue **39** (4-(2-nitro-phenyl)-pyrrole) showed a clear peak on the UV chromatogram at 16.57 min (**A**, figure 6.13) which was consistent with a clear peak on the MS base peak chromatogram that corresponded to an *m/z* of 189 (**B** and **C**, figure 6.13). This is the correct mass of the expected hydrogen adduct of the starting material (see scheme 6.2). The UEA in-house LCMS had previously failed to detect this mass when an identical sample was analysed, instead three mass peaks of 313, 335 and 350 were detected (figure 6.9). These masses could not be explained at the time but the fact that both the high resolution Swansea analysis and now the JIC analysis detected the correct compound strongly infers that the mass detected at UEA was an artefact caused by perhaps ion addition or dimerisation and that the specific ionisation conditions could be to blame for the different results. Even with this new information the formation of the three UEA peaks could not be explained. The *m/z* 313, 335 and 350 peaks represented additions relative to the known *m/z* 193 starting material of +124, +146 and +161. Relative to each other the peaks are *m/z* 313, *m/z* 313+22 and *m/z* 313+37. Again this pattern of ion additions could not be rationalised. Since the JIC LCMS analysis had confirmed the correct mass the reaction product was studied with increased interest.

The UV analysis of the reaction mixture showed very little sign of starting material or product (**A**, figure 6.14). There was a small peak visible at 12.52 minutes but this was also visible in the starting material and also in most of the UV chromatograms of the other substrate analogue as well. Similarly the MS base-peak also showed very little activity apart from the usual broad peak at 2.4 minutes consistent with the hydrogen adduct of the NADH cofactor (*m/z* 664). Targeted MS2 analysis was conducted looking for the starting material (*m/z* 189) and the hydrogen and sodium adducts of the expected chlorinated product (*m/z* 223 and 245 respectively). Very small amounts of the starting material were found (**A**, figure 6.15) and no significant evidence of the hydrogen adduct

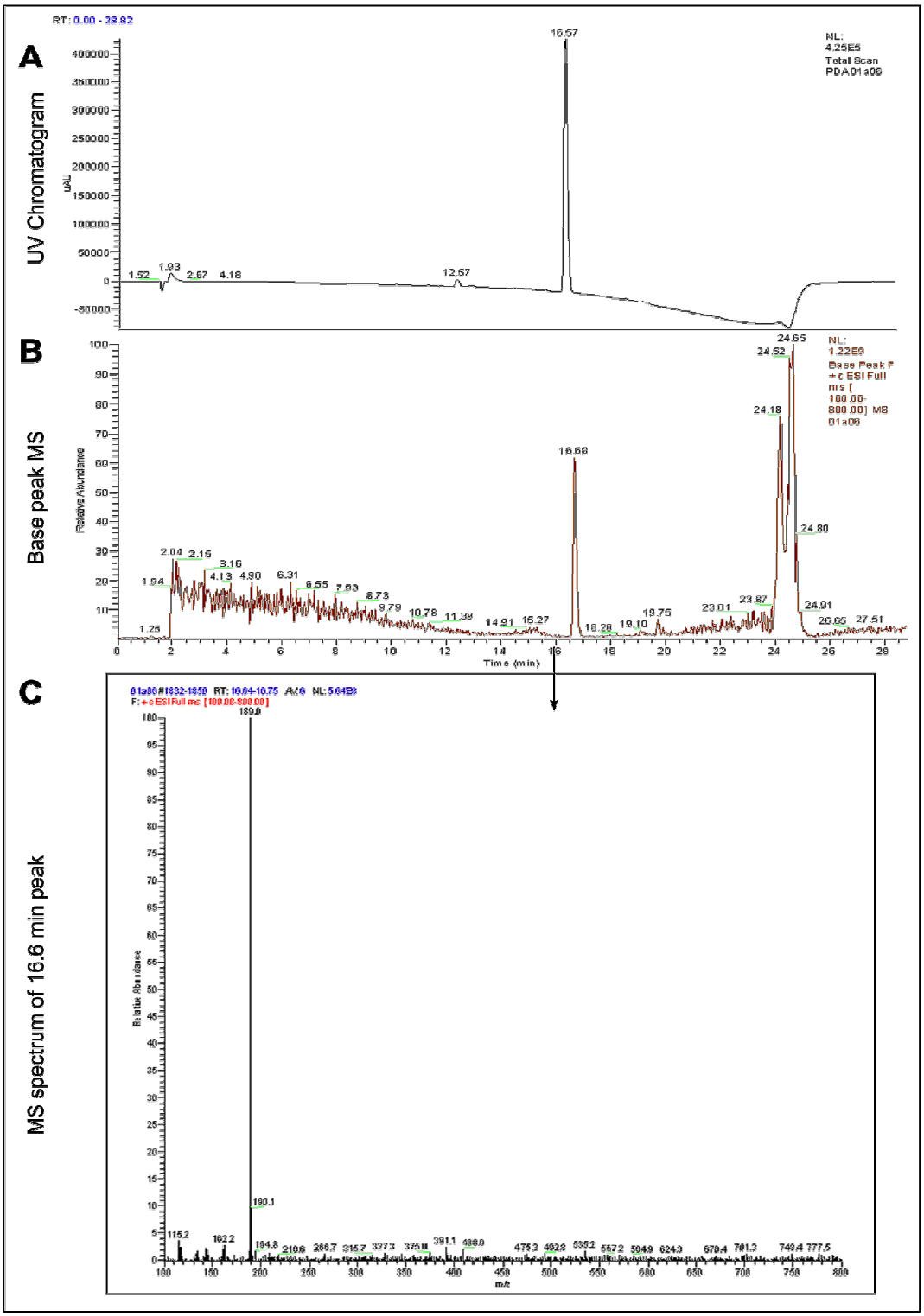


Figure 6.13: JIC Mass spectrometry analysis of compound 39 200 µM standard. (A) UV chromatogram. **(B)** MS base peak chromatogram shows single ion peak at 16.57 min. **(C)** MS analysis of the 16.57 peak from the base chromatogram shows ion corresponding to *m/z* 189.

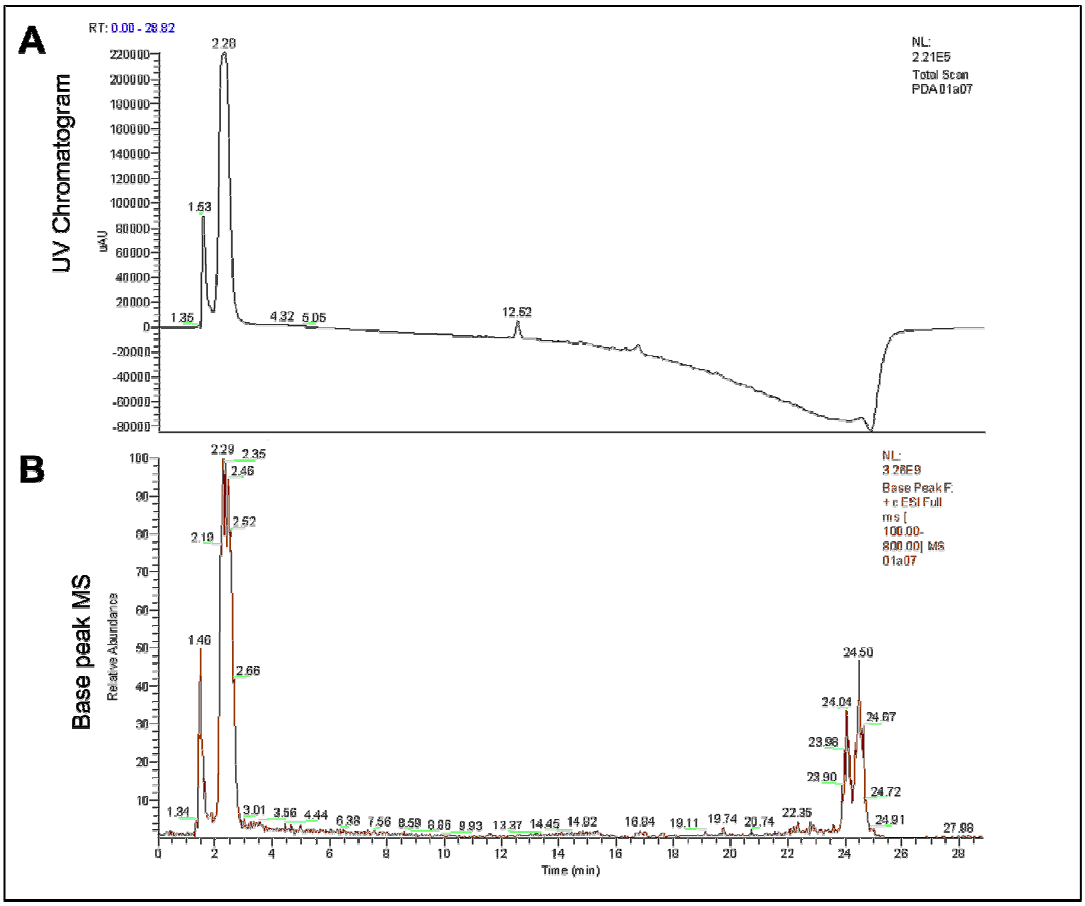


Figure 6.14: JIC Mass spectrometry analysis of compound 39 PrnC reaction mixture. (A) UV chromatogram. **(B)** MS base-peak chromatogram. Neither the UV chromatogram or the MS base peak chromatogram showed any evidence of product or starting material.

of a chlorinated product could be found (**B**, figure 6.15). The targeted MS2 for the sodium adduct did however detect a small peak at 17.16 minutes (**C**, figure 6.15), but this was shown to fragment to *m/z* 226 which also appears in the starting material so is unlikely to be evidence of chlorinated product (**D**, figure 6.15).

Overall no evidence of product was seen in any of the assays analysed. This is counter intuitive to the earlier HPLC results which clearly seemed to indicate new compound formation with the addition of PrnC (figure 6.9). Of the two compounds tested the most convincing evidence of enzyme activity was with the nitro-compound **39**. Although the in-house LCMS analysis failed to detect the 189 *m/z* mass of the correct compound (detected three masses of *m/z* 313, 335 and 350 instead) the higher resolution analysis conducted at the John Innes Centre did detect the correct mass suggesting that the detection method used at UEA was to blame. Furthermore the LCMS PDA data of the assay analysed at UEA did show a secondary peak forming at a later retention time than the starting material (figure 6.10A) which strongly suggested chlorination activity. This peak however was not

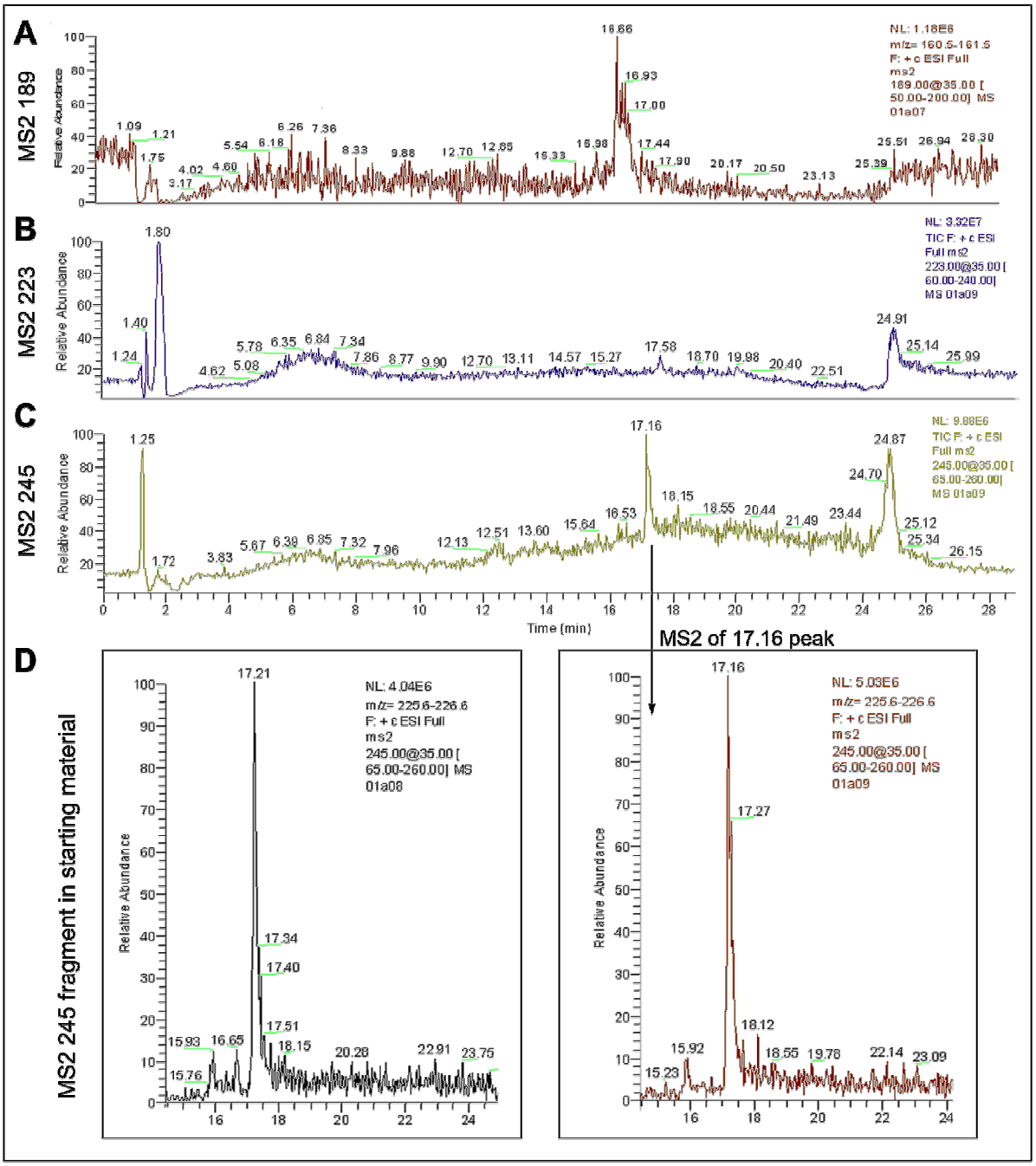


Figure 6.15: JIC Mass spectrometry analysis of compound 39 PrnC reaction mixture. (A) Abundance of *m/z* 189 starting material in the reaction mix. **(B)** Abundance of *m/z* 223 in the reaction mix (M+H of the predicted chlorinated product). **(C)** Abundance of *m/z* 245 in the reaction mix (M+Na of the predicted chlorinated product) and the MS2 fragmentation analysis of this peak to *m/z* 226. **(D)** MS2 showing abundance of *m/z* 245 to 226 in the starting material sample.

associated with any detected ion suggesting again that the ionisation method used was not appropriate.

6.2.4 Repeats of PrnC assays with substrate analogue 39

Due to this tentative evidence of enzyme activity the nitro compound **39** (4-(2-nitro-phenyl)-pyrrole) was added to a triplicate set of PrnC assays together with no PrnC, no PrnF, no flavin, no NADH and no substrate controls. The triplicates were prepared separately to ensure no mistakes were made in preparation. The assays were again analysed by HPLC using method LC_gradTrpMeOH.

This new set of PrnC assays generated a few expected results; the NADH negative control (**D**, figure 6.16) showed a clear single peak at 12.5 minutes which was consistent with the peak of the starting material (labelled as peak **a** on figure 6.16) and the no substrate control showed a broad initial peak at 2.5 minutes consistent with NADH and no 12.5 minute substrate peak (**B**, figure 6.16). The PrnC negative control showed peaks at 2.5 and 12.5 minutes consistent with the NADH and un-reacted starting material, although several smaller peaks were also visible just prior to the substrate peak at approximately 12 minutes (labelled as peaks **b** and **c** on **A**, figure 6.16). These peaks were not visible in the no-substrate control and may represent a decomposition product suggesting the substrate may not be stable under the assay conditions. The reaction mixture showed a single peak of the starting material at 12.5 minutes and suggested no evidence of product formation, which was contrary to the earlier assays which did show new peak formation. However a strange result emerged with the negative flavin and flavin reductase assays; these two controls clearly showed the formation of two major peaks in the region of the substrate although the retention time relative to the PrnC negative control is shifted downstream slightly, giving them a retention time more similar to the two postulated decomposition peaks (**b** and **c**) identified in the no PrnC control. The flavin reductase negative control showed a major peak at around 12 minutes (peak **b**) and a minor peak at 11.5 minutes (peak **c**) (**B**, figure 6.16). The $FADH_2$ negative control (**C**, figure 6.16) showed two peaks of identical height at 11.5 and 12 minutes (**c** and **b** respectively). These peaks and their retention times matched very closely the multiple peaks that were detected in the first set of PrnC assays conducted with this substrate (figure 6.6) but their presence in the flavin negative controls and not in the actual assay mixture is unexpected, as is the shift in retention time which would normally indicate that the compounds had become more polar, not something that would be expected with a chlorinated product. A possible peak can be seen in both negative assays at around 12.5 minutes, suggesting that the two new peaks may indeed relate to possible decomposition products and not

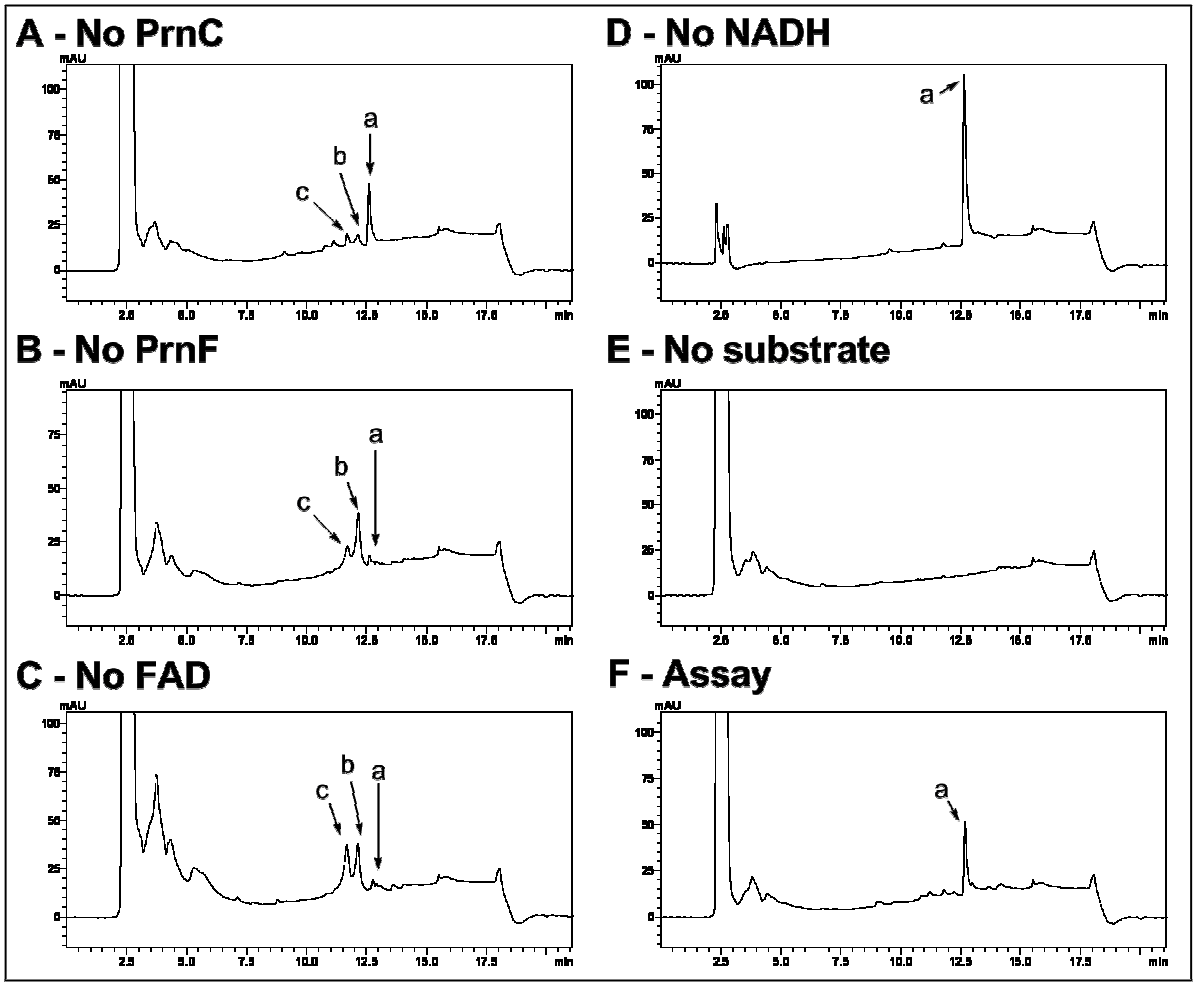


Figure 6.16: HPLC chromatograms of PrnC assays and controls conducted with compound 39. (A) No enzyme (PrnC) control. The starting material (labelled peak a) can be seen at 12.5 minutes, two smaller peaks appear earlier and may represent decomposition products (peaks b and c). **(B)** No flavin reductase (PrnF) control. The same three peaks can be seen as were observed in the control. However the ratio between them is different. **(C)** No additional flavin cofactor control. Peaks a,b and c are again visible. **(D)** No additional NADH cofactor control. Only the single peak corresponding to starting material can be seen. **(E)** No substrate control. **(F)** Complete assay mixture. Only starting material can be seen.

the substrate itself or any chlorinated product which would be expected to have a retention time later than 12.5 minutes.

The results of these assays are inconclusive with possible decomposition of the substrate occurring and appearance of major peaks in two of the negative controls and none in the actual reaction mixture. These results were reproducible throughout all of the triplicates which were prepared at different times, and thus the possibility of sample contamination was ruled out indicating that this is

a genuine result with this synthetic compound. The lack of apparent activity except in the absence of reduced flavin is very peculiar as PrnC is a $FADH_2$ dependent enzyme. Previous assays did show

multiple peak formation in the presence of $FADH_2$ and PrnF. These results suggest that the addition of reduced flavin or reducing the flavin *in situ* is perhaps inhibiting the reaction, or that the reduced flavin reacts with the substrate leading to its decomposition. Further study is needed to establish what it occuring. In an attempt to understand the origin of the possible product or decomposition peaks (**b** and **c**), LCMS analysis was conducted on all of the samples.

As with the previous set of LCMS analysis conducted at UEA no sign could be found of the expected starting material (*m/z* 189) or that of the chlorinated product (*m/z* 223). As before the major ion detected was *m/z* 313 with minor peaks of *m/z* 335, 350 and 376 (**B**, figure 6.17). However the quantity of these ions was not consistent throughout the different samples with a relatively large abundance in the negative PrnF and $FADH_2$ controls (**A** and **B**, figure 6.17), a much smaller amount in the reaction mixture (**D**, figure 6.17) and no trace of starting material in the NADH negative control (**C**, figure 6.17). This inconsistency in detection of the starting material from samples that should contain equal amounts, together with the incorrect detected mass does indicate potential problems with the ionisation conditions. The PrnF negative control also showed the potential starting material eluting from the reverse phase column at around 2 minutes and not the 7 minutes observed in the other compounds. This was probably due to the reverse phase column not being probably equilibrated into the starting conditions of 5% methanol from the 50% storage conditions.

Despite the incorrect mass being detected it was known from the JIC analysis that the correct mass is present inside the samples but was not being properly detected by the instrument at UEA. Therefore an additional ion was targeted that represented the chlorinated product of the 313 ion that was being detected (*m/z* 347), just in case this was a dimerisation event or an alternative ion adduct. In most of the samples there was no detection of this ion either, however the $FADH_2$ negative control showed a peak on the total ion chromatogram at approximately 8 minutes when the *m/z* 347 was targeted (labelled as peak 2 on **C**, figure 6.17), this shift in retention time would be consistent with chlorinated product. The MS spectrum (**2C,** figure 6.17) gave a new major peak at *m/z* 347 which could possibly correspond to the starting material plus the 34 mass units of chlorine. The isotope pattern of this ion also strongly indicates the presence of chlorine with a +36 peak appearing to be occurring at roughly 35% abundance that does not appear in the isotope pattern of the starting material. Interestingly there were also a number of smaller peaks of *m/z* 369 and 384. These two

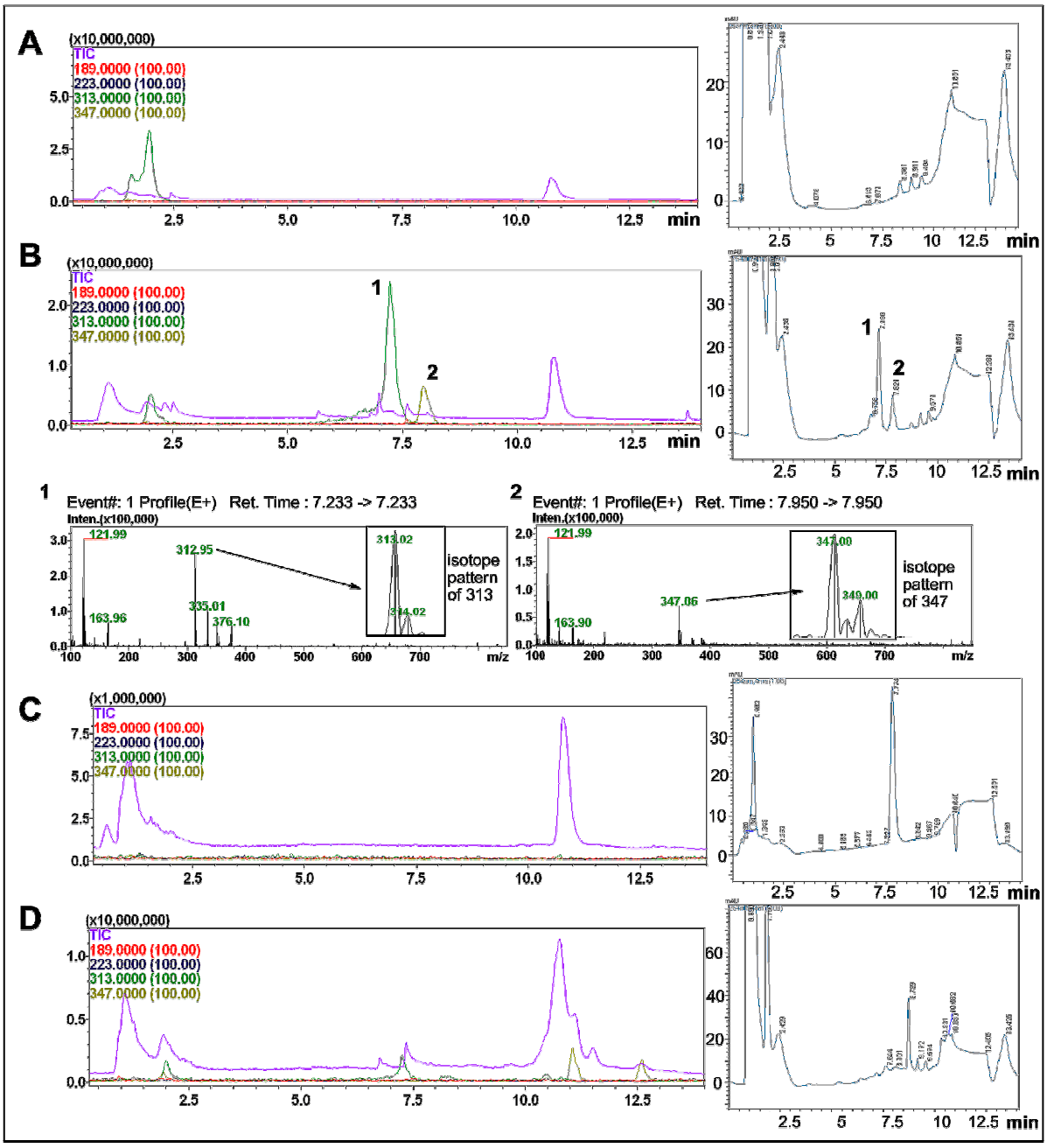


Figure 6.17 UEA LCMS analysis of the repeat PrnC assays with substrate 39. Total ion chromatograms of the assays are shown with the abundance of the known starting material *m/z* 189 and possible chlorinated product *m/z* 223 are shown amplified 100 times. The abundance of *m/z* 313 previously detected on this instrument with the standard has also been amplified and the possible chlorinated product of this ion has also been amplified (*m/z* 347). **(A)** Total ion and PDA chromatogram for the flavin reductase negative control. **(B)**Total ion and PDA chromatograms of the $FADH_2$ negative control and individual MS spectrum for the peaks labelled 1 and 2 including the

isotope pattern for each ion. **(C)**Total ion and PDA chromatograms of the negative NADH control. **(D)** Total ion and PDA chromatograms of the reaction mixture. The large TIC peak at 10.5-11 minutes was contamination within the instrument and was present in the blank.

additional masses represent the addition of 34 mass units to the minor peaks *m/z* 335 and 350 present in the starting material (**1C**, figure 6.17). These additional peaks also exhibit an isotope pattern indicative of chlorine. This strongly suggests that chlorinated product may have been formed and that this product may exhibit a similar adduct profile or be dimerising in the same way as the starting material. This evidence is very preliminary however and unconfirmed at this stage. The really interesting thing is why this compound would be appearing in the negative flavin control and not in the actual assay mix. The emergence of this peak fits with the HPLC data where the new peaks appeared only in the absence of reduced FADH. Interestingly, although there were also new peaks in the negative PrnF control when analysed by the HPLC the *m/z* 347 ion did not appear with the LCMS analysis, although the quantity of peak c was much greater in the $FADH_2$ control than in the PrnF control suggesting that the chlorinated product may have been below detectable levels in the PrnF control when analysed by the LCMS.

6.3 Conclusions and future work

The preliminary results of the activity assays of PrnC is not very encouraging but a major contributing factor is the failure to obtain the natural substrate **35** and the ionisation and detection problems experienced with the other substrate analogues. The only potential positive result was with compound **39** and the tentative mass of chlorinated starting material detected by the LCMS in the $FADH_2$ negative control and this was based on the detected mass of *m/z* 313 and not of the actual mass of the starting material but the isotope pattern clearly indicated the incorporation of chlorine. To be able to fully test PrnC and study its enzyme kinetics and substrate flexibility, the natural substrate would have to be synthesised and confirmed to be of the correct structure, having this confirmed natural substrate would be essential to progress the work on this enzyme further. In addition to this the ionisation problems with the substrate analouges would have to be solved possibly with the utilisation of negative ion mode ESI or APCI (atmospheric pressure chemical ionisation) instead of the positive mode ESI method used for this work, however it should be noted that the JIC LCMS was run with positive mode ESI without any detection issues but optimisation of the UEA instrument will perhaps require different conditions and should be studied.

In addition, improvements perhaps can be made to the production of PrnC itself to obtain more purified protein. Many optimisation methods were trialled within this study on the enzyme

expressed within *E. coli* but even the optimised methods selected did not afford the purified protein at great yield. The closest known sequence homolog to PrnC is another flavin-dependent halogenase (from an *Actinomycete* sp.) known as HalA which shares 42% homology.[196] When this enzyme was isolated it was found to be produced in a largely insoluble form within *E. coli*. *Pseudomonas aureofaciens* CAN was found to be a more suitable alternative for the production of good levels of soluble protein. It may be worthwhile in the future to explore alternative production hosts for PrnC including species of *Pseudomonas* to try to optimise production of protein.

Despite the many setbacks and problems, significant progress has however been made towards the production and characterisation of PrnC during the course of this project. If future work focuses on improving the production of soluble protein and continuing to improve the synthesis of the natural substrate then the potential of PrnC as an adaptable pyrrole halogenase may be explored further.

Chapter 7:

Materials and Methods

7.1 General equipment

Microbial cultures were incubated in a Gallenkamp INR-200 orbital shaker incubator or an Innova 4300 incubator shaker (New Brunswick Scientific). For incubations below ambient temperature an Innova 42 refrigerated orbital shaker incubator was used (New Brunswick Scientific).pH measurements were taken using a Fisherbrand Hydrus 300 pH meter. Centrifugation was carried out using a Beckmann Avanti J-25 centrifuge fitted with a JLA 9.1000 rotor (16800g max), a JA-12 rotor (23200g max) or a JA-20 rotor (48400g max). Alternatively centrifugation was carried out with a Thermo Scientific IEC CL30R centrifuge fitted with a T41 swing-out rotor (3082g max) or a Fisher Scientific accuSpin microcentrifuge (16060g max). Biofilms were spin-coated using a Jouan C4.22 centrifuge fitted with a swinging bucket rotor. Cultures were sonicated using a Status US200 ultra-sonicator (Philip Harris Scientific) fitted with a Titanteller TT13 13 mm titanium flat tip (Bandelin). PCR reactions were carried out using a Techne TC-512 thermocycler. HPLC analysis was performed on a Shimadzu Prominence HPLC system. This was comprised off a degasser unit (DGU-20A3), a liquid chromatograph (LC-20AB) fitted with an autosampler (SIL-20A). Compounds were detected by a PDA diode array detector (SPD-M20A). Mass spectrometry was carried out using either a Shimadzu single quadrupole LC-mass spectrometer equipped with an electrospray ionisation source at the University of East Anglia or a ThermoFinnigan DecaXP$_{plus}$ ion trap and Surveyor HPLC system at the John Innes Centre (Norwich, UK). DNA sequencing was carried out by the Cambridge DNA Sequencing Facility (Cambridge University, UK). Optical densities and UV spectra were obtained with a Lambda Bio+ spectrophotometer (Perkin-Elmer). Freeze drying was performed on a MicroModulyo freeze drier (Thermo Savant).

7.2 General procedures

Chemicals, buffer and media components were purchased from BD Biosciences, Melford, ForMedium, Fisher Scientific, Sigma-Aldrich or Alfa-Aesar unless otherwise stated. For long term storage microorganisms were stored at -80°C as a 20% (v/v) glycerol stock in H_2O. Microorganisms were cultured under sterile conditions using a Faster BH-EN class II vertical laminar airflow cabinet or on the bench adjacent to a Bunsen flame. Materials and media were sterilised with a Boxer Benchtop Denley autoclave (121°C for 20 minutes at 1.3 bar) or were sterilised by filtration through a 0.2 µm membrane.

7.3 Media

Media	*Composition*	
Luria-Bertani (LB) broth	Tryptone Yeast extract NaCl H_2O	10 g 5 g 10 g Final volume 1 L
LB/2	LB media diluted by half into sterile distilled water	
M63 Minimal Media	KH_2PO_4 $(NH_4)_2SO_4$ Succinic acid Glucose $MgSO_4.7H_2O$ (50 mg/ml) $FeSO_4.2H_2O$ (50 mg/ml) H_2O	13.6 g 2.0 g 2.5 g 2.0 g 4 ml 0.01 ml Final volume to 1 L
ZY	Tryptone Yeast extract H_2O	10 g 5 g 925 ml
20X NPS	$(NH_4)_2SO_4$ KH_2PO_4 NaH_2PO_4 H_2O	6.6 g 13.6 g 14.2 g 90 ml pH adjusted to 6.75 with 10N NaOH
50X 5052	Glycerol Glucose α-Lactose H_2O	25 g 2.5 g 10 g 73 ml
ZYP-0.8G	ZY media 1M $MgSO_4$ 40% glucose 20x NPS media	93 ml 0.1 ml 2 ml 5 ml $MgSO_4$ added before 20X NPS to avoid precipitation
ZYP-5052	ZY media 1M $MgSO_4$ 50X 5052 20X NPS	372 ml 0.4 ml 8 ml 20 ml

Media recipes were prepared with distilled water and then autoclave sterilised. Antibiotic and mineral solutions were also prepared in distilled water and filter sterilised through a 0.2 µm membrane. Agar solid media was prepared by the addition of 1.5 % bacteriological agar prior to being autoclaved.

7.4 Buffers and stock solutions

Solution	*Function*	*Composition*
General Solutions		
2.5 mM dNTPs		25 µl of each 10 mM stock of dATP, dGTP, dTTP and dCTP added together to a 100 µl final volume

10% Ammonium persulfate		1g of ammonium persulfate in 10 ml of H_2O	
Ethidium Bromide		A 10mg/ml solution of EtBr in H_2O	
40% Glycerol (v/v)		200 ml of glycerol added to 500 ml of H_2O	
10 M KOH		112g of KOH in 200 ml of H_2O. Add KOH slowly while stirring in an ice bath.	
10x Ligation buffer		Tris Base $MgCl_2$ DTT ATP	300mM 100mM 100mM 10mM pH adjusted to 7.8 with HCl
10 M NaOH		80g of NaOH in 200 ml of H_2O. Add NaOH slowly while stirring in an ice bath.	
10% SDS (w/v)		Add 100g of SDS to 1000 ml of H_2O. Stir solution at 60°C until all solid has dissolved	
1M Tris-HCl Buffer		121.14g of Tris Base in 800 ml of H_2O. pH adjusted to required value with 37% HCl then volume topped up to 1000 ml. Different concentration Tris buffers made by adjusting mass of tris base added or by dilution of 1M stock.	
Solutions for Gel Electrophoresis			
50X TAE	Concentrated Electrophoresis running buffer	EDTA Glacial acetic acid Tris-base	0.05 M 1 M 2 M pH adjusted to 8.0 with HCl
TAE	Ready to use diluted Electrophoresis running buffer prepared by diluting 20 ml of 50X TAE into 1000 ml		
SDS-PAGE Solutions			
Coomassie Brilliant Blue staining solution	For visualising proteins on SDS-PAGE gels	Coomassie Brilliant Blue R-250 Methanol H_2O Glacial acetic acid	0.25 g 500 ml 400 ml 100 ml
Destaining Solution	For visualising proteins on SDS-PAGE gels	Methanol H_2O Glacial acetic acid	500 ml 400 ml 100 ml
Gel Drying Solution		Methanol Glycerol Acetic acid H_2O	400 ml 100 ml 75 ml 425 ml
12% Resolving Gel Solution	Acrylamide gel for separation of proteins. Makes 2 Gels	30% Acrylamide (+0.8% Bisacrylamide) 1.5 M Tris (pH 8.8) 10% SDS (w/v) H_2O Just before pouring: 10% Ammonium persulfate TEMED	4.0 ml 2.5 ml 0.1 ml 3.2 ml 0.1 ml 4 µl
SDS-PAGE Running Buffer		Glycine Tris Base SDS	250 mM 25 mM 0.1% (w/v)

5% Stacking Gel	Acrylamide gel for loading of protein samples. Makes 2 gels	30% Acrylamide (+0.8% Bisacrylamide) 1.5 M Tris (pH 6.8) 10% SDS (w/v) H_2O Just before pouring: 10% Ammonium persulfate TEMED	0.5 ml 0.38 ml 0.03 ml 2.1 ml 0.03 ml 4 µl
2x SDS Loading Dye	Loading Dye for Protein Sample	1M Tris-Cl pH 6.8 10% SDS Bromophenol blue Glycerol H_2O (1M Dithiothreitol The above solution was stored without DTT at room temperature. Before use a 4 ml aliquot was taken and 1 ml of 1M DTT added. This was split into 1 ml portions and stored at -20°C.	5 ml 20 ml 0.1 g 10 ml 5 ml 10 ml)
Solutions for Competent Cell Preparation			
KMES I	Competent cell Preparation solution	$CaCl_2$ KMES $MgCl_2$ $MnCl_2$ H_2O	60 mM 25 mM 5 mM 5 mM pH adjusted to 5.8 with KOH Final volume to 200 ml Autoclaved
KMES II	Competent cell storage solution	KMES storage buffer is prepared by the addition of 10% glycerol to KMES I	
Alkaline Lysis Solutions for DNA isolation			
Alkaline Lysis I	Re-suspension buffer	D-glucose monohydrate Tris-base EDTA	50 mM 25 mM 10 mM pH adjusted to 8.0 with HCl Autoclaved and stored at 4 °C
Alkaline Lysis II	Lysis Buffer	10 M NaOH 10% SDS (w/v)	0.2 ml 1 ml Solution should be freshly prepared and stored at room temperature
Alkaline Lysis III	Neutralisation Buffer	5M Potassium acetate Glacial Acetic Acid H_2O	60 ml 11.5 ml 28.5 ml Autoclaved and stored at 4 °C
EB Buffer	DNA Storage Buffer	10 mM Tris-HCl (pH 8.5)	
Chemical Cell Lysis Buffers			
Lysis Buffer 1	Chemical Lysis Buffer	Tris-Cl NaCl Glycerol Supplemented with 2 mg/ml lysozyme before use	50 mM 25mM 5% pH adjusted to 8.0 with HCl

Lysis Buffer 2	Chemical Lysis Buffer with EDTA	Tris-Cl NaCl EDTA Glycerol Supplemented with 2 mg/ml lysozyme before use	50 mM 25mM 2.5 mM 5% pH adjusted to 8.0 with HCl
NiNTA Purification Buffers			
NiNTA Lysis Buffer	Lysis Buffer	NaH_2PO_4 NaCl Imidazole	50 mM 300 mM 10 mM pH adjusted to 7.8 with 10M NaOH
NiNTA Wash Buffer	Wash Buffer	NaH_2PO_4 NaCl Imidazole	50 mM 300 mM 40 mM pH adjusted to 7.8 with 10M NaOH
NiNTA Elution Buffer	Elution Buffer	NaH_2PO_4 NaCl Imidazole	50 mM 300 mM 300 mM pH adjusted to 7.8 with 10M NaOH
Protein Storage Buffer	Long term protein storage buffer	Hepes NaCl Glycerol	20 mM 50 mM 10 % pH adjusted to 7.4 with 10M NaOH
Ni-NTA Purification Optimisation Buffers			
Reduced NaCl Lysis Buffer		NaH_2PO_4 NaCl Imidazole	50 mM 150 mM 10 mM pH adjusted to 7.8 with 10M NaOH
Reduced Imidazole Lysis Buffer		NaH_2PO_4 NaCl	50 mM 150 mM pH adjusted to 7.8 with 10M NaOH
Tryptophan Synthase Biotransformation Buffers			
Tryptophan Synthase Lysis Buffer	Generation of tryptophan synthase cell free lysates	KH_2PO_4 EDTA PLP	100 mM 5 mM 0.1 mM pH adjusted to 7.8 with 10M KOH
Potassium Phosphate Reaction Buffer	Tryptophan synthase biotransformation buffer	KH_2PO_4 L-serine PLP	100 mM 7 mM 0.2 mM pH adjusted to 7.0 with 10M KOH
Low Phosphate Reaction Buffer		KH_2PO_4 L-serine PLP	10 mM 7 mM 0.3 mM pH adjusted to 7.0 or 7.8 with 10M KOH

7.5 Antibiotics

Antibiotic	*Stock Concentration*	*Culture Concentrations*
Ampicillin sodium salt (Amp)	100 mg ml^{-1} in H_2O	100 µg ml^{-1}
Kanamycin sulfate (Kan)	50 mg ml^{-1} in H_2O	50 µg ml^{-1}
Chloramphenicol (Clm)	25 mg ml^{-1} in EtOH	25 µg ml^{-1}

Prior to making solutions kanamycin sulfate was stored at room temperature, ampicillin sodium salt at 5°C and chloramphenicol was stored at -20°C. Aqueous stock solutions of antibiotics were sterilised by filtration through a 0.2 µm membrane. The ethanolic chloramphenicol solution was used without further sterilisation. All stock solutions were stored at -20°C. For most applications, unless otherwise stated, the stock solutions were diluted 1000x into culture medium to give the final concentration of antibiotic.

7.6 Plasmids

Plasmid	*Description; Relevant Genotype*	*Source*
pBAD/HisA	4.1 kb vector for protein expression in *E. coli* (with optional C-terminal His6 tag) araBAD promoter; *araC*; Amp^R	Invitrogen
pET21a(+)	5.4 kb vector (optional C-terminal His6 tag) for protein expression in *E. coli* T7 promoter; l*acI*; Amp^R	Novagen
pET28a(+)	5.3 kb vector (N-terminal His6 tag with optional C-tag) for protein expression in *E. coli* T7 promoter; l*acI*; Kan^R	Novagen
pNYCOMPS-LIC-FH10T ccdB	6.8 kb vector (with N-terminal flag epitope, His10 tag) for protein expression T7 promoter; FH10T; *ccdB;* Clm^R; Kan^R	DNASU Plasmid Repository
pPro24-gfp	Propionate induced 5.7 kb vector for expression of GFP fusion proteins prpBCDE promoter; *prpR*; *gfpuv*; Amp^R	Addgene
pGro7	L-arabinose induced vector containing groEL-groES chaperone araBAD promoter; *araC*; *groES*; *groEL*; Clm^R	Takara
pJKE7	L-arabinose induced vector containing dnaK-dnaJ-grpE chaperone araBAD promoter; *araC*; *dnaK*; *dnaJ*; *grpE*; Clm^R	Takara

pTF16	L-arabinose induced vector containing tig chaperone araBAD promoter; *araC; tig*; Clm^R	Takara
pSTB7	Constituently active vector expressing tryptophan synthase from *Salmonella typhimurium* Trp promoter; *trpABC*; Amp^R	Isolated from culture from American Type Culture Collection (ATCC 37845)
pMW03	1.7 kb BglII/MfeI restricted PCR fragment (Primers PrnC-bgl, PrnC-R1)containing *prnC* gene cloned into BglII/EcoRI restricted pBAD-HisA	This study
pMW04	1.6 kb BglII/MfeI restricted PCR fragment (Primers PrnC-bgl, PrnC-R2)containing C-terminally truncated(Δ50bp) *prnC* gene cloned into BglII/EcoRI restricted pBAD-HisA	This study
pMW05	1.6 kb BglII/MfeI restricted PCR fragment (Primers PrnC-bgl, PrnC-R3)containing C-terminally truncated (Δ84bp) *prnC* gene cloned into BglII/EcoRI restricted pBAD-HisA	This study
pMW06	1.7 kb NdeI/EcoRI restricted PCR fragment (Primers PrnC-1, PrnC-R1) containing *prnC* gene cloned into the corresponding site on pET28a(+)	This study
pMW12	Quikchange mutagenesis on pMW06 to increase size of N-terminal His6 tag to His8 tag (mutagenesis primers pET28a-8xHis-F1, pET28a-8xHis-R1)	This study
pMW17	Fragment containing tryptophan synthase genes *trpA* and *trpB* restricted out of pSTB7 (HindIII/EcoRI) and ligated into corresponding restriction sites of pET28a(+)	This study
pMW18	1194 bp HindIII/XhoI restricted PCR fragment (Primers TRPSYN-B-F1, TRPSYN-B-R1) containing tryptophan synthase *trpB* gene cloned into corresponding sites of pET21a(+)	This study
pMW19	806 bp NdeI/XhoI restricted PCR fragment (Primers TRPSYN-A-F1, TRPSYN-A-R1) containing tryptophan synthase *trpA* gene cloned into corresponding sites of pET28a(+)	This study
pMW21	806 bp NcoI/XhoI restricted PCR fragment (Primers TRPSYN-A-F2, TRPSYN-A-R1) containing tryptophan synthase *trpA* gene cloned into pET28a(+) before His6 tag.	This study
pMW22	PCR amplified fragment containing *prnC* (Primers PrnC_LIC-F1, PrnC_LIC-R1a) was cloned into pNYCOMPS-LIC-FH10T-ccdB using ligation independent cloning methodology.	This study
pSG22	*prnA* tryptophan-7-halogenase gene PCR amplified from *Pseudomonas fluorescens Pf-5* gDNA cloned into pET21a(+)	SG*
pSG24	*prnF* flavin reductase gene PCR amplified from *Pseudomonas fluorescens Pf-5* gDNA cloned into pET21a(+)	SG*
pSG28	*prnA* tryptophan-7-halogenase gene PCR amplified from *Pseudomonas fluorescens Pf-5* gDNA cloned into pET28a(+)	SG*
pSG49	*prnA* gene PCR amplified from *Pseudomonas fluorescens Pf-5* gDNA cloned into pPro(s)-GFP vector to create PrnA-GFP fusion protein	SG*

*These plasmids obtained for this study were designed by my colleague Dr Sabine Grüschow for a separate but related project

7.7 Micro-organisms

Strain	***Description; Relevant Genotype***	***Source***
Pseudomonas sp.		
Pseudomonas fluorescens Pf-5	Pyrrolnitrin Producer. Genome template for pyrrolnitrin biosynthetic genes.	
Escherichia coli sp.		
XL1-Blue MR	Used for general cloning and DNA propagation. Δ(*mcr*A)183 Δ(*mcr*CB-*hsd*SMR-*mrr*)173 *end*A1 *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac*.	Stratagene
BL21 (DE3)	Used for protein production under the control of most promoters F⁻*omp*T *hsdS*$_B$($r_B^-m_B^-$) *gal dcm* (DE3)	Novagen
DH10B	Used for general cloning and DNA propagation. Also used to express proteins controlled under the araBAD promoter. F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ-*rps*L *nup*G	Invitrogen
PHL644	Biofilm forming strain. Contains ompR mutation that promotes curli formation resulting in biofilm formation. *ara*D139 D(*arg*F-*lac*)U169 *rps*L150 *rel*A1 *flb*B5301 *deo*C1 *pts*F25 *rbs*R *mal*A-*kan omp*R234	Vidal *et al.* (1998)[42]
CB149	Tryptophan synthase host strain containing pSTB7 plasmid Δ*trp* EDCBA2 *hsdR*514 (rk- mk+) *sup*E44 *sup*F58 *lac*Y1 *lac*U169 *gal*K2 *gal*T22.	ATCC 37845
PHL644 MW-002	Biofilm strain containing pSTB7 plasmid for expression of tryptophan synthase	This study
XL1-Blue MR MW-005	Used for storage and propagation of pMW03 plasmid	This study
XL1-Blue MR MW-006	Used for storage and propagation of pMW04 plasmid	This study
XL1-Blue MR MW-007	Used for storage and propagation of pMW05 plasmid	This study
BL21 (DE3) MW-008	Expression strain for pMW03 PrnC halogenase under araBAD promoter	This study
BL21 (DE3) MW-009	Expression strain for pMW04 truncated PrnC haogenase under araBAD promoter	This study
BL21 (DE3) MW-010	Expression strain for pMW05 truncated PrnC halogenase under araBAD promoter	This study
BL21 (DE3) MW-012	Expression strain for pMW06. PrnC halogenase under T7 promoter	This study
BL21 (DE3) MW-016	Expression strain for pMW12. PrnC halogenase fusion protein mutant with His8 tag	This study
DH10B MW-021	DNA propagation and protein expression strain for pMW03. PrnC under araBAD promoter.	This study
BL21 (DE3) MW-022	Expression strain for pMW12 co-expressed with pGro7 chaperone protein	This study

BL21 (DE3) MW-023	Expression strain for pMW12 co-expressed with pJKE7 chaperone proteins	This study
BL21 (DE3) MW-024	Expression strain for pMW12 co-expressed with pTf16 chaperone protein	This study
DH10B MW-029	Cloning and DNA propagation strain for pMW17. Tryptophan synthase genes restricted out of pSTB7	This study
BL21 (DE3) MW-030	Expression strain for pMW17. Tryptophan synthase genes restricted out of pSTB7	This study
BL21 (DE3) MW-031	Expression strain for pMW18	This study
DH10B MW-032	Cloning and DNA propagation strain for pMW19	This study
DH10B MW-033	Cloning and DNA propagation strain for pMW20	This study
BL21 (DE3) MW-034	Expression strain for pMW18 (tryptophan synthase β subunit) and pMW19 (α subunit tagged)	This study
DH10B MW-035	Cloning and DNA propagation strain for pMW21	This study
BL21 (DE3) MW-036	Expression strain containing pMW18 and pMW21 for co-expression of both α and β subunits of tryptophan synthase	This study
DH10B MW-037	Cloning and DNA propagation strain for pMW22	This study
Bl21 (DE3) MW-038	Expression strain for pMW22. PrnC in LIC vector	This study
BL21 (DE3) MW-039	Expression strain for pSG22. PrnA	This study
PHL644 MW-040	Strain for expressing PrnA in a biofilm	This study
BL21 (DE3) MW-041	Strain for expressing PrnA-gfp fusion protein	This study
PHL644 MW-042	Strain for expressing pSG49, PrnA-gfp fusion protein in a biofilm	This study
CB149 MW-043	Host strain for pSTB7 tryptophan synthase genes	ATCC 37845
BL21 (DE3) RG-5066	Strain for expressing pSG24 PrnF flavin reductase	SG*
BL21 (DE3) RG-5076	Strain for expressing pSG28 PrnA halogenase	SG*

*These strains obtained for this study were constructed by my colleague Dr Sabine Grüschow for a separate but related project

7.8 Primers

Primer	***Sequence***	***Restriction site***	***Source***
PrnC-Bgl	5'-TTT*AGATCT*TGACTCAGAAGAGCCCCGC	BglII	This study
PrnC-R1	5'-TT*CAATTG*TTACTTCTTCAGAGCCAAGCCG	MfeI	This study
PrnC-R2	5'-TT*CAATTG*TTATCACGCGAACATCGGGTTGTAG	MfeI	This study
PrnC-R3	5'-TT*CAATTG*TTATCACACCTCTGCCGGCGCG	MfeI	This study
PrnC-1	5'-AGATTATC*CATATG*ACTCAGAAGAGCCCCGCGAAC	NdeI	SG*
pET28a-8xHis-F1	5'- CAGCAGCCATCATCATCACCATCATCATCACAGCAGC		This study
pET28a-8xHis-R1	5'- GCTGCTGTGATGATGATGGTGATGATGATGGCTGCTG		This study

TRPSYN-B-F1	5'- TTT*AAGCTT*ATGACAACACTTCTCAACCC	HindIII	This study
TRPSYN-B-R1	5'- TT*CTCGAG*GATTTCCCCTCGCGCTT	XhoI	This study
TRPSYN-A-F1	5'- TTTTTTT*CATATG*GAACGCTACGAAAATTTATTTGC	NdeI	This study
TRPSYN-A-R1	5'- TTT*CTCGAG*TTATGCGCGGCTGGC	XhoI	This study
TRPSYN-A-F2	5'- TT*CCATGG*GCGAACGCTACGAAAATTTATTTGC	NcoI	This study
PrnC_LIC-F1	5'- TATTTTCAATCCTACGGAATGACTCAGAAGAGCCCCG		This study
PrnC_LIC-R1a	5'- TTCCCTCAATATTATACGCCTTACTTCTTCAGAGCCAAGCC		This study

Restriction sites are indicated in italics. *These primers obtained for this study were designed by my colleague Dr Sabine Grüschow for a separate but related project. Primers were designed to be between 15 and 25 base pairs in length with a base stacking T_m of between 55 and 65 °C. The melting temperatures (T_m) of the forward and reverse primers were kept within 4°C of each other.

7.9 General molecular biology procedures

7.9.1 *In silico* procedures

DNA sequences were obtained from either the EMBL-EBI (European bioinformatics institute) or the NCBI GenBank nucleotide database (National Center for Biotechnology Information). Nucleotide sequence homologues were identified by NCBI blastn software using the nucleotide collection database. Protein sequences were obtained from the EMBL-EBI protein database. The BLAST software used to search for protein sequence homologues was either the NCBI blastp software using the non-redundant protein sequences database (nr) or the SIB BLAST Network Service (Swiss Institute of Bioinformatics). Multiple sequence alignments were performed with ClustalW2 software (EMBL-EBI). DNA sequences were translated into protein sequences using the ExPASy translate tool (Swiss institute bioinformatics). Protein secondary structure was predicted using the online PSIPRED predictor from the Bloomsbury Centre for Bioinformatics from University College London. Restriction endonuclease sites within nucleotide sequences were identified with NEBcutter software V2.0 (New England Biolabs). The theoretical melting temperatures (T_m) of PCR primer oligos were calculated using the Biomath T_m calculator (Promega); base-stacking melting temperatures were calculated under PCR master mix conditions adjusting for Mg^{2+} concentration (1.5 mM). Protein molarity calculations were converted from milligrams per millilitre using the Biomath Protein Molar Conversion calculator (promega).

7.9.2 Standard cell culture conditions

E. coli cultures were grown under the following conditions unless otherwise stated. *E. coli* cells from a stock culture (stored at -80°C) were streaked onto an LB-agar plate using a sterile toothpick. Single colonies on the plate were achieved by the method of dilution streaking (following general protocol in Sambrook & Russel 2001)[226] and incubating the agar plate at 37°C for 12-15 hours. From this plate single colonies were picked and used to inoculate a 10 ml LB starter culture (supplemented with antibiotic if appropriate) which was grown in an orbital shaker incubator (37°C, 180 rpm) for approximately 12 hours. This starter culture was then diluted 100 fold into fresh LB media of the required volume for the main culture. Small scale cultures (<20 ml) were grown in 50 ml screw cap centrifuge tubes. Pre-sterilised media was added to these tubes under aseptic conditions. Larger cultures (≥ 20 ml) were grown in conical flasks, with bungs in the top, which were autoclaved containing media prior to use.

7.9.3 DNA purification

7.9.3.1 High quality plasmid purification

Plasmids needed for sequencing and/or transformation were purified using Plasmid Mini Kits (Qiagen) using a modification of the manufacturer's protocol. All buffers referred to are provided by the manufacturer. To improve DNA yield a larger culture than stated in the protocol was used. A 20 ml main culture of LB, inoculated with the required strain of *E. coli* starter culture, was incubated under standard *E. coli* growth conditions overnight (approximately 15 hours). The culture was then centrifuged (1500g, 15min, 4 °C) and the resulting pellet re-suspended in 0.6 ml of Buffer P1. The mixture was then aliquoted into two 0.3 ml portions. 0.3 ml of lysis buffer P2 was then added to each portion and mixed by inversion. This was followed by incubation at room temperature for 5 minutes. 0.3 ml of neutralisation buffer P3 was then added to each aliquot; the contents were mixed and incubated on ice for a further 5 minutes. The mixture was then centrifuged (16060g, 10 min, 4 °C) to remove the unwanted insoluble matter. The contents of the two tubes were then combined and applied to an equilibrated Qiagen tip 20 column. From this point the standard manufacturer's protocol was followed.

7.9.3.2 Plasmid purification by alkaline lysis

A 10 ml main culture of *E. coli* was grown under standard conditions overnight. From this, 1 ml of culture was taken and centrifuged (16060g, 30 sec) in a 1.5 ml microcentrifuge tube and the supernatant removed. The pellet of cells was re-suspended in 100 µl of cold alkaline lysis solution 1

and vortexed to mix. 200 µl of alkaline lysis solution 2 was added. The contents were mixed by inversion and then placed on ice. 150 µl of cold alkaline lysis solution 3 was added and the contents again mixed by inversion. After 3-5 minutes on ice the contents were centrifuged (16060g, 5 min) and the supernatant collected. The DNA was recovered by precipitation using 2 volumes of ethanol, vortexed and left to stand for 2 minutes before centrifugation (16060g, 30 min). The pellet was washed with 1 ml of 70% ethanol and the tube inverted a few times before being centrifuged again (16060g, 10 min). The supernatant was removed and the tube left to dry slightly before the DNA was dissolved in 30 µl of EB buffer. 5 µl of DNA solution was set aside to be analysed via agarose gel electrophoresis.

7.9.3.3 Other DNA purification methods

Other DNA purifications (from PCR reactions, restriction digests or ligation reactions) were carried out using a QIAquick PCR Purification Kit (Qiagen). Purification of a specific size fragment of DNA from a mixture was achieved by first separating out the different DNA fragments by gel electrophoresis and excising out the slice of agarose containing the DNA using a scalpel. The DNA was then purified using a QIAquick Gel Extraction Kit (Qiagen). Both Qiagen kits were carried out according to the manufacturer's instructions.

7.9.4 Agarose gel electrophoresis

TAE buffer was made as a 50x stock and diluted fifty-fold when required. 0.8 g of electrophoresis grade agarose was added to TAE buffer (100 ml) in a conical flask and heated in the microwave (2 min, 90% power). The solution was mixed and returned to the microwave (2 min, 90% power). The resulting gel was cooled slightly (approximately 60 °C) before 2.5 µl of ethidium bromide solution (10% *w/v*) was added. The mixture was swirled to ensure good dispersal of the ethidium bromide. The gel was then poured into an electrophoresis rack fitted with two well-forming combs. The gel was then allowed to solidify before the combs were removed. The agarose gel was submerged into an electrophoresis tank containing TAE buffer.

Samples for analysis were prepared by the addition of 6x marker dye (1 part dye added to 5 parts DNA solution). 5 µl of 1kb DNA ladder (promega) was loaded into the first well using a micropipette. The samples to be analysed were then added into the adjacent wells (generally 5 µl for analytical purposes). Electric current was then applied (120 volts). After the DNA samples had run for long enough to enable satisfactory separation the current was removed. The DNA was visualised by viewing under UV light.

To enable DNA purification from an electrophoresis gel a high quality agarose reagent was used. The gel tray and the electrophoresis tank were washed before use and fresh TAE buffer used to minimise cross contamination. Five prongs on the comb were taped up to create a larger well to enable the loading of greater volumes of DNA into a single well. After satisfactory separation the band that corresponded to the required DNA fragment was visualised under UV light and excised from the gel using a scalpel and the resulting slice weighed. A Gel Extraction kit (Qiagen) was then used to purify the required DNA fragment. This kit was carried out according to the manufacturer's instructions.

7.9.5 Polymerase chain reaction

7.9.5.1 Specific PCR conditions

PCR Target	***Use in this Study***	***Template DNA***	***Primers***	***Annealing Temperature***	***Extension Time***
PrnC	Cloning PrnC into pBAD/HisA	*P. fluorescens* Pf-5 genomic DNA	PrnC-Bgl, PrnC-R1	68°C	10 min
			PrnC-Bgl, PrnC-R2	68°C	
			PrnC-Bgl, PrnC-R3	72°C	
	Cloning PrnC into pET28a(+)		PrnC-1, PrnC-R1	60°C	
	Producing PrnC with long overhangs to enable LIC cloning		PrnC_LIC-F1, PrnC_LIC-F2	60°C	6 min
TrpA	Tryptophan synthase α subunit with N-terminal His6 tag	pSTB7 plasmid	TRPSYN-A-F1, TRPSYN-A-R1	55°C	3 min
	Tryptophan synthase α subunit with no His6 tag		TRPSYN-A-F2, TRPSYN-A-R1		
TrpB	Tryptophan synthase β subunit with C-terminal His6 tag		TRPSYN-B-F1, TRPSYN-B-R1		

7.9.5.2 General PCR protocol

A typical PCR reaction involved first an initial denaturation (2 min at 94°C). This was followed by 30 cycles of denaturation (1 min at 94°C), annealing (1 min at primer specific temperature) and extension (72°C, time dependent on gene). After the 30 cycles the PCR was completed by a final extension step (10 min at 72°C) and the samples held at 4°C until ready to be used.

To establish optimal PCR conditions duplicate small scale initial reactions were performed at 50 µl scale using a range of annealing temperatures; 55, 60, 65 and 70°C. A typical reaction contained *Pfu* DNA polymerase buffer (5 µl, 10x concentrated stock solution, Promega), dNTPs (4 µl of 2.5 mM dNTPs stock), forward and reverse primers (0.5 µl of each), template DNA (1-2 µl), and *Pfu* DNA polymerase enzyme (0.5 µl, 3U/µl, Promega). To half the reactions 2.5 µl of DMSO was also added to a final concentration of 5% (v/v). The reactions were completed to 50 µl with sterile H_2O. With *Pfu* DNA polymerase an estimated extension time of 2 min per kb of template DNA was used.

The initial PCR reactions were analysed via agarose gel electrophoresis and the conditions that produced the optimum amount of DNA of the target size were scaled up to a 100 µl reaction. This was split up into individual 33.5 µl reaction volumes which were combined after the PCR was complete. The amplified DNA of required size was then usually gel purified. Every PCR reaction contained within this study was enhanced with the addition of 5% DMSO which was included in every scaled up PCR.

7.9.5.3 Colony PCR

Colony PCR was generally used to identify single colonies that contained the desired gene after competent cell transformation. For every four colonies picked a PCR solution of Biomix Red reagent was made (Bioline). 25 µl of Biomix Red (2x concentrated stock solution) was added to both the relevant forward and reverse primers (0.5 µl of each) and 2.5 µl DMSO (5% final volume (v/v)).

12.5 µl of sterile water was added to each of four PCR tubes. 4 colonies of the transformation plate were numbered and each colony was picked using a sterile toothpick. The toothpick was then quickly lowered into one of the PCR tubes containing sterile water and removed. The same toothpick was then used to inoculate 0.5 ml of LB in a 1.5 ml microcentrifuge tube (supplemented with appropriate antibiotic). These cultures were incubated in an orbital shaker incubator for 4-6 hours (37 °C, 180 rpm).

7.1 µl of the Biomix Red PCR solution prepared above was added to each tube to complete the PCR reactions. PCR reactions were then performed using annealing temperatures established during

initial PCR trials of the gene. Biomix Red polymerase operates at 15-30 seconds per kb of template DNA. Therefore the extension time was shortened to half of that used with Pfu polymerase.

After the PCR reactions were complete they were analysed using agarose gel electrophoresis. If one or more colonies showed amplified DNA of the correct size, one was chosen and the corresponding 0.5 ml culture was used to inoculate a 10 ml culture of LB (+ antibiotic) which was incubated in an orbital shaker incubator for 12-16 hours (37°C, 180rpm). From this culture a permanent bacterial stock was taken and stored at -80°C.

7.9.6 Restriction digests

7.9.6.1 General restriction digest protocol

DNA was digested with restriction endonuclease enzymes (Roche, Fermentas or NEB) according to the manufacturer's instructions using the recommended buffer. If more than a single enzyme was used then the buffer that showed greatest compatibility was selected. The volume of the reaction was then adjusted with sterile H_2O and incubated at 37°C for approximately 1 hour before analysis by gel electrophores. Most reactions were inactivated by heating to 65°C for 5 minutes. A PCR purification kit was used to purify the restricted DNA (Qiagen).

7.9.6.2 Restriction digest of plasmids

Plasmid DNA was digested using the general restriction digest protocol in a 100 µl final volume. To ensure good levels of digestion, reactions were digested overnight at 37°C. After this time 1 µl of shrimp alkaline phosphatase (SAP) was added to dephosphorylate the cut ends of the DNA to prevent internal ligation of the plasmid. The reaction was incubated at 37°C for 1 hour. The reaction was stopped with, and the DNA purified using a PCR purification kit.

7.9.7 Ligation of insert and plasmid

50-100 ng of vector DNA was incubated with approximately twice as much insert DNA in a solution containing 2 µl ligation buffer (10x concentrated stock) and 1U of T4 DNA ligase (Roche, 1U/µl). The solution was completed to a 20 µl final volume with sterile H_2O. The solution was kept on ice and allowed to reach room temperature slowly as the ice melted overnight. If a ligation reaction was unsuccessful a repeat reaction with the addition of 2.5 mM ATP (final concentration) was sometimes successful.

A non-insert control was usually prepared simultaneously. This consists of an identical ligation reaction but without the addition of insert DNA. This control measures the amount of vector self ligation.

A successful ligation was tested either by colony PCR or by the preparation of the plasmid from the *E. coli* host and subsequent double restriction digest to remove a section of DNA containing the inserted gene of known length. The digest was analysed by agarose gel electrophoresis and the size of the excised fragment examined.

7.9.8 Preparation of chemically competent *E. coli* cells

A starter culture of the pertaining *E. coli* strain was grown in LB following the standard protocol. The culture was then diluted fifty-fold into fresh LB media and incubation continued at 37 °C until an optical density OD_{600} of 0.4 was reached (approximately 4 hours).

The culture was then chilled on ice for 10 minutes and subsequently centrifuged (1500g, 10 min, 4 °C). The supernatant was decanted and the pellet re-suspended in equal volumes of pre-chilled KMES buffer 1 (4 °C). This cell suspension was placed on ice for 60 minutes. The cells were then centrifuged (1500g, 10 min, 4 °C) and the pellet re-suspended in 10% of the original culture volume of pre-chilled KMES buffer 2 (4 °C). The cell suspension was split into 100 µl aliquots into pre-chilled 1.5 ml microcentrifuge tubes and stored at -80 °C.

7.9.9 Plasmid transformation into bacterial host

7.9.9.1 General transformation protocol

To transform vector DNA into a host organism, the DNA (7 µl of a ligation mixture or 0.5-1 µl of pure plasmid DNA) was added to a 100 µl aliquot of chemically competent cells. The mixture was allowed to sit on ice for 30 minutes before the cells were heat shocked by submersion into a water bath at 42°C for 45 seconds. The competent cells were then returned to the ice for a further 2 minutes. 0.5 ml of LB media was then added to the cell aliquot and the cells shaken inside an orbital shaker incubator for 45 minutes (37°C, 180 rpm). After this time the cells were spread onto LB-agar plates (supplemented with relevant selection antibiotic) and left to incubate at 37°C for 12-14 hours or until single colonies had appeared.

7.9.9.2 Insertion of multiple plasmids into single host

If more than one plasmid was required inside a single host then one of the desired plasmids was transformed first. From this strain chemically competent cells were prepared again and the transformation repeated with the second plasmid. Separate selection markers on each plasmid (such as resistance to two different antibiotics) can be used to distinguish clones that contain both the desired plasmids.

7.9.10 Other molecular biology protocols

7.9.10.1 Ligation independent cloning (LIC)

7.9.10.1.1 Preparation of LIC vector

The ligation independent vector (pNYCOMPS-LIC-FH10T-ccdB) was purified from a culture of *E. coli* DH10B containing the plasmid according to the standard method of plasmid DNA purification. A 1527 base pair fragment of the plasmid was then removed by digestion with SnaBI (Fermentas) for 3 hours (37°C). SnaBI is a blunt cutting restriction enzyme that leaves behind blunt ended DNA. The excised fragment contains chloramphenicol resistance (Clm^R) and toxic gene that stop undigested vector from propagating in host after bacterial transformation. The blunt ended vector DNA was purified from the excised fragment by agarose gel electrophoresis.

The 3' to 5' exonuclease activity of T4 DNA polymerase was then harnessed. This enzyme excises the DNA from the 3' end in the absence of dNTPs to create a 5' overhang. The size of this overhang is limited by the inclusion of a guanosine (G) nucleotide 15 base pairs from the blunt end of the DNA strand. The rest of the overhang contains only A,T and C base pairs. By supplementing the T4 DNA polymerase reaction with only dGTP and no other nucleotide the polymerase enzyme will excise the DNA from the 3' end until meeting the first G base, causing the polymerase to stall and creating a 15 base pair overhang. This same process occurs at the opposite end of the linearised plasmid resulting in a non-complimentary 17 base pair overhang.

10 µl of plasmid purified pNYCOMPS-LIC-FH10T-ccdB was treated with 1 µl of T4 DNA polymerase (5 U/µl, Fermentas) in 5 µl reaction buffer (5x concentrated). The reaction was supplemented with 8 µl dGTP (2.5 mM) and 1 µl of sterile H_2O. The reaction was incubated for 30 minutes at 22°C. 4 µl of 100 mM EDTA was then added before heating the reaction to 75°C for 20 minutes to denature the polymerase.

7.9.10.1.2 Preparation of LIC insert

Primers were designed to contain complimentary base pair overhangs to the LIC vector after T4 DNA polymerase treatment. This sequence was added to the beginning of previous PrnC primers. A PCR reaction using these primers (PrnC_LIC-F1, PrnC_LIC-R1a) was carried out using the PCR conditions listed above (section 7.9.5.2). The resulting PCR product was gel-purified and T4 DNA polymerase used to create 5' overhangs. Because the overhangs of the insert are complimentary to the vector the overhang contains only A,T and G base pairs and the polymerase will stall once it reaches the first dCTP. 3 µl of purified PCR product was treated with 1 µl of T4 DNA polymerase (5 U/µl, Fermentas) in 5 µl reaction buffer (5x concentrated). The reaction was supplemented with 8 µl dCTP (2.5 mM) and 8 µl of sterile H_2O. The reaction was incubated for 30 minutes at 22°C. 4 µl of 100 mM EDTA was then added before heating the reaction to 75°C for 20 minutes to denature the polymerase.

7.9.10.1.3 LIC ligation protocol

3 µl of digested insert DNA was added to 6 µl of prepared vector DNA. A no insert control was also prepared by supplementing 3 µl of H_2O in place of insert. The reactions were incubated at room temperature for 10 minutes. 3 µl of 100 mM EDTA was added and the reactions heated to 75°C. The reactions were then cooled slowly back to room temperature. 2 µl of the mixture was then used to transform chemically competent *E. coli* DH10B cells. The following day successful transformants were identified using colony PCR and double restriction digest. Final confirmation of inserted genes was achieved by sending the purified plasmid for full sequencing.

7.9.10.2 Performing Quikchange site-directed mutagenesis

7.9.10.2.1 Designing Quikchange primers

Site-directed mutagenesis of pMW06 was performed to increase the length of the His6 tag to His8. This was performed based on the Quikchange protocol (Stratagene). Mutagenic primers were designed based on the guidelines in the protocol. Two primers were designed that contained the desired mutation in the middle of the primer with 10-15 base pairs of correct sequence on either side that annealed to the same sequence but on opposite strands of the plasmid. Primers were designed to be between 25 and 45 bases in length with a melting temperature (T_m) of ≤ 78°C. The T_m was calculated using the following equation (where N is the primer length not including inserted/deleted bases): $Tm = 81.5 + 0.41 (\%GC) – 675 / N$

The optimal GC content should be 40% and the primer should terminate in one or more GC bases. The resulting primers were designated pET28a-8xHis-F1 and pET28a-8His-R1.

7.9.10.2.2 Quikchange mutagenesis protocol

The plasmid pMW06 was purified as detailed previously. PCR reactions were prepared as follows using either 50 ng or 500 ng of DNA. pMW06 template DNA (0.25 and 2.5 µl respectively)was added to *Pfu* DNA polymerase buffer (5 µl), mutagenic primers (1 µl of each) and dNTPs (4 µl of 25 mM stock). *Pfu* DNA polymerase (1 µl) was added and H_2O added to a final volume of 50 µl. A no primer control was also added.

A modified PCR method was used. This consisted of an initial denaturation step (95°C, 30 seconds) followed by 18 cycles of denaturation (95°C, 30 seconds), annealing (55°C, 60 seconds) and extension (68°C, 15 minutes). A final extension completed the PCR cycle (68°C, 30 minutes). DpnI endonuclease was then added (1 µl) to each reaction to digest the methylated template DNA and incubated for 1 hour (37°C). The no primer control was used to determine the efficiency of this digest. 5 µl of the PCR mixture was used to transform chemically competent *E. coli* DH10B cells. After overnight incubation there were many cells visible on the control plate and many more on the 500 ng plate. A second round of DpnI digestion yielded a clearer difference between the control plate and the PCR plate. 4 colonies were picked from the 500 ng plate and grown in 10 ml of LB. The resulting culture was used to prepare pure plasmids which were sent for DNA sequencing to identify transformants carrying the mutation in the histidine tag. Two of the colonies were found to contain the additional histidine residues.

7.10 General proteomic methods

7.10.1 Protein production

7.10.1.1 Optimisation of arabinose induction of the araBAD promoter

Using standard culture conditions, a series of five 10 ml main LB cultures were inoculated with a starter culture of *E. coli* (BL21 or DH10B transformed with the pBAD based plasmid of choice). The cultures were incubated at 37 °C for 4 hours. Once the OD_{600} of the cultures had reached ≈ 0.5, 1 ml of culture was removed and centrifuged (16060g, 1 min) to act as a pre-induction control. Different final concentrations of arabinose were added to each of the cultures (0.2, 0.02, 0.002, 0.0002 and 0.00002% arabinose final concentration). The cultures were incubated for a further 4 hours. 1 ml was then taken from each of the cultures and centrifuged (16060g, 1 min). The pellets were analysed

by SDS-PAGE analysis to check for varying levels of protein expression. The percentage of arabinose that produced the most protein was carried forward to larger scales.

7.10.1.2 Optimised production of protein under araBAD promoter

Main *E. coli* BL21 or DH10B cultures (transformed with the pBAD plasmid of choice) were prepared from starter cultures grown in the standard way. The main cultures were left to incubate at 37 °C in an orbital shaker incubator until an OD_{600} of approximately 0.5 was reached (around 3 hours). Protein expression was then induced with arabinose (final concentration determined by initial optimization efforts) and left to incubate for 20 hours at 16 °C. After this time the cultures were centrifuged (1500g for 10 ml culture, 3315g for 500 ml culture, 20 min, 5 °C), the media removed and the pellet stored at -20°C for at least 12 hours before protein purification or SDS-PAGE analysis.

7.10.1.3 Production of protein under T7*lac* promoter

A main culture of *E. coli* BL21 cells (transformed with plasmid of choice under the control of the T7 promoter) was inoculated with a starter culture prepared under standard culture conditions and incubated in an orbital shaker incubator (37 °C, 180 rpm) until an OD_{600} of between 0.6 and 0.8 was achieved. A 0.5 ml sample was set aside as a pre-induction control before 0.1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) final concentration was added to induce the T7 promoter. The culture was then incubated for 20 hours at low temperature (16 °C, 180 rpm). After this time the cultures were centrifuged (1500g for 10 ml culture, 3315g for 500 ml culture, 20 min, 5 °C), the media removed and the pellet stored at -20°C for at least 12 hours before protein purification or SDS-PAGE analysis.

7.10.1.4 Auto-induction of proteins cloned into pET28a(+)

Single colonies of *E. coli (*transformed with the pET28a(+) based plasmid of choice) were used to inoculate 2 ml of ZYP-0.8G media (+ 50 µg/ml kanamycin) which was incubated in an orbital shaker incubator (37 °C, 180 rpm, 12 hours) until the culture was slightly turbid. 200 µl of this starter culture was used to inoculate 400 ml of ZYP-5052 (+ 50 µg/ml kanamycin). 1 ml was taken at this point to serve as a post-induction control sample for SDS-PAGE analysis. The culture was then returned to an orbital shaker incubator at a lower temperature and higher rpm (24 °C, 220 rpm) for 15 hours. After this time the cultures were centrifuged (1500g for 10 ml culture, 3315g for 500 ml culture, 20 min, 5 °C), the media removed and the pellet stored at -20°C for at least 12 hours before protein purification or SDS-PAGE analysis.

7.10.1.5 Production of protein under the propionate promoter

A main culture of *E. coli* BL21 or PHL644 (containing a pPro24(s)-gfp based plasmid) was prepared from a starter culture grown under standard conditions. This main culture was incubated in an orbital shaking incubator (37°C, 180 rpm) until the OD_{600} reached ≈0.6. Sodium propionate was then added to a final concentration of 20 mM (200 µl of 1M stock added to every 10 ml) and incubation continued for 5 hours for 10 ml cultures or switched to a lower temperature for 500 ml cultures and incubated for longer (16°C, 20 hours, 180 rpm). After this time the cultures were centrifuged (1500g for 10 ml culture, 3315g for 500 ml culture, 20 min, 5 °C), the media removed and the pellet stored at -20°C for at least 12 hours before protein purification or SDS-PAGE analysis

7.10.1.6 Co-production of PrnC with chaperone proteins

3 main LB cultures were grown up for each of the 3 chaperone/PrnC constructs (*E. coli* BL21 MW-022, MW-023 and MW-024) using the usual culturing method. The three cultures were supplemented with decreasing concentrations of L-arabinose (0.4, 0.2 and 0.05% final concentration (w/v)) and incubated at 37°C until an OD_{600} of 0.5 was recorded. Production of the PrnC gene was then induced with the addition of 0.1 mM IPTG. Incubation was switched to 16°C for a further 24 hours. Following incubation the cells were pelleted by centrifugation (1500g, 15 min, 4°C) and chemical lysis carried out to generate a cell free lysate. This lysate was subjected to small scale Ni-NTA purification which was analysed by SDS-PAGE. The initial concentration of arabinose that produced the best results was scaled up to 500 ml scale and Ni-NTA purification used to purify the resulting PrnC.

7.10.2 Generation of cell-free lysates

7.10.2.1 Chemical cell lysis

For small scale cultures (< 200 ml) chemical cell lysis was used instead of sonication. A pellet of cells prepared from an original 10 ml culture volume was re-suspended in 1 ml of chemical lysis buffer (+ 2 mg/ml lysozyme). If Ni-NTA purification was not to be used as the next step then chemical lysis buffer containing additional EDTA was used instead. The cell suspension was incubated at 5°C for between 30 to 60 minutes. Once a viscous solution had formed it was passed several times through a G23 gauge needle to loosen the mixture. The culture was then centrifuged to collect the cell debris (14000g, 20 min, 4°C) and the supernatant either used straight away or frozen at -20°C for long term storage

7.10.2.2 Cell lysis by sonication

A pellet of cells prepared from an original 500 ml culture was re-suspended in 40 ml lysis buffer. The cells were harvested again (3315g, 20 min, 5°C) and the buffer removed. This process was repeated once more to remove as much of the growth media as possible. The pellet was once more re-suspended in 40 ml of lysis buffer which was then sonicated (8.20 min 10% pulse 49% power). From this point on work was kept on ice or otherwise carried out at 5°C or below. The lysate was cleared by centrifugation (20442g, 30 min, 4 °C) and the supernatant either used straight away or frozen at -80°C for long term storage.

7.10.2.3 Generation of tryptophan synthase cell lysate

A 500 ml culture of *E. coli* MW-043 was inoculated from a 12 hour starter culture in the usual way, it was then returned to the incubator overnight for 12-15 hours before being centrifuged (3315g, 20 min, 5°C). The resulting pellet of cells was re-suspended in 40 ml of saturated NaCl solution. The cells were harvested again (3315g, 20 min, 5°C) and the buffer removed. This process was repeated once more to remove as much of the growth media as possible. The pellet was then re-suspended in 40 ml of tryptophan synthase lysis buffer and sonicated (8.20 min 10% pulse 49% power). From this point on work was kept on ice or otherwise carried out at 5°C or below. The lysate was cleared by centrifugation (20442g, 30 min, 4 °C) and the supernatant either used straight away or frozen at -80°C for long term storage.

7.10.3 Protein purification

7.10.3.1 Standard Ni-NTA purification protocols

7.10.3.1.1 Ni-NTA Spin purification

Protocol for Ni-NTA purification for 1 ml cleared cell lysate from an original 10 ml culture volume.

80 µl of Ni-NTA slurry (50% resin in ethanol) was equilibrated into Ni-NTA lysis buffer in a 1.5 ml capped microcentrifuge tube. This was achieved by centrifugation of the slurry (16060g, 2 min) to separate the resin from the ethanol storage buffer. The ethanol was removed and the resin re-suspended in an additional 100 µl of lysis buffer. The resin was centrifuged again and the buffer removed. This was repeated three times to remove the ethanol. Cleared cell lysate was then added to the equilibrated Ni-NTA slurry (≈ 40 µl resin). The lysate/resin suspension was then incubated at 4 °C for 1 hour on a slowly rotating wheel (approximately 30 rpm). After this time the resin/cell lysate suspension was centrifuged (14000g, 2 min, 4 °C) and the lysate removed. 30 µl was kept for SDS-

PAGE analysis. The resin was washed with 0.5 ml of Ni-NTA lysis buffer then with 0.3 ml of Ni-NTA wash buffer. The resin suspension was centrifuged after each wash (14000g, 2 min, 4°C), the supernatant removed and 30 µl retained for analysis. Finally the protein was eluted from the resin with 0.2 ml of Ni-NTA elution buffer. The resin was centrifuged and the 30 µl of the elution buffer removed for analysis. All fractions were analysed via SDS-PAGE Gel Electrophoresis.

7.10.3.1.2 Standard NiNTA purification

Protocol for Ni-NTA Purification of 40 ml cell lysate generated from 500 ml original culture volume.

2 ml of Ni-NTA slurry (≈ 1 ml resin) was equilibrated with 2 x 20 ml washes of Ni-NTA lysis buffer followed by centrifugation (1500g, 5 min, 4 °C) each time to collect the resin and allow the supernatant to be pipetted off. The 40 ml cell-free lysate was added to the resin and incubated at 4 °C for 1 hour on a slowly rotating wheel (30 rpm). The resin suspension was then added to a plastic column (Biorad) and the supernatant allowed to flow through under gravity. A Bradford micro-assay was used to measure the amount of protein eluting from the column. The resin in the column was washed with 2 x 15 ml washes of Ni-NTA lysis buffer and samples of the flow through tested for decreasing levels of protein being washed off. 2 x 15 ml washes of Ni-NTA wash buffer were then added to the column and the flow-through collected and analysed for protein elution. This step was continued until little protein was seen coming off the column as visualized by the micro Bradford assay (3 x 15 ml washes normally suffice). The protein binding tightly to the Ni-NTA resin was eluted using approximately 25 ml of elution buffer. The flow-through was collect in 1 ml fractions, 5 µl of each were tested using the micro-Bradford assay. The fractions containing the strongest blue colour and therefore the most protein were combined together. The protein solutions were concentrated down to 1.5 ml (2700g, ≈30 min, 4 °C) using Amicon Ultra 15 centrifugal filter devices with an Ultracel-10 membrane (Millipore). For long term storage the protein solution was topped back to 15 ml with protein storage buffer and concentrated down to 1.5 ml again. This was repeated twice more until the elution buffer had been diluted into storage buffer 1000 times. 50-100 µl aliquots were placed into pre-chilled 1.5 ml microcentrifuge tubes and flash frozen with liquid nitrogen and then stored at -80 °C. Samples of each stage were retained to enable analysis by SDS-PAGE.

7.10.3.2 Optimisation of NiNTA binding by PrnC

7.10.3.2.1 Removal of imidazole during binding step

A 500 ml culture of *E. coli* BL21 MW-015 was grown and protein production under the T7lac promoter induced using the standard protocols. The resulting pellet was re-suspended in 50 ml of

Ni-NTA lysis buffer containing no imidazole. The suspension was split into two aliquots which were pelleted again (2580g, 10 min, 5°C). One of the resulting cell pellets was re-suspended into standard lysis buffer and the other into lysis buffer containing no imidazole. The cells were sonicated in the usual way and the lysate cleared. 1 ml of Ni-NTA resin (equilibrated with lysis buffer with or without imidazole) was added to each of the lysate fractions and standard Ni-NTA purification carried out. The purifications were analysed by SDS-PAGE.

7.10.4 Protein quantification

7.10.4.1 Bradford total protein assay

A 10 mg ml^{-1} solution of Bovine Serine Albumin (BSA) was used to create a number of serial dilutions resulting in 0.2, 0.4, 0.6, 0.8 and 1 mg ml^{-1} solutions of BSA.

1 ml of Bradford reagent (Biorad) was placed into a series of spectrophotometer cuvettes. 20 µl of each of the protein standard solutions were added to a separate cuvette which was lightly agitated to mix. 20 µl of the sample of protein of unknown concentration was also added to a cuvette containing the Bradford reagent. If required, the unknown protein solution was diluted in order for the strength of the colour of the reagent to fall within the range of the BSA standards. The samples were incubated at room temperature for around 20 minutes before the absorbance of each was recorded at 595 nm. The readings were carried out in duplicate.

A standard curve of BSA concentration versus absorbance was constructed using the data. From this standard curve the unknown concentration of the protein sample was deduced from the absorbance readings taking into account any dilution factor that had been introduced along the way.

7.10.4.2 Micro-Bradford assay

100 µl of Bradford reagent (Biorad) was placed into the wells of a 96-well plate. 5-10 µl of the Ni-NTA purification column flow-through was added to one of the wells. The strength of the blue colour indicates the concentration of protein present in the sample.

7.10.5 Analysis of proteins by SDS-PAGE gel electrophoresis

7.10.5.1 Preparation of a 12% acrylamide gel

For the analysis of protein via SDS-PAGE (Sodium dodecyl sulfate polyacrylamide gel electrophoresis) a 12% resolving gel was mixed according to the values in section 7.4. As soon as the ammonium persulfate and TEMED (tetramethylethylenediamine) are added the polymerisation of the gel begins.

Therefore the solution was pipetted into a gel caster as quickly as possible. A thin layer of IPA (propan-2-ol) was then added to the top of the gel before allowing the gel to set (around 30 min to 1 hour). The IPA layer was then decanted and the top of the gel washed with some distilled water. A 5% stacking gel was then mixed following the recipe in section 7.4. The stacking gel was poured on top of the already set resolving gel and a comb inserted to form the wells. The gel was then allowed to polymerise. Once gel was set the comb was removed.

7.10.5.2 Sample preparation and running of SDS-PAGE

Samples for analysis were diluted in equal volumes of 2x SDS loading dye and boiled for 10 minutes. 5-10 µl of sample was then loaded into each of the wells of the SDS gel. Any empty wells were filled with blank loading dye. The SDS-PAGE tank was filled with running buffer and a current of 200 volts was applied until the blue loading dye reached the bottom of the gel. After completion the stacking gel was removed and the resolving gel placed into a Coomassie Brilliant Blue staining solution and left to stain overnight on a shaking platform. To de-stain the gels they were placed into de-staining solution (section 7.4) and heated in the microwave for 2 minutes. The solution was then left to cool before the de-staining solution was removed and replaced by fresh. The process was repeated until satisfactory de-staining had occurred. After this time the acrylamide gels were stored in distilled H_2O.

The gels were dried by placed them into drying solution (section 7.4) and laying them between two film sheets. The slides of the sheets were clamped between two metal frames and allowed to dry. The acrylamide gel could then be cut out and stored. Alternatively after de-staining the gel was scanned into digital format for storage.

7.11 Biofilms and immobilised biocatalysts

7.11.1 Artificially spin-coated biofilm

7.11.1.1 Spin-coated biofilm protocol

Prior to biofilm formation, glass microscope slides (75 mm by 25 mm, VMR) were coated with approximately 4 ml of 0.1% (w/v) poly-lysine (PLL) in water (Sigma), which was then dried overnight in an oven at 60°C. 10 µl of a culture of *E. coli* PHL644 MW-002 was streaked onto an agar plate (supplemented with ampicillin) and incubated at 37°C for 14 hours. Single colonies were picked and used to inoculate 200 ml of ½ LB (1:1 LB:H_2O, supplemented with ampicillin). The culture was incubated in an orbital shaker (30°C, 180 rpm) for 16 hours.

Following incubation, cultures were transferred aseptically into sterile 750 ml polypropylene centrifuge bottles (Beckman Coulter UK Ltd.) containing the PLL-coated glass slides supported on a bed of glass beads (200 g, soda-glass beads, 4 mm diameter) to provide a flat surface to prevent cracking during centrifugation. The slides were centrifuged (1851g, 10 min) in a centrifuge fitted with a swinging bucket rotor. After centrifugation the glass slides were gently placed in 500 ml sterilised wide necked Erlenmyer flasks (Fisher Scientific) containing 70ml of M63 medium (supplemented with ampicillin). The spin coated biofilms were incubated in an orbital shaker incubator (30 °C, 70 rpm), set at a low speed to minimise bacteria shearing from the biofilm, for a maturation period of 7 days.

7.11.1.2 Testing stability of the spin-coated biofilm

In order to utilise the engineered biofilm as an immobilised catalyst, its stability to the reaction buffer and incubation rpm was tested. Buffer stability tests were carried out by placing the engineered biofilm (matured for 7 days) into the reaction buffer. The M63 biofilm maturation medium was carefully removed from the biofilm-coated slide using a syringe. Un-adhered, planktonic cells were removed from the biofilm by gentle re-submersion of the slides and washing in aliquots of reaction buffer (2 x 50 ml), which was then also removed. The washed biofilm slides were then submerged in 70 ml of potassium phosphate reaction buffer and placed into an orbital shaker incubator (30°C, 70 rpm) and incubated for up to 45 hours. At intervals, 1 ml aliquots of reaction buffer were removed and the OD_{600} measured in order to determine the amount of planktonic cells that had left the biofilm. A biofilm coated slide was totally re-suspended into 70 ml of reaction buffer and the OD_{600} of this used to represent the value of 100% dissociation.

The stability of the engineered biofilm to agitation was assessed by re-submerging a spin-coated biofilm into reaction buffer as above and incubating at 70 rpm for 5 hours. After 5 hours the agitation was increased to 100 rpm for 1 hour before the agitation was increased again to 150 rpm for a further hour. At intervals, 1 ml aliquots of reaction buffer were removed and the refractive index of the solution or the OD_{600} measured in order to determine the amount of planktonic cells that had left the biofilm.

7.11.1.3 UEA spin-coating protocol

7.11.1.3.1 Full size glass slides

The spin-coating method of biofilm formation developed in Birmingham University labs was adapted for use back at UEA.

Glass slides (75 mm by 25 mm, VMR) were placed onto a bed of glass beads (150g, soda-glass beads, 4 mm diameter) inside 1000 ml centrifuge tubes (Beckman). The tubes and their contents were then autoclaved. A layer of poly-lysine was applied to the top of the glass slide under sterile conditions and dried overnight at 40°C.

A 200 ml culture of *E. coli* PHL644 MW-002 (+ ampicillin) was inoculated from a starter culture and allowed to incubate in an orbital shaker incubator for 16 hours (37°C, 180rpm). The culture was then poured into the pre-prepared centrifuge tubes and the culture centrifuged (1865g, 10 min, 4°C) in a fixed-angle centrifuge rotor. The coated slides were then handled in the same way as in the standard spin-coating biofilm formation protocol.

7.11.1.3.2 Reduced size glass slides

Glass slides (75 mm by 25 mm, VMR) were cut into smaller squares (1.5 cm in length) or into rounds (2 cm diameter). Glass beads and the reduced size glass slides were autoclaved. Approximately 10 ml of glass beads were placed into a 50 ml screw capped centrifuge tube and a reduced size glass slide adjusted to sit on top. 0.3 ml of poly-lysine was applied to the surface of the slides and the lid loosely fastened. The slides were allowed to dry at 37°C for 24 hours.

A 200 ml culture of *E. coli* PHL644 MW-002 (+ ampicillin) was inoculated from a starter culture and allowed to incubate in an orbital shaker incubator for 16 hours (37°C, 180rpm). 20 ml of the culture was then poured into the 24 pre-prepared centrifuge tubes and the culture centrifuged (1500g, 10 min, 4°C). The LB media was carefully removed using a syringe. Using tweezers the glass slides were transferred into sterile 250 ml conical flasks. 2x 10 ml of M63 was used to gently wash the slides before the media was removed with a syringe. 30 ml of fresh M63 media was then added (+ ampicillin) and the slides incubated in an orbital shaker incubator at low rpm (28°C, 60 rpm, 7 days).

7.11.2 Generation of naturally deposited biofilms

Natural biofilms were generated by harvesting the 16-hour cultures (generated above) by centrifugation (1851g, 15 mins), and re-suspending the bacteria in 70 ml of M63 medium in a 500 ml wide necked Erlenmyer flask into which the PLL coated glass slide was introduced. The slide was incubated for 7 days (30°C, 70 rpm) in an orbital shaker incubator.

7.11.3 Immobilisation of tryptophan synthase onto Ni-NTA resin

7.11.3.1 Restriction digest of *trpA* and *trpB* out of pSTB7 and ligation into pET28a(+)

The plasmid pSTB7 was purified from *E. coli* CB149 MW-043. The genes representing the two subunits of tryptophan synthase are flanked by HindIII and EcoRI restrictions sites. 30 µl of purified plasmid was digested with 4 µl of HindIII and EcoRI (Roche 10U/µl) in 52 µl of H_2O and 10 µl buffer (Roche Buffer B). The restriction digest was incubated at 37°C overnight before being run on agarose gel to separate out the linearised plasmid (6.3 kb) and excised genes (1.8 kb). The band corresponding to the excised genes was gel purified and ligated into pET28a(+) that had been doubly digested with the same enzymes. The ligation mix contained 5 µl of vector DNA with 9 µl of the purified excised *trpA* and *trpB* genes. Standard ligation protocol was followed and the finished vector (pMW17) was transformed into *E. coli* DH10B for long term storage (MW-029) and *E. coli* BL21 (MW-030) for protein expression. This would have the result of producing a His6 fusion *trpB* gene which can be used to immobilise the β-subunit of tryptophan synthase onto Ni-NTA resin.

7.11.3.2 Cloning *trpA* and *trpB* into pET vectors using PCR

The *trpB* gene of tryptophan synthase from *Salmonella enterica* sv Typhimurium TB1533 was amplified from the plasmid pSTB7 using the TRPSYN-B-F1 and R1 primers. PCR conditions are shown in section 7.9.5.2. A restriction digest was carried out on the PCR product with HindIII and XhoI which was subsequently ligated into the corresponding restriction sites of pET21a(+) (Novagen) to give pMW18. This provided a route to the C-His-tagged β-subunit of tryptophan synthase under the control of the T7 promoter and ampicillin selection marker.

The *trpA* gene of Tryptophan Synthase was also amplified from pSTB7 using the TRPSYN-A-F1 and R1 primers. A restriction digest was carried out on the PCR product with NdeI and XhoI restriction enzymes and subsequent ligation into pET28a(+) (Novagen) was performed to give pMW19. This provided a route to the N-terminally His6 tagged α-subunit of tryptophan synthase.

Also the *trpA* gene of Tryptophan Synthase was amplified from pSTB7 using the TRPSYN-A-F2 and R1 primers. A restriction digest was carried out on the PCR product with NcoI and XhoI restriction enzymes and subsequent ligation into pET28a(+) (Novagen) was performed to give pMW21. The NcoI restriction site is upstream of the N-terminal His-tag sequence in pET28a(+) therefore trpA was expressed under the T7 promoter and kanamycin resistance marker without a poly-histidine tag.

pMW17 was transformed into *E. coli* DH10B for long term storage at -80°C (MW-029) and *E. coli* BL21 for protein expression (MW-030). pMW18 was transformed into *E. coli* BL21 for protein expression (MW-031). Chemically competent cells were constructed from *E. coli* BL21 MW-031 and either pMW19 or pMW21 were co-transformed into this strain to form *E. coli* BL21 MW-034 and MW-036 respectively. The host expressing both genes was obtained by dual selection with ampicillin and kanamycin.

7.11.3.3 Immobilisation of tryptophan synthase subunits

LB (10 ml) supplemented with appropriate antibiotic(s) was inoculated with a His6 tryptophan synthase fusion protein containing strain (*E. coli* BL21 MW-030, MW-034 or MW-036) and incubated in an orbital shaker incubator (37°C, 180 rpm) for 12 hours. This starter culture was then used to inoculate 500 ml of LB (+ appropriate antibiotics) which was then also placed in an orbital shaking incubator (37°C, 180 rpm). Once an OD_{600} of between 0.6-0.8 was reached protein production was induced by the addition of 0.1 mM IPTG and growth continued (37°C, 180 rpm) for 4 hours. The cells were then collected by centrifugation (3315g, 20 min, 5°C) and the cell free extract prepared using the general method. The His-tagged tryptophan synthase subunits were then incubated with Ni-NTA (Ni-nitrilotriacetic acid) resin using the standard protocol for Ni-NTA purification (section 7.10.3.1.2). After incubation the resin was then sequentially washed by re-suspension in 10 ml of lysis buffer followed by 10 ml of wash buffer to remove any weakly binding protein impurities from the resin, collecting the resin via centrifugation prior to each wash. The resin-bound protein was then used directly for biotransformation or quantification analysis.

7.11.4 Quantification of biocatalyst amounts

7.11.4.1 Quantification of dry cell biomass of the spin-coated biofilm by plate drying

The total biomass was determined for 14 slides that had been coated with engineered biofilm and left to mature for 7 days. This was achieved by disrupting and re-suspending the biofilms in sterile water into pre-weighed centrifuge tubes using a vortex mixer for 30 minutes then centrifuged (13165g, 15 min). The supernatant was removed and the biomass allowed to dry at 100^{o}C for 24 hrs. The mass was measured every 5 to 8 hours, the dry biomass being determined when the mass stopped decreasing.

7.11.4.2 Quantification of the dry cell biomass of 1ml of planktonic cell suspension

LB (10 ml) supplemented with ampicillin was inoculated with *E. coli* PHL644 MW-002 and incubated in an orbital shaker incubator (37°C, 180 rpm). After the culture reached an OD_{600} of 0.6 it was diluted 100-fold into 500 ml of fresh LB supplemented with ampicillin and incubation continued. Once growth had reached stationary phase (OD_{600} > 1) the cells were harvested by centrifugation (7529g, 20 min, 5°C) and the pellet was washed and re-suspended in 50 ml of potassium phosphate reaction buffer. The dry mass of cells was directly measured by taking aliquots of various volumes of the cell suspension, centrifuging them (16060g, 60 seconds) and drying the cells in an oven (60°C) for 4 days in pre-weighed vessels until a constant mass was reached. The masses of the dried cells were then obtained. This demonstrated a linear relationship between the volume of cell suspension and the dry cell biomass.

7.11.4.3 Quantification of biomass of planktonic cells by Bradford total protein assay

LB (10 ml) supplemented with ampicillin was inoculated with *E. coli* PHL644 MW-002 and incubated in an orbital shaker incubator (37°C, 180 rpm). After the culture reached an OD_{600} of 0.6 it was diluted 100-fold into 500 ml of fresh LB supplemented with ampicillin and incubation continued. Once growth had reached stationary phase (OD_{600} > 1) the cells were harvested by centrifugation (7259g, 20 min, 5°C) and the pellet was washed and re-suspended in 50 ml of potassium phosphate reaction buffer. Aliquots of a range of volumes (0.2 – 2 ml) were taken from this suspension of cells. The cells were collected by centrifugation (16060g, 60 seconds) and re-suspended in 1 ml of water. To release cellular proteins, 100 µl of 10% (w/v) NaOH was added and the sample incubated at 30°C for 30 minutes. The samples were then heated at 100 °C for 15 minutes and then allowed to cool. 80 µl of 2M HCl was then added to neutralise the pH. The samples were centrifuged (16060g, 10 minutes). The supernatant was removed, diluted tenfold with water. The protein concentration in these samples was assessed by Bradford assay. The amount of cell suspension equal to 80 mg dry mass (total dry biomass of biofilm) was calculated to be 6 ml using the assumption that the average total protein content in cells is approximately equal to 50% dry cell weight.[64]

7.11.4.4 Quantification of tryptophan synthase immobilised onto Ni-NTA resin

The quantity of tryptophan synthase catalyst loaded onto the Ni-NTA resin was assessed by eluting the purified protein off the resin with 10 ml of Ni-NTA elution buffer. The amount of protein that had been immobilised onto the resin was determined by a Bradford assay (Biorad) and the relative amount of the α and β subunits was estimated by SDS-PAGE analysis.

7.11.5 Biotransformations

7.11.5.1 Tryptophan synthase biotransformations

7.11.5.1.1 Tryptophan synthase cell lysate biotransformations

A cell free lysate was prepared from a culture of *E. coli* CB149 MW-043 following the method of tryptophan synthase lysate preparation. For long term storage the lysate was stored at -80°C or freeze-dried and stored at 5°C. 5-bromo-, 5-chloro- or 5-fluoroindole (2 mM) was added to a 250 ml conical flask containing 50 ml of potassium phosphate reaction buffer. 3 ml of prepared *E. coli* CB149 MW-043 cell free lysate was added to dialysis tubing and added to the reaction. The reaction was placed into an orbital shaking incubator (37°C, 180 rpm) for 48 hours. The reaction was either analysed by HPLC or purified by reverse-phase chromatography on C18 silica.

For C18 silica purification the unreacted indole was removed by washing with 2x 100ml ethyl acetate. The aqueous layer was reduced in volume under reduced pressure to approximately 10-20 ml. This was applied to 30 g of C18 silica (sigma) which was washed with H_2O until nothing could be seen eluting off the column by TLC visualised with ninhydrin. The L-halotryptophan was recovered by washing with 10 ml aliquots of MeOH until no further elution off the column could be seen. The MeOH was removed *in vacuo* to yield the pure L-halotryptophan. The HCl salt of the halotryptophan was made by re-dissolving in 2M HCl and removing the acid *in vacuo* to yield the crystalline salt.

5-fluoro-tryptophan was yielded as a pale pink crystalline solid (18mg, 0.081 mmol, yield: 83%). **^{1}H NMR** (D_2O): δ ppm 3.29 (dd, *J*=15.4, 7.1 Hz, 1H), 3.36 (dd, *J*=15.4, 5.4 Hz, 1H), 4.23 (t, *J*=6.2 Hz, 1H), 6.94 (td, *J*=9.3, 2.4 Hz, 1H), 7.24 (d, *J*=2.4 Hz, 1H), 7.26 (s, 1H), 7.36 (dd, *J*=8.9, 4.5 Hz, 1H);**MS (ESI):** *m/z:* (M+H) 222.95 (100%), 223.95 (14%); **UV/λmax**: 278 nm.

5-chlorotryptophan was yielded as a pink crystalline solid (11.9mg, 0.05 mmol, yield: 50%). **^{1}H NMR** (D_2O): δ ppm 3.24 (dd, *J*=15.2, 7.1 Hz, 1H), 3.31 (dd, *J*=15.4, 5.4 Hz, 1H), 4.21 (t, *J*=6.2 Hz, 1H), 7.09 (d,

J=9.7 Hz, 1H), 7.21 (s, 1H), 7.32 (d, *J*=8.7 Hz, 1H), 7.51 (s, 1H); **MS(ESI):** *m/z:* (M+H) 238.96 (100%), 239.96 (12.54%), 241.02 (33.42%); **UV/λmax**: 280 nm.

5-bromotryptophan was yielded as a sandy brown crystalline solid (6.5mg, 0.023 mmol, yield: 23%). **^{1}H NMR** (D_2O): δ ppm 3.28 (dd, *J*=15.4, 7.3 Hz, 1H), 3.36 (dd, *J*=15.4, 5.3 Hz, 1H), 4.20 (t, *J*=6.3 Hz, 1H), 7.23 (s, 1H), 7.26 (d, *J*=8.7 Hz, 1H), 7.33 (d, *J*=8.7 Hz, 1H), 7.74 (s, 1H); **MS(ESI):** *m/z:* (M+H) 282.90 (100%), 283.90 (12.83%), 284.90 (95.61%), 285.90 (11.98%); **UV/λmax**: 287 nm.

7.11.5.1.2 Comparison of tryptophan synthase activity of planktonic cultures and the cell free lysate of *E. coli* PHL644

Two 500 ml cultures of *E. coli* MW-002 were grown overnight using the standard conditions of tryptophan synthase cell lysate preparation. Following overnight growth the cells were collected *via* centrifugation (3315g, 20 min, 5°C) and the supernatent removed. Each pellet was then washed with saturated sodium chloride solution (40 ml) followed by an additional centrifugation (3315g, 20 min, 5°C) to collect the cells. The supernatant was removed and the pellet re-suspended in tryptophan synthase lysis buffer (40 ml). One of the resulting cell suspensions was sonciated to generate the cell free lysate (8.20 min, 10% pulse, 49% power); the other was kept as a planktonic cell suspension.

5 ml of either the cell lysate or the cell suspension was placed inside dialysis tubing which was added to tryptophan synthase biotransformation buffer (50 ml) containing 5-bromoindole (5 mM). In addition the effect of 2.5% (v/v) DMSO or acetonitrile was examined by adding 1.75 ml of either solvent to the biotransformation reaction. The yields of the biotransformation reactions were assessed in the usual tryptophan synthase biotransformation method by extraction with ethyl acetate and purification by reverse phase chromatography.

7.11.5.1.3 Effect of buffer composition on tryptophan synthase activity of planktonic cultures of *E. coli* PHL644

Cells from a 20 ml overnight culture of *E. coli* MW-002 grown in M63 media were collected via centrifugation (1500g, 15 min, 4°C) and the media removed. The cell pellet was re-suspended into 20 ml of tryptophan synthase buffer containing either 100 mM or 10 mM KH_2PO_4 at either pH 7.0 or 7.8. 2 ml of the resulting planktonic cell suspension was added to a further 18 ml of the re-suspension buffer. The biotransformation was supplemented with 5-bromoindole (5 mM) and 5% DMSO. The reactions were incubated at for 3 days (37°C, 180 rpm). Following this a 1 ml aliquot was

taken, centrifuged (16060g, 5 min) to remove the planktonic cells. The yield of the biotransformation reaction was assessed by HPLC analysis (LCgrad_MeOH).

7.11.5.1.4 Comparison of planktonic cell biotransformations to spin-coated biofilm catalyst

10 ml of LB supplemented with ampicillin was inoculated with *E. coli* PHL644 MW-002 and incubated in an orbital shaker (37°C, 180 rpm). After the culture reached an OD_{600} of 0.6 it was diluted 100-fold into 500 ml of fresh LB supplemented with ampicillin and incubation continued. The cells were harvested by centrifugation (3315g, 20 min, 5°C) once growth had reached stationary phase (OD_{600} > 1) and the pellet was washed and re-suspended in 50 ml potassium phosphate reaction buffer.

It was determined that 12 ml of cell suspension contained approximately 140-160 mg of dry cell biomass (see determination of planktonic cell biomass, section 7.11.4). This volume of cell suspension was added to a 250 ml conical flask containing 58 ml of potassium phosphate reaction buffer, giving a final volume of 70 ml. The flask was supplemented with 0.7 M DMSO and 2 mM 5-chloroindole. The biotransformation reactions were placed into an orbital shaker incubator (30°C, 70 rpm) and incubated under the same conditions as the biofilm mediated biotransformations. 0.5 ml aliquots of the reaction buffer were taken every hour for the first 7 hours and then at regular intervals thereafter. Any reaction in the samples was stopped by centrifugation (16060g, 5 min) to collect any planktonic cells in suspension. The concentration of 5-chlorotryptophan in each of the aliquots was determined by HPLC analysis (LCgrad_MeOH).

7.11.5.2 Biofilm mediated biotransformations

7.11.5.2.1 Biofilm mediated tryptophan synthase biotransformation

The biofilm maturation M63 medium was carefully removed from a 7 day mature spin-coated biofilm covered slide using a syringe. Un-adhered, planktonic cells were removed from the biofilm by gentle re-submersion and washing in aliquots of reaction buffer (2 x 50 ml), which was then also removed. The washed biofilm slide was then submerged in 70 ml of reaction buffer supplemented with 0.7 M DMSO and either 2 mM 5-chloroindole (0.0212 g), 2 mM 5-fluoroindone (0.0189 g), or 2 mM 5-bromoindole (0.0274g). The biotransformation reactions were placed into an orbital shaker incubator (30°C, 70 rpm), set at a low speed to minimise cell shearing, and incubated for 30 hours. 0.5 ml Aliquots of the reaction buffer were taken every hour for the first 7 hours and then at regular intervals thereafter. Any reaction in the samples was stopped by centrifugation (16060g, 5 min) to

collect any planktonic cells in solution. The concentration of 5-halotryptophan in each of the aliquots was determined by HPLC analysis (LCgrad_MeOH).

7.11.5.2.2 Recycling the Biocatalytic Biofilm

Biotransformation reactions with the biofilm were set up as previously. After 10 hours reaction time the reaction mix was carefully removed using a syringe. The biofilm was carefully re-submerged in 70 ml of potassium phosphate reaction buffer which was then carefully removed again. This was repeated to remove any transformed tryptophan and un-reacted indole. Fresh reaction buffer was added (70 ml) and a fresh amount of haloindole (2 mM) added. The reaction was then resumed for another 10 hours. After this time the recycling protocol was repeated once more.

7.11.5.3 Biotransformation with Ni-NTA immobilised tryptophan synthase

Tryptophan synthase was immobilised onto the Ni-NTA resin as previously described. Having been washed with lysis and wash buffers the resin was further washed with 10 ml of reaction buffer before the resin was collected and the buffer removed. The amount of catalyst immobilised onto the resin was calculated as described earlier and the biotransformation reaction scaled down appropriately so as to maintain an enzyme concentration that was equivalent to the average calculated enzyme quantity present in the biofilm (0.6 mg/ml).

To begin the biotransformation the immobilised tryptophan synthase was re-suspended in the appropriate volume of reaction buffer suitable to the quantity of immobilised enzyme. This was supplemented with 0.7 M DMSO and 2 mM 5-chloroindole. The biotransformation was incubated in an orbital shaker (37 °C, 180 rpm) for 30 hours and aliquots taken throughout this period to monitor the reaction by HPLC analysis. After this time the resin was again collected by centrifugation (1500g, 5 min, 4°C) and the biotransformation buffer removed. The resin was washed successively with 10 ml of potassium phosphate reaction buffer, Ni-NTA lysis buffer and then Ni-NTA wash buffer. The immobilised protein was then removed from the resin by a final wash with 10 ml Ni-NTA elution buffer. The total amount of protein remaining on the resin was again measured by a Bradford assay and the relative amounts of the two subunits estimated if appropriate by SDS-PAGE analysis.

7.11.5.4 Flavin-dependent halogenase activity assays

7.11.5.4.1 Preparation of purified halogenases and flavin reductase

500 ml of LB (+ appropriate antibiotic) was inoculated with *E. coli* BL21 containing the desired halogenase construct from a starter culture in the usual way. Cultures were incubated for 4 hours in an orbital shaker incubator (37 °C, 180rpm) before being induced with 0.1 mM final concentration IPTG. Incubation was continued at 16 °C for a further 24 hours. Cell lysates were prepared by washing, re-suspending and sonicating the cultures as previously described (section 7.10.2) and Ni-NTA purification was carried out.

7.11.5.4.2 Tryptophan-7-halogenase whole-cell catalysed biotransformations

A 500 ml LB culture of *E. coli* transformed with a PrnA containing plasmid (BL21 MW-039, BL21 MW-041, PHL644 MW-042 or BL21 RG-5076) was prepared and protein production induced using the general culturing method. After growth the centrifuged bacterial pellet was washed in Ni-NTA lysis buffer (50 ml) before being pelleted again (1500g, 10 min, 4°C) and re-suspended in potassium phosphate buffer (10 mM KH_2PO_4, pH 7.2). 10 ml of this cell suspension was added into dialysis tubing which was tied at both ends. The tubing was then placed into a conical flask (250 ml) containing potassium phosphate buffer (10 mM KH_2PO_4 pH 7.2), tryptophan (1 mM) and NaCl (25 mM). The reactions were placed into an orbital shaker incubator (28°C, 180 rpm) for 24 hours. A 100 µl sample of the reaction was taken at time zero and at 2, 6 and 24 hours and the concentration of 7-chloro-tryptophan established by HPLC analysis (LCgrad_MeOH).

7.11.5.4.3 Flavin reductase assays

Flavin reductase enzyme PrnF was Ni-NTA purified from cultures of *E. coli* BL21 RG-5066. Flavin reductase enzyme solution (0.3 µM) was added to an assay containing NADH (200 µM), FAD (30 µM), NaCl (50 mM), Tris-HCl (20 mM, pH 7.5) in a total volume of 1 ml. The PrnF enzyme was added half way through the time course and the decreasing absorbance of the assay was followed at 340 nm over either 20 or 30 minutes, corresponding to the oxidation of NADH. An assay containing no FAD was used as a blank.

7.11.5.4.4 Purified tryptophan halogenase assays

Tryptophan Halogenase enzyme PrnA was purified from cultures of *E. coli* BL21 RG-5076. 100 µl reactions were carried out using Ni-NTA purified enzymes without further purification with appropriate controls. PrnA (25 µM) was added to an assay containing NADH (5 mM), FAD (10 µM), NaCl (100 mM), L-tryptophan (1 mM) and flavin reductase (2.5 µM PrnF) in Tris-HCl buffer (20 mM, pH 7.5). Protein storage buffer was used instead of protein solution for a no-enzyme control. The reactions were incubated at 30°C for 18 hours. The reactions were stopped by the addition of equal volumes of 10% formic acid and centrifuged (16060g, 15 min) before HPLC analysis (LCgrad_MeOH).

7.11.5.4.5 Preliminary purified PrnC assays

PrnC enzyme was purified from cultures of *E. coli* BL21 MW-012. 100 µl reactions were carried out using Ni-NTA purified PrnC and PrnF enzymes using a range of pyrrole as a substrate (10mM stock solutions in DMSO) with appropriate controls. 4 µM of PrnC and 1 µM of PrnF were added to an assay containing NADH (1 mM), FAD (10 µM) and Pyrrole substrate (1 mM) in Tris-HCl buffer (15 mM, pH 7.5). The effect of the addition of 10% isopropyl alcohol was also examined. A no-enzyme control contained blank protein storage buffer instead of enzyme containing buffer. The reactions were set up in 1.5 ml microcentrifuge tubes and carried out at 30 °C for 30-60 minutes. Equal volumes of 10% formic acid were added to quench the reactions and the tubes were centrifuged (16060g, 15 min). The samples were analysed by HPLC (method: LCgrad_MeOH) and LCMS (method: LC-Xbridge-Pac).

7.11.5.4.6 PrnC activity assays

PrnC was purified from cultures of *E. coli* BL21 MW-022 or MW-038. 100 µl reactions were carried out using Ni-NTA purified enzymes without further purification with appropriate negative controls. The standard PrnC assay contained NADH (5 mM), FAD (10 µM), NaCl (100 mM) in 100 µl of Tris-HCl buffer (20 mM, pH 7.5) and supplemented with flavin reductase enzyme PrnF (2 µM). Potential substrates were added to a final concentration of 1 mM. Reactions were started with the addition of the PrnC enzyme at various concentrations. Initial assays of activity used PrnC-LIC at 28 µM and PrnC-pGro at 80 µM. The standard PrnC assay utilised PrnC-LIC at a concentration of 7 µM and PrnC-pGro at 19 µM. Protein storage buffer was used instead of protein solution for no-enzyme controls. The reactions were incubated at 30°C for up to 18 hours. The reactions were stopped by the addition of equal volumes of 10% formic acid and centrifuged (16060g, 15 min) before HPLC analysis (using gradients LCgrad_MeOH or LCgrad_Pyrrole) or LCMS analysis (using gradient LC-XBridge-Pac at UEA or LC-JIC-Method1 at JIC).

7.12 Analytical methods

7.12.1 Biofilm analysis

7.12.1.1 Biofilm analysis using atomic force microscopy (AFM)

Biofilm-coated glass slides were removed from the media after 3-10 days of growth and a 7 x 7 mm area was carefully excised and firmly secured in a BioCell (JPK, Germany) containing 2 ml of M63 medium in order to perform the wet mode experiments at 30°C. The temperature in the BioCell was allowed to equilibrate at 28°C and then mounted in a NanoWizard II AFM (JPK, Germany) incorporating a CellHesion module (JPK, Germany), providing a lateral scan range of 100 µm x 100 µm, and a vertical range of 90 µm. A minimum of 10 force measurements were collected on at least three different surface locations, the peak vertical deflection of the cantilever (nm) was converted to a peak force (N) via multiplication of the deflection by the manufacturer specified cantilever spring constant of 0.9 N m^{-1}. In this study, microfabricated rectangular Si cantilevers (450 µm length) were used with pyramidal oxide sharpened tips (MikroMasch, Estonia) and a scan velocity of 5 µm s^{-1} was employed throughout. Each force measurement was repeated for five different biofilm regions on the same sample and averaged over three separately grown biofilms giving a total usually in excess of 100 force curves.

7.12.1.2 Environmental scanning electron microscopy (ESEM) and vertical scanning inteferometry (VSI)

Biofilm samples were also examined using environmental scanning electronic microscopy (ESEM). The methods used to treat the biofilm prior to ESEM and VSI imaging have been previously discussed and standardised by Bozzola and Russell (1992)[227]. The biofilm samples were immersed in 2.5% glutaraldehyde for one day in order to preserve the structure of living tissue with no alternation from the living state. After the primary fixation with glutaraldehyde, the biofilm samples were dehydrated using 50%, 70%, 90% 100% ethanol and 100% dried ethanol for 15 mins on each concentration and repeated twice. After fixation and rinsing, the samples were critical point dried using a critical point dryer (Agar Scientific) and then mounted onto microscope stubs for Pt coating using an Emscope SC 500 sputter coater (Emscope, Ashford, UK), and examined using a XL-30 FEG ESEM (Cambridge Instruments, Cambridge, UK). The Pt-coated biofilm glass slides were placed into the motorised operating stage of a MicroXAM interferometer (Omniscan, UK), which operated using a white light source and a 50X objective lens. Three-dimensional topographical surface profiles were acquired with a resolution of 0.2 µm $pixel^{-1}$, which were subsequently stitched together. Scanning

Probe Image Processor software (Image Metrology, Denmark) was employed for the analysis of acquired images. Surface profiles were acquired for five different biofilm regions on the same sample. The average roughness was calculated on pixel by pixel basis from five different sampling areas with a resolution of 419 μm x 312 μm (2095 pixels x 1560 pixels). The biofilm thickness was also measured using inteferometry. A thin line was cut through the biofilm to the glass slide using a sharp razor and the thickness was measured relative to the glass slide from 5 different sampling areas. The razor did not cut into the glass slide.

7.12.2 HPLC and LCMS Methods

7.12.2.1 LCgrad_MeOH and HPLC analysis of tryptophan and halotryptophan

HPLC samples were run on a Shimadzu HPLC with a ZORBAX (SB-C18 4.6 mm x 15 cm) column run with methanol versus water run at a rate of 0.7 ml min^{-1}. Both solvents were acidified with 0.1% formic acid and run using the gradient described in the table below.

LCgrad_MeOH HPLC gradient showing increase of mobile phase (methanol) over time.

Time (minutes)	*% Methanol*
0	10
0.5	10
12.5	95
15	95
16	10
21	10

The concentration of halo-tryptophan was calculated based on a standard curve of HPLC peak area versus concentration. Samples of pure L-halo-tryptophan (5-chloro, 5-bromo or 5-fluoro) of known concentration, ranging from 0.125 mM to 2 mM, were analysed by HPLC using the method described above. A standard curve of the relationship between HPLC peak area and concentration was constructed and the relationship between peak area and concentration was found to be linear. From this standard curve the unknown concentration of halo-tryptophan in the biotransformation samples was determined and an estimation of the overall yield was calculated. To confirm the accuracy of the HPLC concentration estimation a standard tryptophan synthase cell lysate biotransformation of 5-chlorotryptophan was performed. A sample of this reaction was analysed by HPLC and a yield calculated from the peak areas. The same sample was applied to reverse-phase silica and column purified using the tryptophan synthase method. From this the overall yield of the reaction was calculated and compared to the estimated HPLC value.

The different 5-halo-tryptophans were distinguished on the HPLC by retention time (r.t) and PDA spectrum/λ_{max} values: **5-fluoro-tryptophan** - r.t: 7.74 min, λ_{max}: 278 nm; **5-chloro-tryptophan** - r.t: 9.4 min, λ_{max}: 280 nm; **5-bromo-tryptophan** – r.t: 9.83 min, λ_{max}: 287 nm.

7.12.2.2 LCgrad_Pyrrole and HPLC analysis of PrnC substrates

HPLC samples were run on a Shimadzu HPLC with a ZORBAX (SB-C18 4.6 mm x 15 cm) column run with methanol versus water run at a rate of 0.7 ml min^{-1}. Both solvents were acidified with 0.1% formic acid and run using the gradient described in the table below.

LCgrad_Pyrrole HPLC gradient showing increase of mobile phase (methanol) over time.	
Time (minutes)	***% Methanol***
0	10
0.5	10
20.5	95
25	95
26	10
27	10

7.12.2.3 UEA LCMS analysis

Mass spectrometry at UEA was carried out using a Shimadzu single quadrupole LC-mass spectrometer equipped with an electrospray ionisation source. Separation was on a C18 XBridge column (2.1 x 100 mm, 3.5 µm, Waters) using the following gradient of acetonitrile versus 0.1% formic acid in water with a flow rate of 0.35 ml min^{-1} .

LC-XBridge-Pac gradient showing increase of mobile phase (methanol) over time.	
Time (minutes)	***%Acetonitrile***
0	10
0.5	10
9	95
11	95
11.5	2
14	2

Compounds were detected using a PDA detector and electrospray MS in positive ion mode. UV spectra were collected from 200-600nm.

7.12.2.4 JIC LCMS analysis of PrnC standard substrates and assay mixtures

Standard 5 mM solutions were made of the PrnC substrates in 5% DMSO. These were diluted 25 fold with 20% MeOH prior to analysis. The reaction mixtures were expected to be about 1mM, so were diluted ten fold with 20% MeOH. The samples were centrifuged and the supernatant transferred to glass inserts for analysis. Initial broad screening runs demonstrated no problems caused by the DMSO, but also little sign of product. Therefore new more concentrated samples were prepared by diluting original samples 3 fold for the targeted runs.

7.12.2.4.1 Broad screening method

Samples were run on a Surveyor HPLC attached to a DecaXPplus ion trap MS (both Thermo). Separation was on a 150×2mm 4µ PolarRP column (Phenomenex) running the following gradient of methanol versus 0.1% formic acid in water, at 0.3 ml min^{-1} and 30°C:

LC_JIC_Method1 gradient showing increase of mobile phase (methanol) over time.	
Time (minutes)	***% Methanol***
0	2
1	2
20	95
22	95
22.5	2
28.8	2

Detection was by UV and electrospray MS. UV spectra were collected from 200-600nm. MS analysis was performed with duplicate runs in positive and negative mode (only positive data was used for data analysis as the negative data added nothing). The instrument was set up to collect full scan data from *m/z* 100-800 and data-dependent zoom-scans and MS2 of the most abundant precursors in the range 100-400. MS2 was carried out at 35% collision energy and with an isolation width of *m/z* 4.0, with dynamic exclusion to ensure that after three spectra had been collected for a precursor ion, it would be ignored for 0.5min in favour of the next most abundant precursor ion. Spray chamber conditions were 50 units sheath gas, 5 units aux gas, 350°C capillary temperature, and 3.8kV spray voltage using a steel needle kit.

7.12.2.4.2 Targeted method

The targeted method was identical to the screening method but without zoom-scans and data-dependent MS2 scans were exchanged for targeted scans of the masses expected for the hydrogen and sodium adducts of the products. Individual methods were used for each expected starting material and product to ensure that there were enough scans across any chromatographic peak (too many different scan events slow down the instrument and make it difficult to recognise a peak as genuine when it contains only one or two scans)

.

Appendix 1: Production and Analysis of 5-halo-tryptophan Standards

Production of pure 5-halo-tryptophans

To enable the assignment of new HPLC peaks formed during the biotransformation reactions with tryptophan synthase in this study, pure quantities of the 5-series of halo-tryptophans (5-fluoro, 5-chloro and 5-bromo) were synthesised using the tryptophan synthase biotransformation method contained within this study. The halo-tryptophans were purified by reverse-phase chromatography to yield crystalline solids. These solids were characterised by ^{1}H NMR and LCMS to verify their purity so they could be used as standard samples for HPLC analysis.

Characterisation of pure 5-fluoro-tryptophan by ^{1}H NMR and LCMS and HPLC

^{1}H NMR:

(D_2O): δ ppm 3.29 (dd, *J*=15.4, 7.1 Hz, 1H), 3.36 (dd, *J*=15.4, 5.4 Hz, 1H), 4.23 (t, *J*=6.2 Hz, 1H), 6.94 (td, *J*=9.3, 2.4 Hz, 1H), 7.24 (d, *J*=2.4 Hz, 1H), 7.26 (s, 1H), 7.36 (dd, *J*=8.9, 4.5 Hz, 1H)

LCMS:

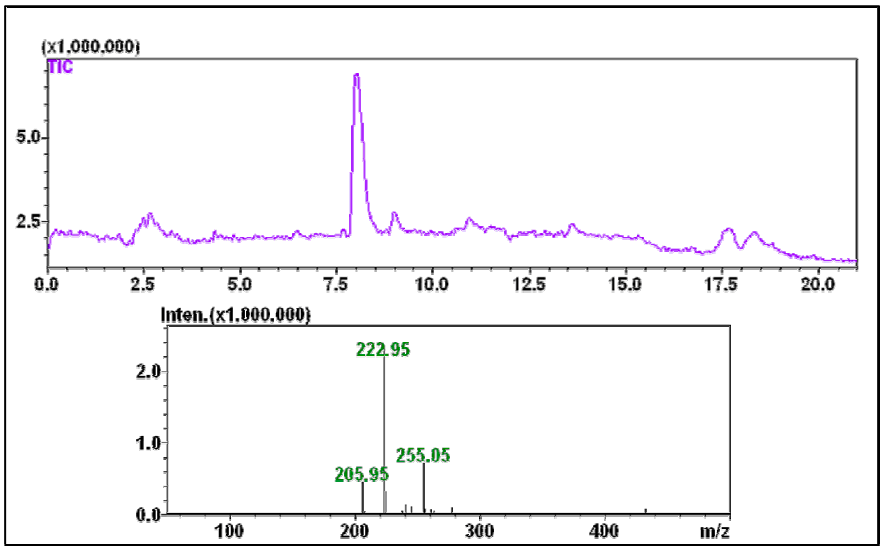


Figure A1.1: Total ion chromatogram and MS analysis of the major ion peak visible on the LCMS analysis of 5-fluoro-tryptophan

The calculated molecular mass of 5-fluoro-tryptophan ($C_{11}H_{11}N_2O_2F$) is 222.08. The total ion chromatogram for the sample of pure 5-fluoro-tryptophan showed a single peak at around 7.7 minutes (figure A1.1). This peak corresponded to a major ion of *m/z* 222.95 which represents the hydrogen adduct of the product (M+H).

HPLC:

Analysis of the 5-fluoro-tryptophan product by HPLC (LC_gradMeOH) gave a clear peak with a retention time of 7.6 minutes (figure A1.2). The PDA spectrum of the compound shows a λ_{max} value of 278 nm. The retention time and the PDA spectrum can be used to identify 5-fluoro-tryptophan in an unknown sample.

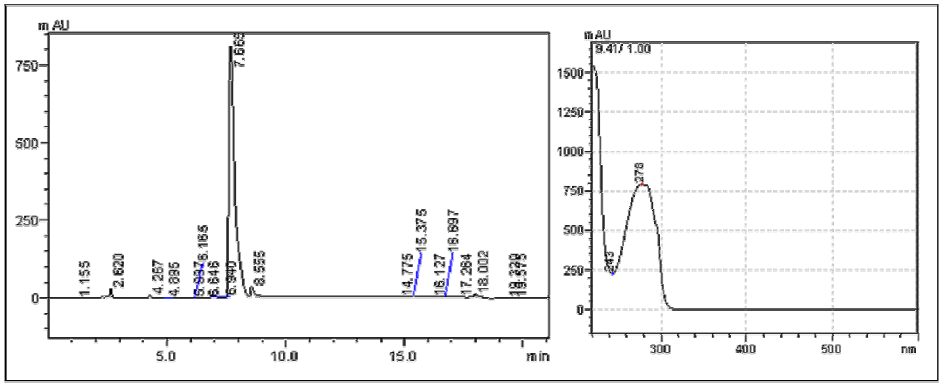


Figure A1.2: HPLC trace and PDA spectrum of 5-fluoro-tryptophan showing the λ_{max} as 278

The concentration of 5-fluoro-tryptophan in an unknown sample can be calculated from the HPLC peak area if a number of standard solutions of 5-fluoro-tryptophan of known concentrations (0.125-5 mM) are analysed. Plotting the resulting integrated peak areas versus concentration enables the construction of a standard curve with a gradient of y=5E+06x (figure A1.3). This can be used to estimate 5-fluoro-tryptophan concentration in an unknown sample. For increased accuracy the peak areas are calculated at the specific λ_{max} of 278 nm.

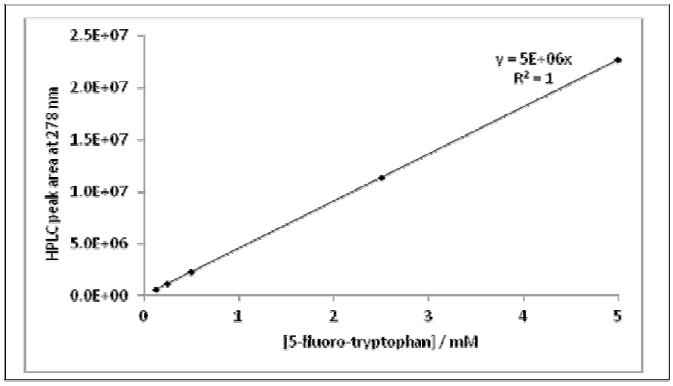


Figure A1.3: Standard curve constructed of 5-fluoro-tryptophan concentration versus HPLC peak area at 278 nm.

Characterisation of pure 5-chloro-tryptophan by ^{1}H NMR and LCMS and HPLC

^{1}H NMR:

(D_2O): δ ppm 3.24 (dd, *J*=15.2, 7.1 Hz, 1H), 3.31 (dd, *J*=15.4, 5.4 Hz, 1H), 4.21 (t, *J*=6.2 Hz, 1H), 7.09 (d, *J*=9.7 Hz, 1H), 7.21 (s, 1H), 7.32 (d, *J*=8.7 Hz, 1H), 7.51 (s, 1H)

LCMS:

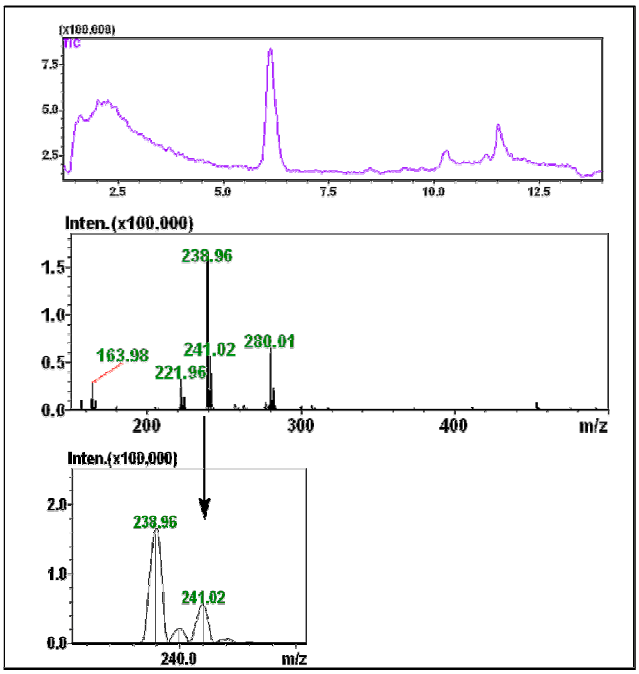


Figure A1.4: Total ion chromatogram and MS analysis of the major total ion peak visible on the LCMS analysis of 5-fluoro-tryptophan. The isotope pattern of the major ion is shown.

The calculated molecular mass of 5-chloro-tryptophan ($C_{11}H_{11}N_2O_2Cl$) is 238.67. The total ion chromatogram for the sample of pure 5-chloro-tryptophan showed a single peak at around 6 minutes (figure A1.4). This peak corresponded to a major ion of *m/z* 238.96 (100%) which represents the hydrogen adduct of the product (M+H). The isotope pattern of this ion strongly suggests the incorporation of chlorine with two isotope peaks at 238.96 (100%) and 241.02 (33.42%) which correspond to the 35C and 37C isotopes of chlorine which have relative abundances of 100 and 32.1%.

HPLC:

Analysis of the 5-chloro-tryptophan product by HPLC (LC_gradMeOH) gave a clear peak with a retention time of 9.3 minutes (figure A1.5). The PDA spectrum of the compound shows a λ_{max} value of 280 nm. The retention time and the PDA spectrum can be used to identify 5-chloro-tryptophan in an unknown sample.

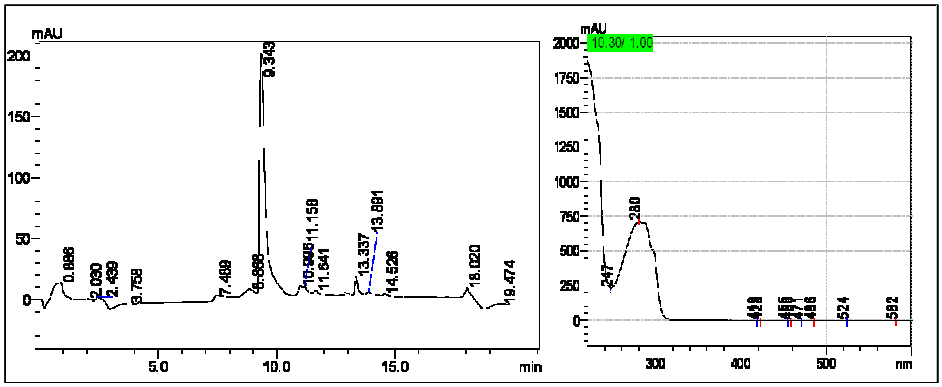


Figure A1.5: HPLC trace and PDA spectrum of 5-chloro-tryptophan showing the λ_{max} as 280

The concentration of 5-chloro-tryptophan in an unknown sample can be calculated from the HPLC peak area if a number of standard solutions of 5-chloro-tryptophan of known concentrations (0.125 – 2.5 mM) are analysed. Plotting the resulting integrated peak areas versus concentration enables the construction of a standard curve with a gradient of y=5E+06x (figure A1.6). This can be used to estimate 5-fluoro-tryptophan concentration in an unknown sample. For increased accuracy the peak areas are calculated at the specific λ_{max} of 278 nm.

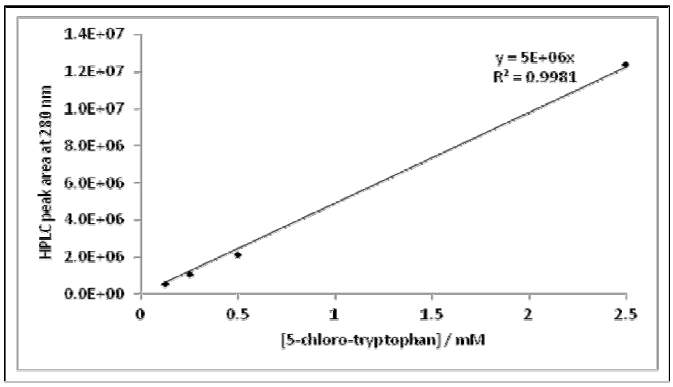


Figure A1.6: Standard curve constructed of 5-chloro-tryptophan concentration versus HPLC peak area at 280 nm.

Characterisation of pure 5-bromo-tryptophan by ^{1}H NMR and LCMS and HPLC

^{1}H NMR:

(D_2O): δ ppm 3.28 (dd, *J*=15.4, 7.3 Hz, 1H), 3.36 (dd, *J*=15.4, 5.3 Hz, 1H), 4.20 (t, *J*=6.3 Hz, 1H), 7.23 (s, 1H), 7.26 (d, *J*=8.7 Hz, 1H), 7.33 (d, *J*=8.7 Hz, 1H), 7.74 (s, 1H)

LCMS:

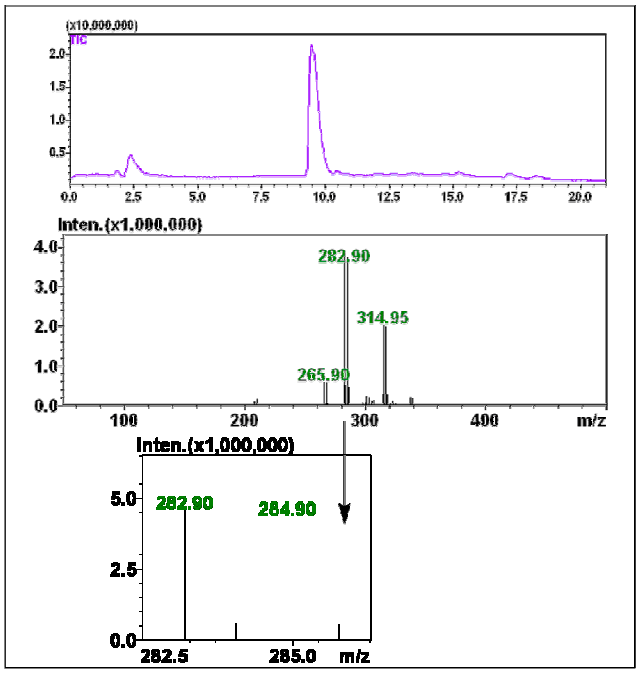


Figure A1.7: Total ion chromatogram and MS analysis of the major total ion peak visible on the LCMS analysis of 5-bromo-tryptophan. The isotope pattern of the major ion is shown.

The calculated molecular mass of 5-bromo-tryptophan ($C_{11}H_{11}N_2O_2Br$) is 282.0. The total ion chromatogram for the sample of pure 5-bromo-tryptophan showed a single peak at around 10 minutes (figure A1.7). This peak corresponded to a major ion of *m/z* 282.90 (100%) which represents the hydrogen adduct of the product (M+H). The isotope pattern of this ion strongly suggests the incorporation of bromine with two isotope peaks at 282.90 (100%) and 284.90 (95.61%) which correspond to the 79Br and 81Br isotopes of bromine which share almost the same abundance and have relative abundances of 100 and 95.61% each.

HPLC:

Analysis of the 5-bromo-tryptophan product by HPLC (LC_gradMeOH) gave a clear peak with a retention time of 9.59 minutes (figure A1.7). The PDA spectrum of the compound shows a λ_{max} value of 287 nm. The retention time and the PDA spectrum can be used to identify 5-bromo-tryptophan in an unknown sample.

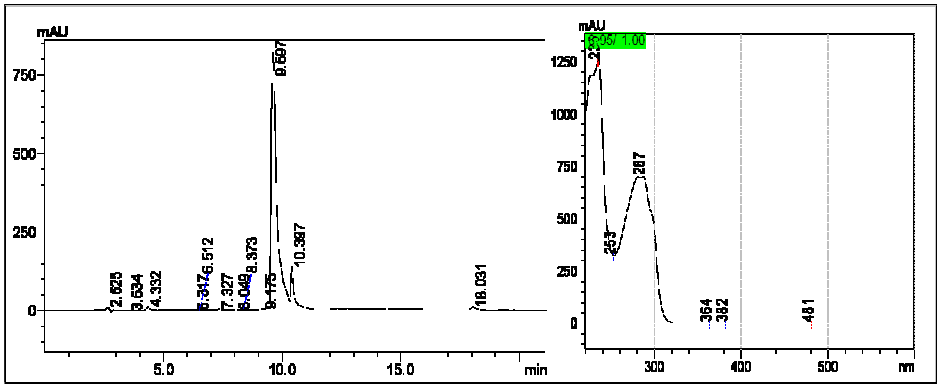


Figure A1.8: HPLC trace and PDA spectrum of 5-bromo-tryptophan

The concentration of 5-bromo-tryptophan in an unknown sample can be calculated from the HPLC peak area if a number of standard solutions of 5-bromo-tryptophan of known concentrations (0.125 – 2.5 mM) are analysed. Plotting the resulting integrated peak areas versus concentration enables the construction of a standard curve with a gradient of y=5E+06x (figure A1.6). This can be used to estimate 5-bromo-tryptophan concentration in an unknown sample. For increased accuracy the peak areas are calculated at the specific λ_{max} of 287 nm.

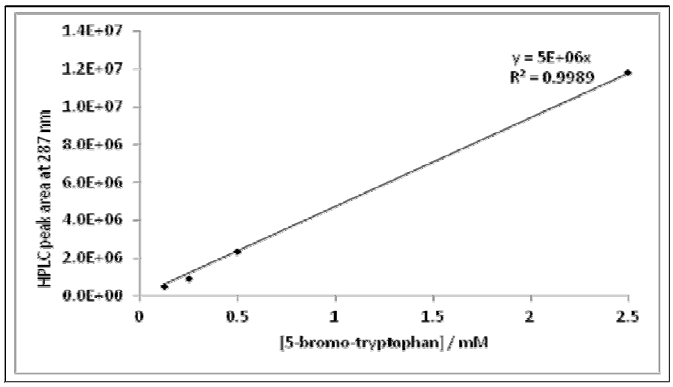
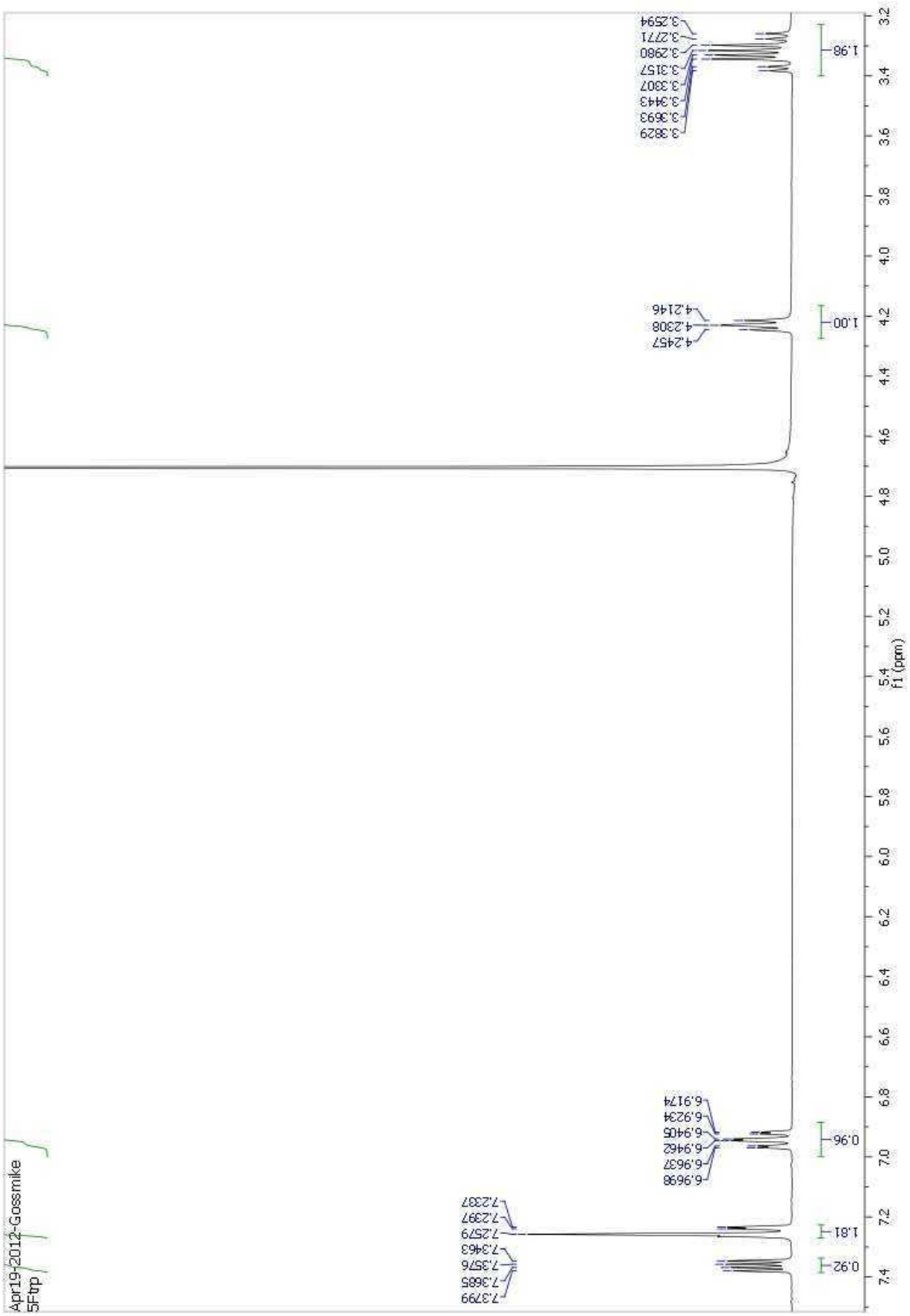
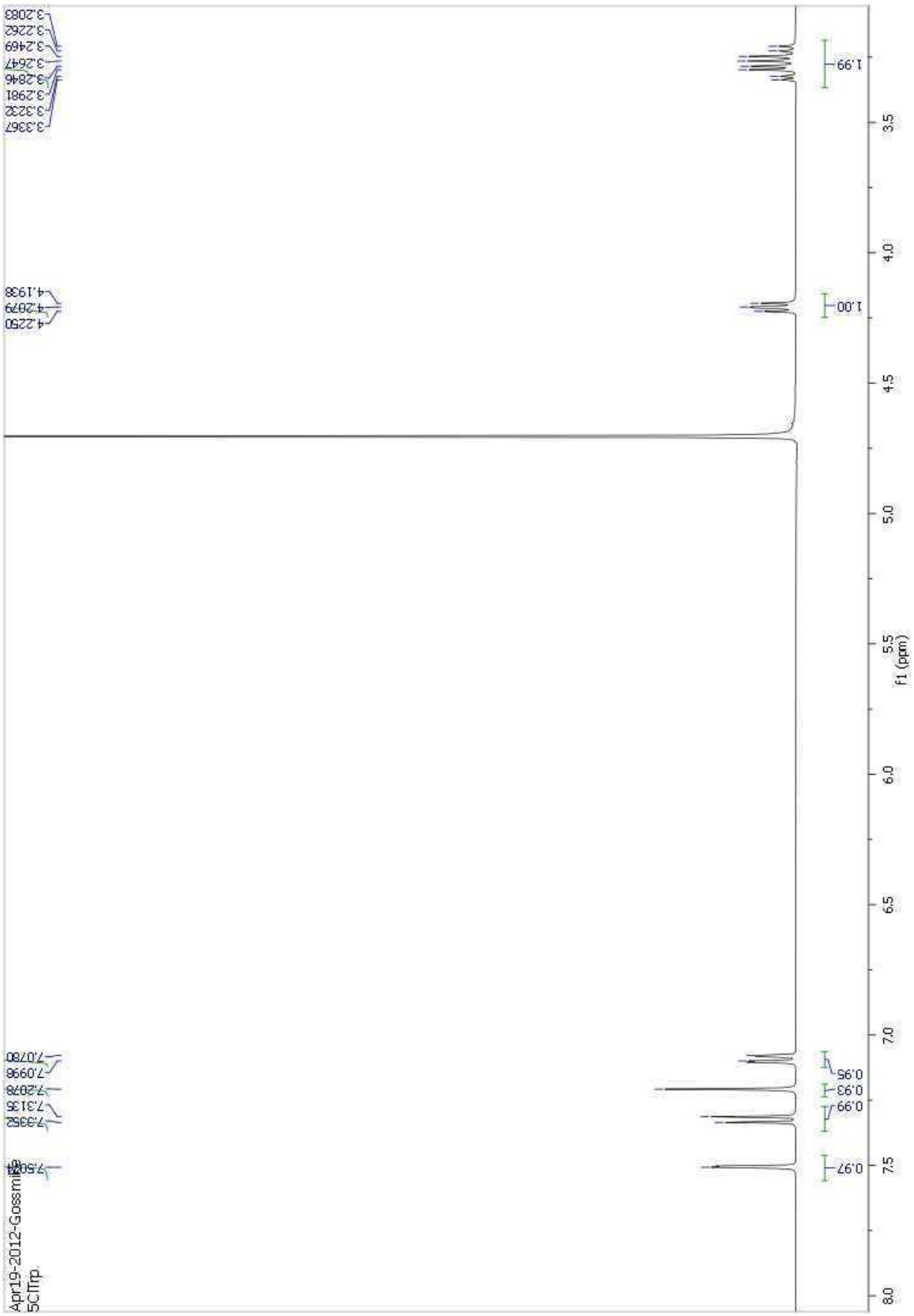


Figure A1.9: Standard curve constructed of 5-bromo-tryptophan concentration versus HPLC peak area at 278 nm.

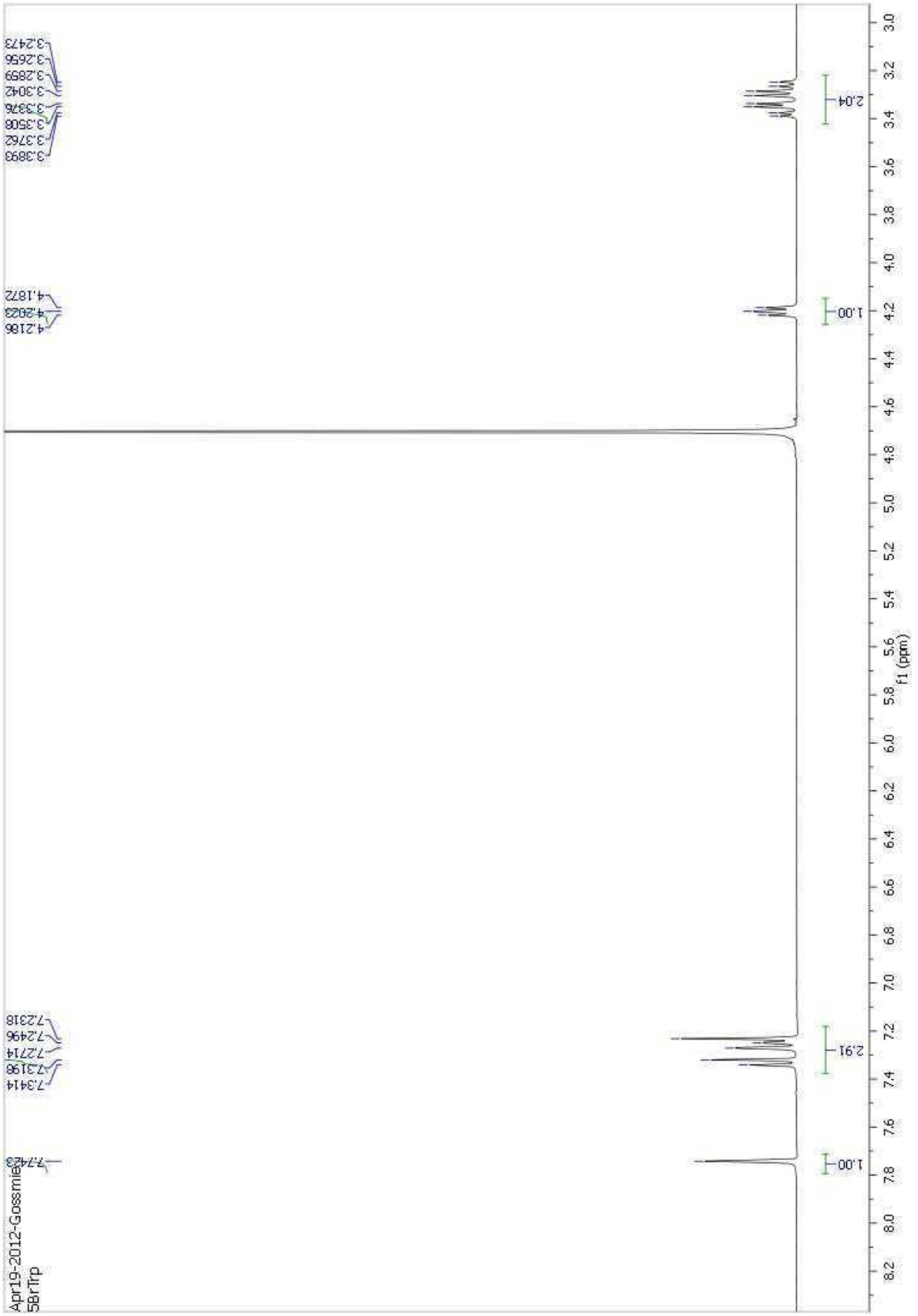
^{1}H NMR of 5-fluoro-tryptophan



^{1}H NMR of 5-chloro-tryptophan



^{1}H NMR of 5-bromo-tryptophan



Appendix 2: Initial PrnC Assays with Variety of Pyrrole Substrates

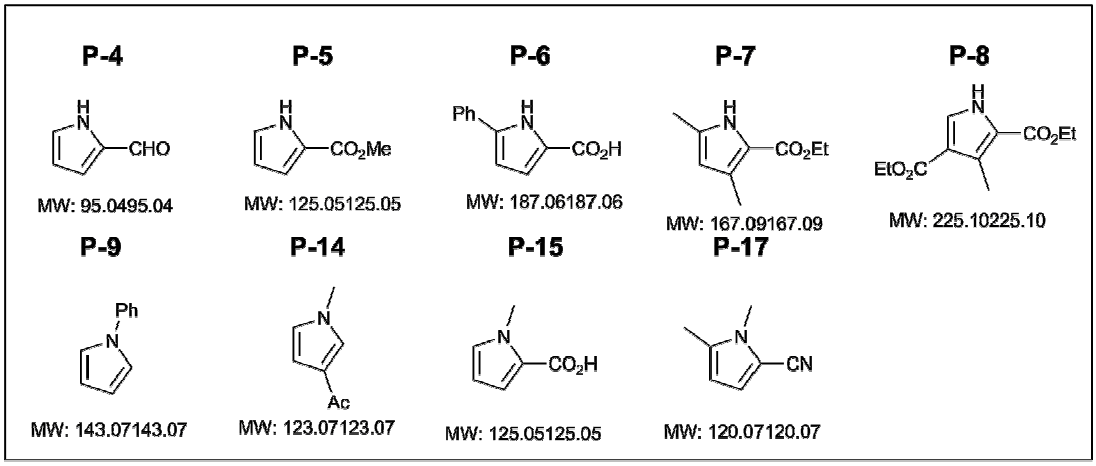
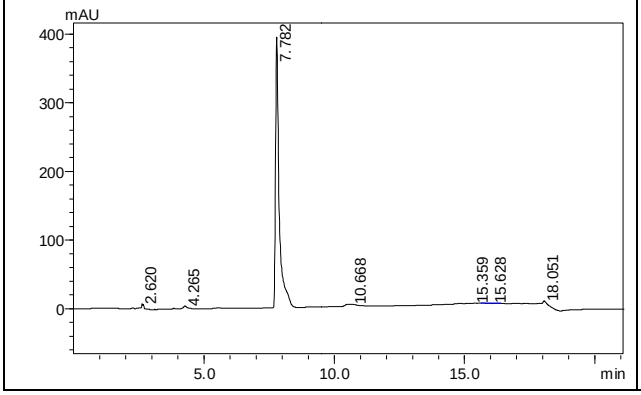
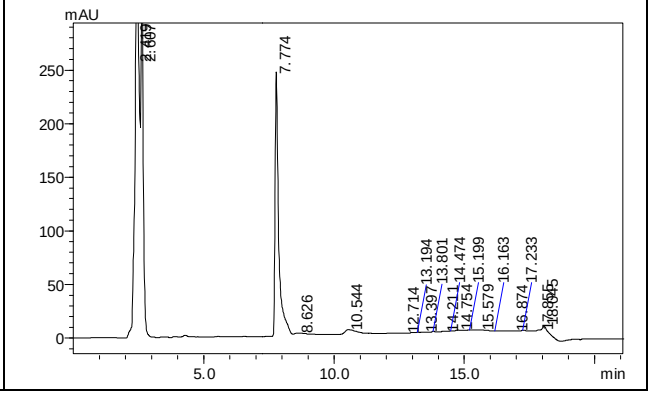
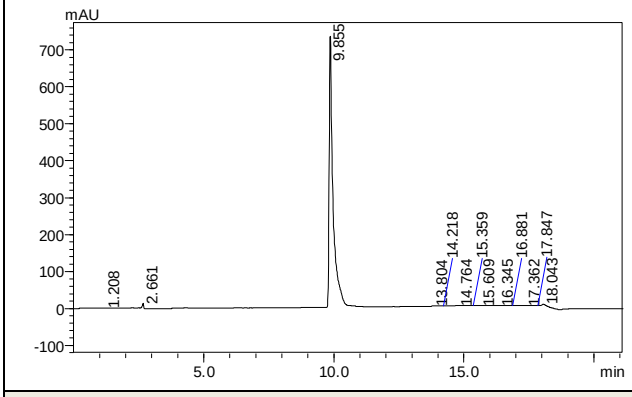
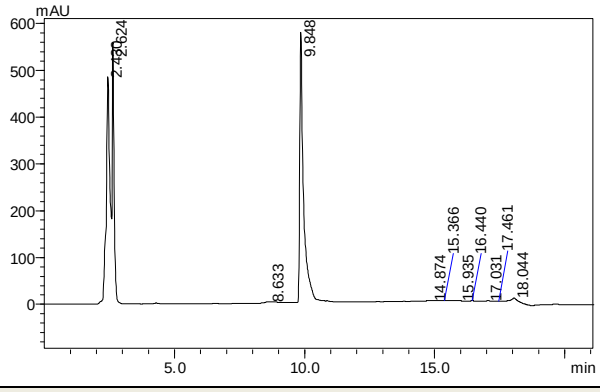
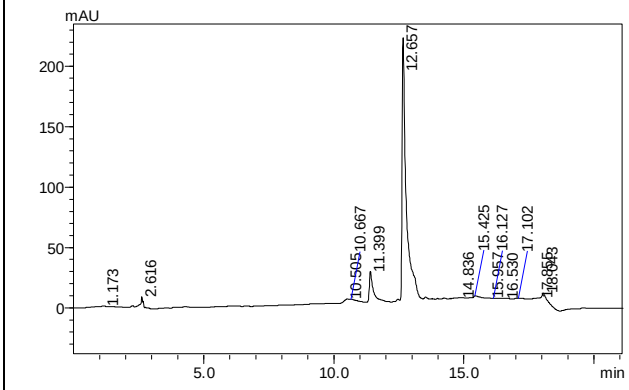
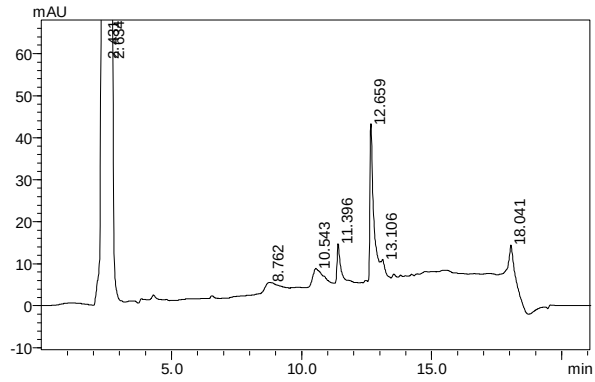
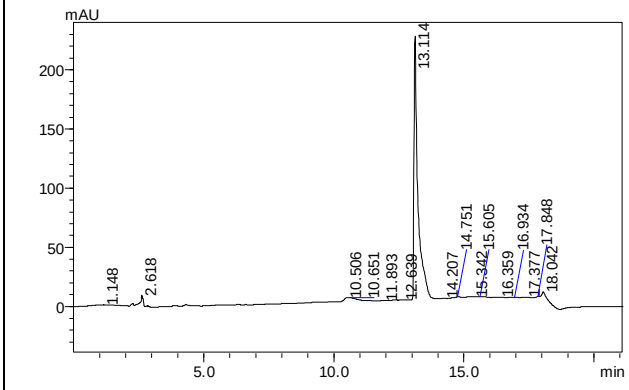
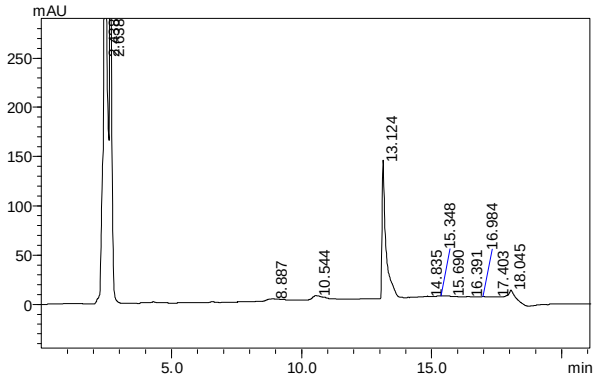
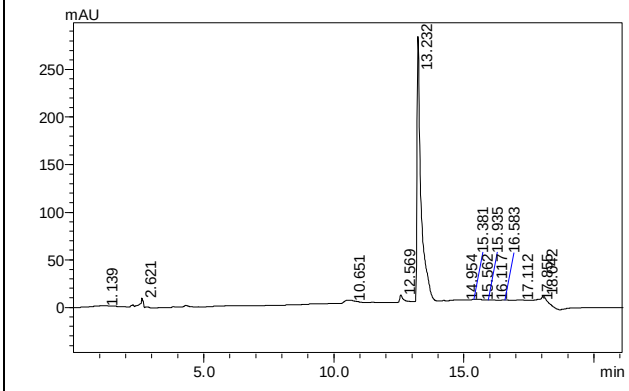
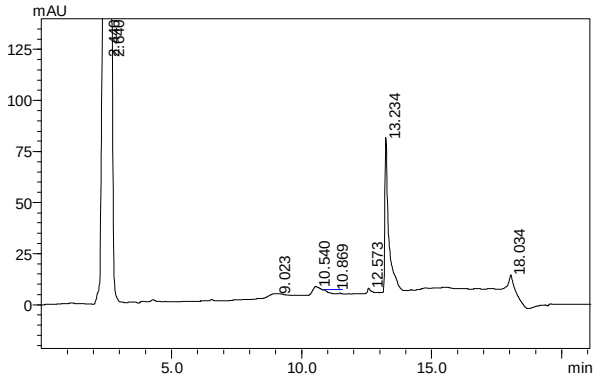


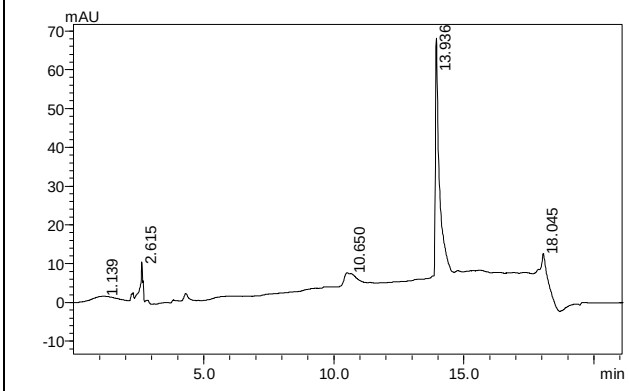
Figure A2.1: Range of pyrrole substrates tested with purified PrnC enzyme from chapter 5.2.3.2.1.

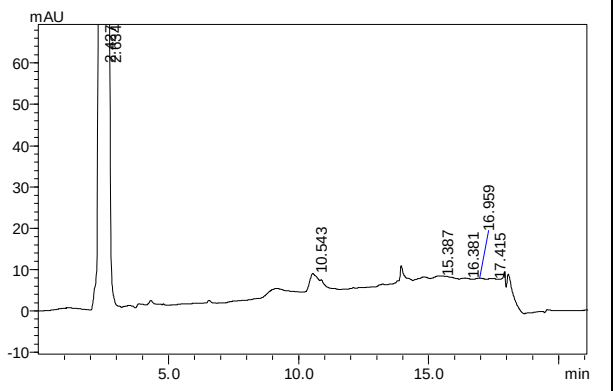
As described in chapter 5.2.3.2.2 a selection of available pyrrole derivatives were trialled as possible substrates for PrnC. The candidate pyrroles were examined for how easy they were to detect with the HPLC as this would be the method used to determine product formation. Nine pyrrole related compounds were found to be amenable for this process as they produced clear peaks on the HPLC (figure A2.1). No evidence of chlorinated product could be detected with any of the pyrroles tested. The table below shows HPLC traces of the standard compounds and the reaction mixture. The large peak at 2.5 min in assay mixtures is attributed to the NADH cofactor.

HPLC trace of starting material	HPLC trace of assay mixture
Pyrrole **P4**	
	

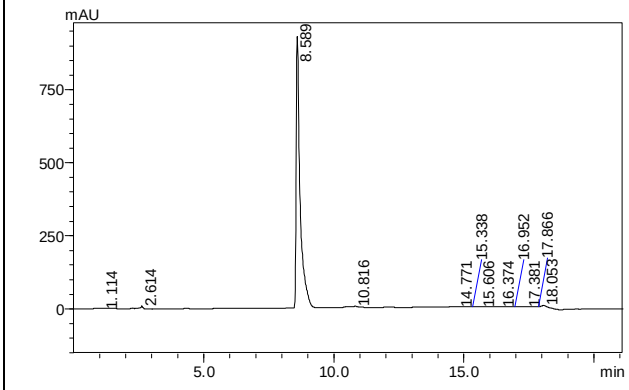
HPLC trace of starting material	HPLC trace of assay mixture
Pyrrole **P5**	
	
Pyrrole **P6**	
	
Pyrrole **P7**	
	
Pyrrole **P8**	
	

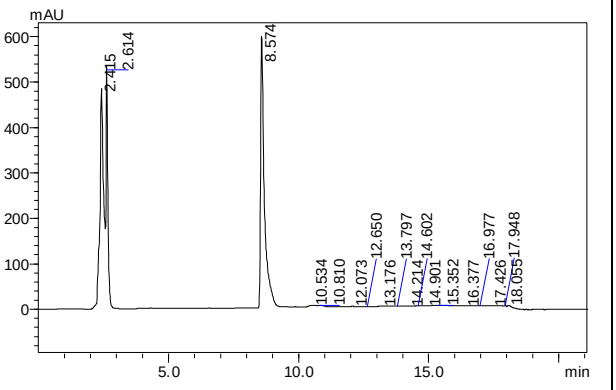
Pyrrole **P9**



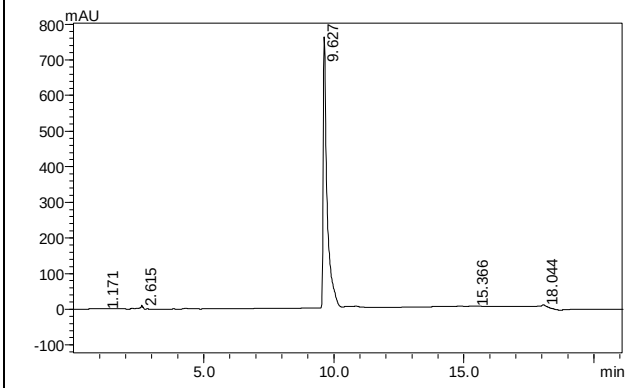


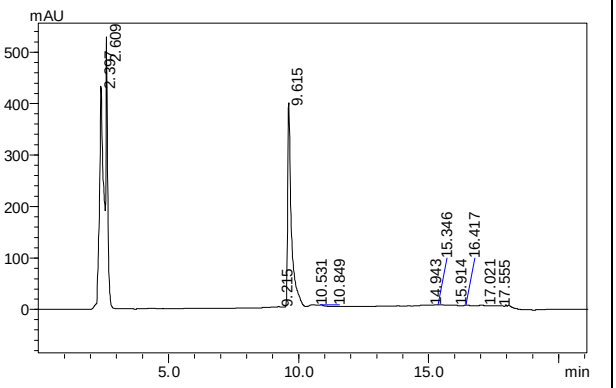
Pyrrole **P14**



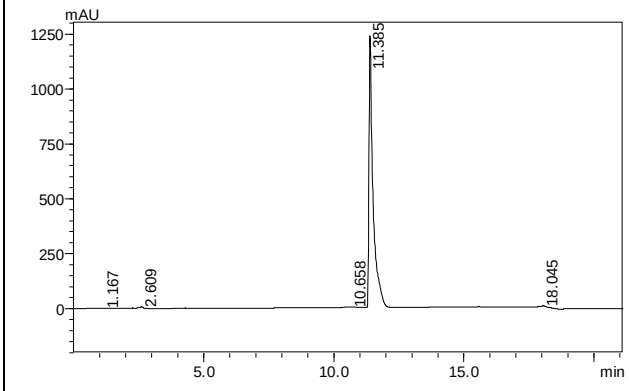


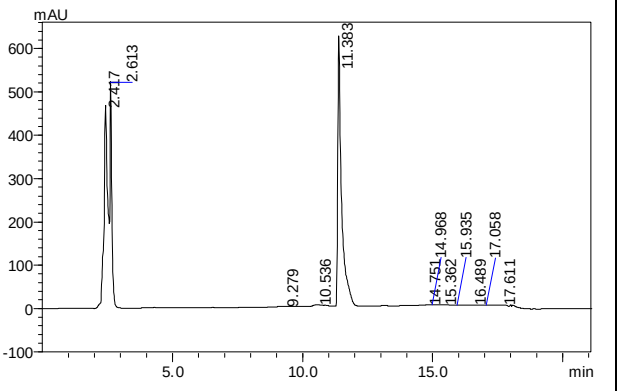
Pyrrole **P15**





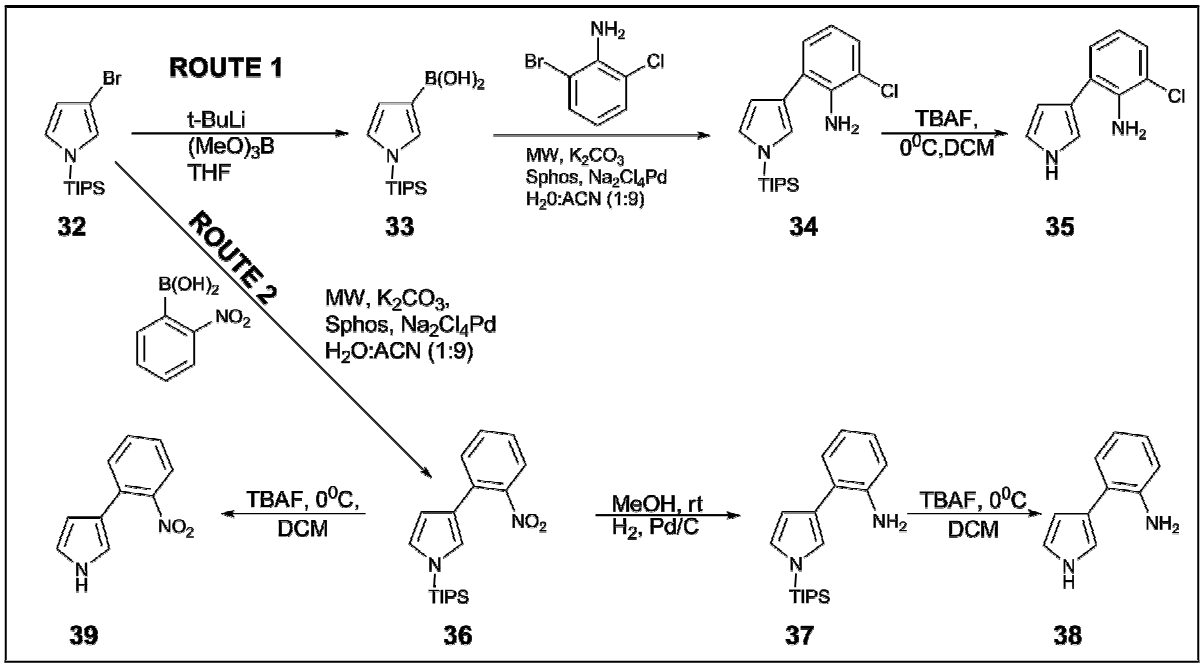
Pyrrole **P17**





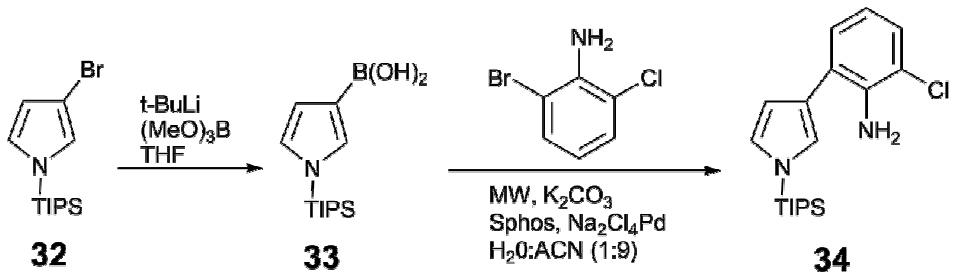
Appendix 3: Chemical Synthesis of PrnC Substrate Analogues

The following synthesis was carried out by Antoine Abou Fayad with the aim of providing compounds for the assessment of the activity of PrnC.



Scheme A3.1: General scheme of the synthetic routes to the PrnC natural substrate mono-dechloro-aminopyrrolnitrin 35 and unnatural analogues 38 and 39.

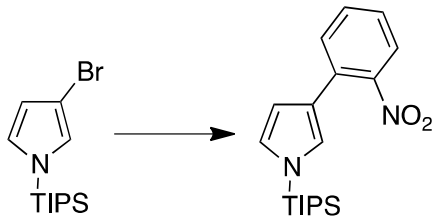
(4-(2-amino-3-chloro-phenyl)-pyrrole) (monodechloroaminopyrrolnitrin)



The protected, brominated pyrrole, **32**, was converted into the boronic acid. Suzuki cross-coupling with 2-bromo-6-chloroaniline resulted in the *N*-TIPS protected monodechloroaminopyrrolnitrin, **34**. The TLC of this reaction showed several spots and the purification could not be accomplished without decomposition. Therefore the crude material was taken directly through the next 2 steps before purification.

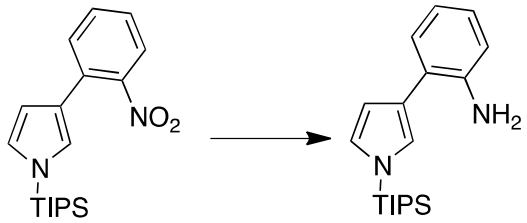
(1-(triisopropylsilyl)-1*H*-pyrrol-3-yl)boronic acid (**33**) (200mg), Na_2Cl_4Pd (4 mol%), Sphos (2.5 equiv. to Pd catalyst), K_2CO_3 (5 equiv.), and 2-bromo-6-chloroaniline (1.2 equiv.) were placed in a microwave vial and purged with N_2. Degassed water:acetonitrile (1:9) and a magnetic stirrer bar. The reaction was placed into the microwave. Continuous microwave irradiation of 250 W was used for 90 minutes; the temperature was increased from room temperature to 90°C and then kept constant for the duration. The mixture was allowed to cool to room temperature. TLC (Solid phase: silica on aluminum plate, eluent: hexane: ethyl acetate 9:1, visualization: UV, iodine) showed the formation of compound (**34**) with an Rf of 0.628. The reaction mixture was diluted in ethyl acetate (30ml) and extracted with saturated sodium bicarbonate (2x20ml). The organic layer was dried over magnesium sulfate, filtered, and then concentrated. This yielded a compound that did not match the desired target after full characterization.

4-(2-nitrophenyl)-1-(triisopropylsilyl)-1*H*-pyrrole



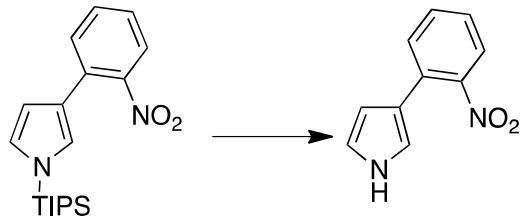
4-bromo-1-(triisopropylsilyl)-1*H*-pyrrole (**32**) (200mg, 0.66mmol), Na_2Cl_4Pd (4 mol%), Sphos (2.5 equiv. to Pd catalyst), K_2CO_3 (5 equiv.), and 2-nitrobenzene boronic acid (1.2 equiv.) were placed in a microwave vial and purged with N_2 . Degassed water:acetonitrile (1:9) and a magnetic stir bar. The vial was then placed into the microwave. Continuous microwave irradiation of 250 W was used for 90 minutes, the temperature increasing from room temperature to 90°C and then kept constant for the duration. The mixture was then allowed to cool to room temperature. TLC (Solid phase: silica on aluminum plate, eluent: hexane: ethyl acetate 9:1, visualization: UV, iodine) was taken and showed the formation of compound (**36**) with an Rf of 0.628. Afterwards, reaction mixture was diluted in ethyl acetate (30ml) and extracted with saturated solution of sodium bicarbonate (2x20ml). The organic layer was dried over magnesium sulfate, filtered, and then concentrated. The crude material was purified using column chromatography (Solid phase: silica, eluent (hexane: ethyl acetate 13:1 with 5% triethylamine) affording 4-(2-nitrophenyl)-1-(triisopropylsilyl)-1*H*-pyrrole (**36**) as a thick dark yellow oil (156mg, 0.453mmol, yield: 68%). ^{1}H NMR ($(CD_3)_2CO$): δ ppm 1.132 (d, *J* =7.4, 18H), 1.547 (dq, *J* = 7.7, 15.3, 3H), 6.413 (dd, *J* = 1.6, 2.6, 1H), 6.908 (t, *J* = 2.3, 1H), 7.056 (t, *J* = 1.5,1H), 7.407 (ddd, *J* = 1.5, 7.6, 8.0, 1H), 7.584 (td, *J* = 1.5, 8.0, 1H), 7.664 (ddd,*J* = 1.1, 3.1, 4.6, 1H); ^{13}C NMR ($(CD_3)_2CO$): δ ppm 151.278, 133.202, 132.271, 130.979, 128.235,127.147, 124.716, 124.632, 122.742, 111.943, 19.081, 13.232; HRMS (ESI) *m/z*: calculated for $C_{19}H_{28}N_2O_2Si$: 344.1920. Found 344.1927.

2-(1-(triisopropylsilyl)-1*H*-pyrrol-3-yl)aniline



4-(2-nitrophenyl)-1-(triisopropylsilyl)-1*H*-pyrrole (50mg, 0.145mmol) was dissolved in methanol (5ml) in a 25ml round bottom flask and Pd-C (20% w/w) was carefully added to the solution (under hydrogen atmosphere with a pressure of 1 atm). the reaction mixture was left stirring for 2 hours at room temperature. A TLC (Solid phase: silica on aluminum plate, eluent: hexane: ethyl acetate 9:1, visualization: UV, iodine) was taken after the 2 hours and it showed the formation of compound (**37**) with an Rf of 0.510. The reaction mixture was filtered through a plug of celite (eluent: warm methanol). The solution was concentrated and purified using column chromatography (Solid phase: silica, eluent(hexane: ethyl acetate 13:1 with 5% triethylamine) affording 2-(1-(triisopropylsilyl)-1*H*-pyrrol-3-yl)aniline (35mg, 0.111mmol, yield: 76%) as a yellow oil. . ^{1}H NMR ($(CD_3)_2CO$):δ 1.162 (d, *J* =8.54 Hz, 18H), 1.582 (td, *J* = 14.86, 7.82 Hz, 3H), 6.514 (dd, *J* = 2.74, 1.60 Hz, 1H), 6.632 (d, *J* =1.10 Hz, 1H), 6.734 (m, 1H), 6.946 (m, 2H), 7.041 (s, 1H), 7.172 (dd, *J* = 7.67, 1.35 Hz,1H); ^{13}C NMR ($(CD_3)_2CO$): δ ppm 143.603, 129.624, 126.865, 124.634, 123.817, 122.512,122.227, 118.468, 115.421, 110.634, 17.821, 11.635; HRMS (ESI) *m/z*: calculated for $C_{19}H_{30}N_2Si$: 314.2178. Found 314.2187.

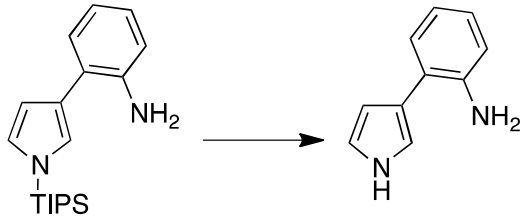
4-(2-nitrophenyl)-1*H*-pyrrole:



4-(2-nitrophenyl)-1-(triisopropylsilyl)-1*H*-pyrrole (100mg, 0.29mmol) was dissolved in anhydrous THF (20ml) in a 50ml round bottom flask that was wrapped in aluminum foil due to the light sensitivity. Tetra-*n*-butylammonium fluoride (TBAF) (2equiv.) was added to the reaction and mixture was left stirring at room temperature for 30 minutes. A TLC (Solid phase: silica on aluminum plate, eluent: hexane: ethyl acetate 4:1, visualization: UV, iodine) was taken showing formation of compound (**38**) with an Rf value of 0.16 and no trace of starting material was found. Reaction was quenched with aqueous ammonium chloride solution (5ml). Afterwards, reaction mixture was diluted in ethyl acetate (30ml) and extracted with aqueous ammonium chloride solution (10ml) three times. Organic

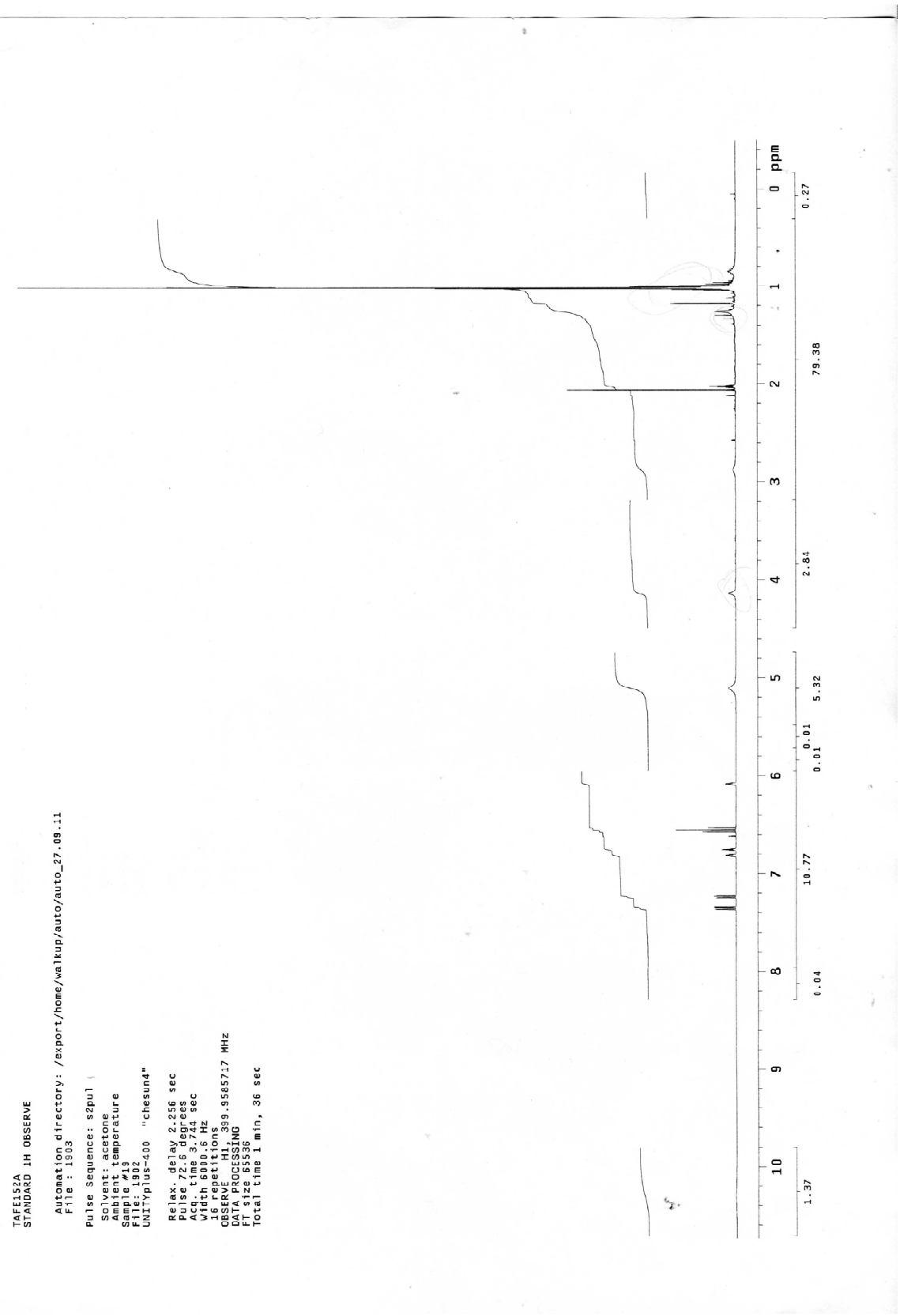
layer was then dried over magnesium sulfate, filtered, and then concentrated. The crude material was filtered over a silica plug (eluent: 6:1 hexane: ethyl acetate (100ml) followed by 1:1 hexane: ethyl acetate (100ml)), affording 4-(2-nitrophenyl)-1*H*-pyrrole as a dark yellow oil. ^{1}H NMR ($(CD_3)_2CO$): 6.252 (dd, *J* = 1.7, 2.5, 1H) , 6.872 (dd, *J* = 1.8, 2.7, 1H), 7.053 (d, *J* = 1.6, 1H), 7.382 (ddd, *J* = 1.5, 7.5, 8.0, 1H), 7.571 (ddd, *J* = 1.2, 7.6, 8.5, 1H), 7.634 (m, 1H), 10.28 (s, 1H); ^{13}C NMR ($(CD_3)_2CO$) δ ppm 142.782, 141.823, 129.356, 126.513, 125.079, 121.254, 117.756, 116.886, 113.163, 101.895; HRMS (ESI) *m/z*: calculated for $C_{10}H_8N_2O_2$: 188.06. Found 188.045

2-(1*H*-pyrrol-3-yl)aniline

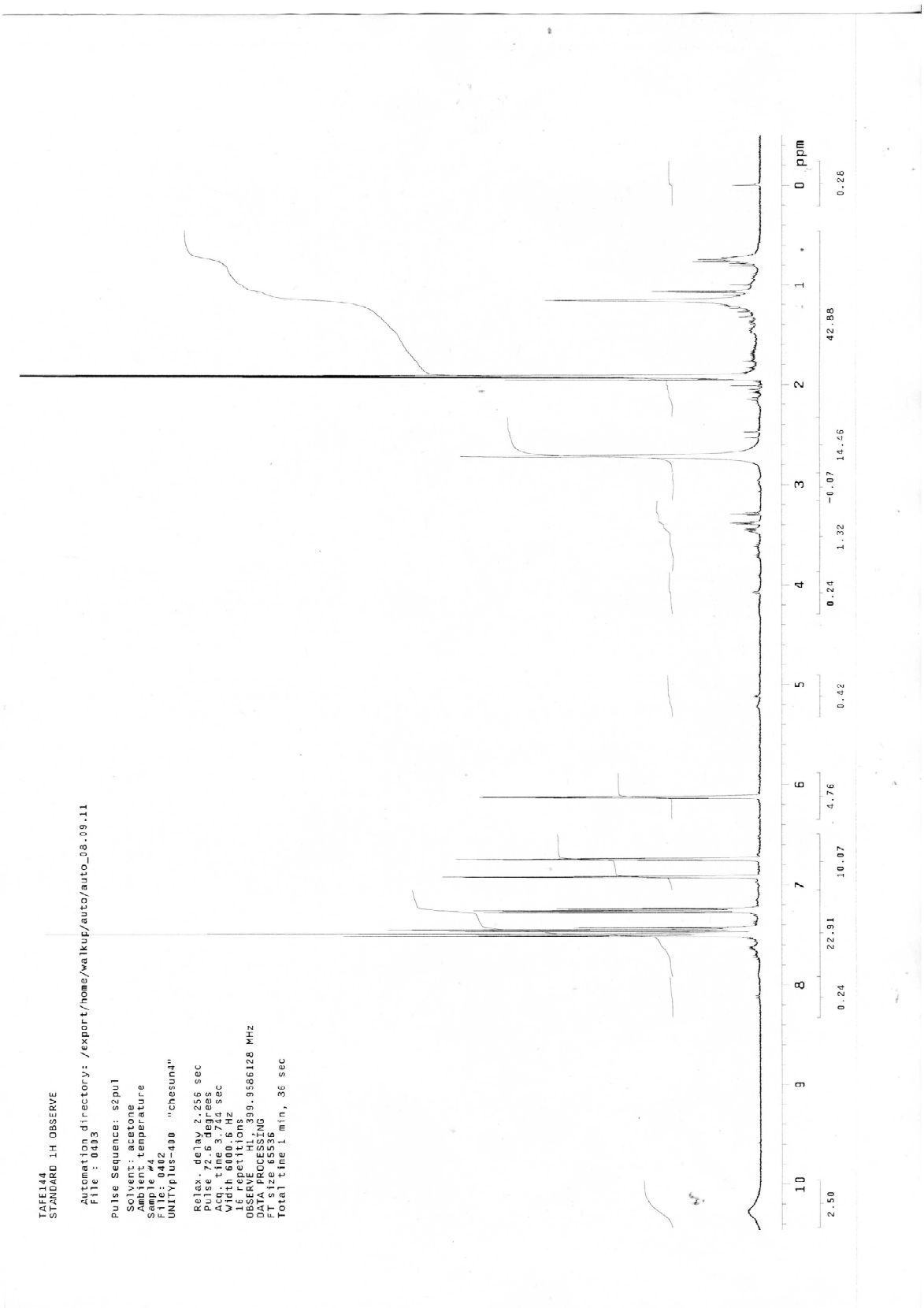


2-(1*H*-pyrrol-3-yl)aniline was prepared same way as described above to yield compound (**39**) with an Rf of 0.21 as a yellow sticky oil. ^{1}H NMR ($(CD_3)_2CO$):δ 6.343 (dd, *J* =2.5, 4.1, 1H), 6.622 (td, *J* = 1.2, 7.4, 1H), 6.746 (dd, *J* = 1.0, 7.9, 1H), 6.878 (dd, *J* = 2.6, 4.7, 1H), 6.941 (td, *J* = 1.5, 7.9, 1H), 7.021(dd, *J* = 1.8, 4.1, 1H), 7.149(dd, *J* = 1.5, 7.5, 1H); ^{13}C NMR ($(CD_3)_2CO$) δ ppm 146.681, 131.029, 128.184, 123.121, 123.594, 120.142, 119.096, 117.772, 116.920, 109.454; HRMS (ESI) *m/z*: calculated for $C_{10}H_{10}N_2$: 158.0844. Found 158.0847.

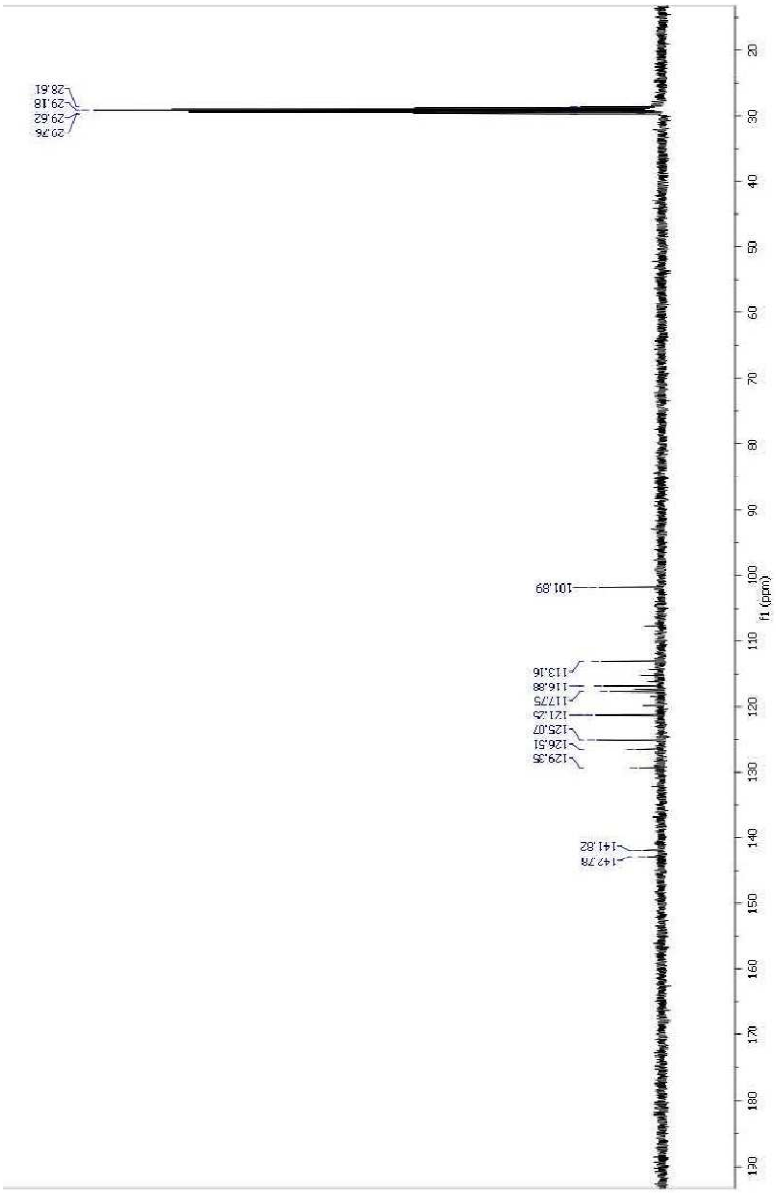
^{1}H NMR of compound 35 (4-(2-amino-3-chloro-phenyl)-pyrrole) (monodechloroaminopyrrolnitrin)



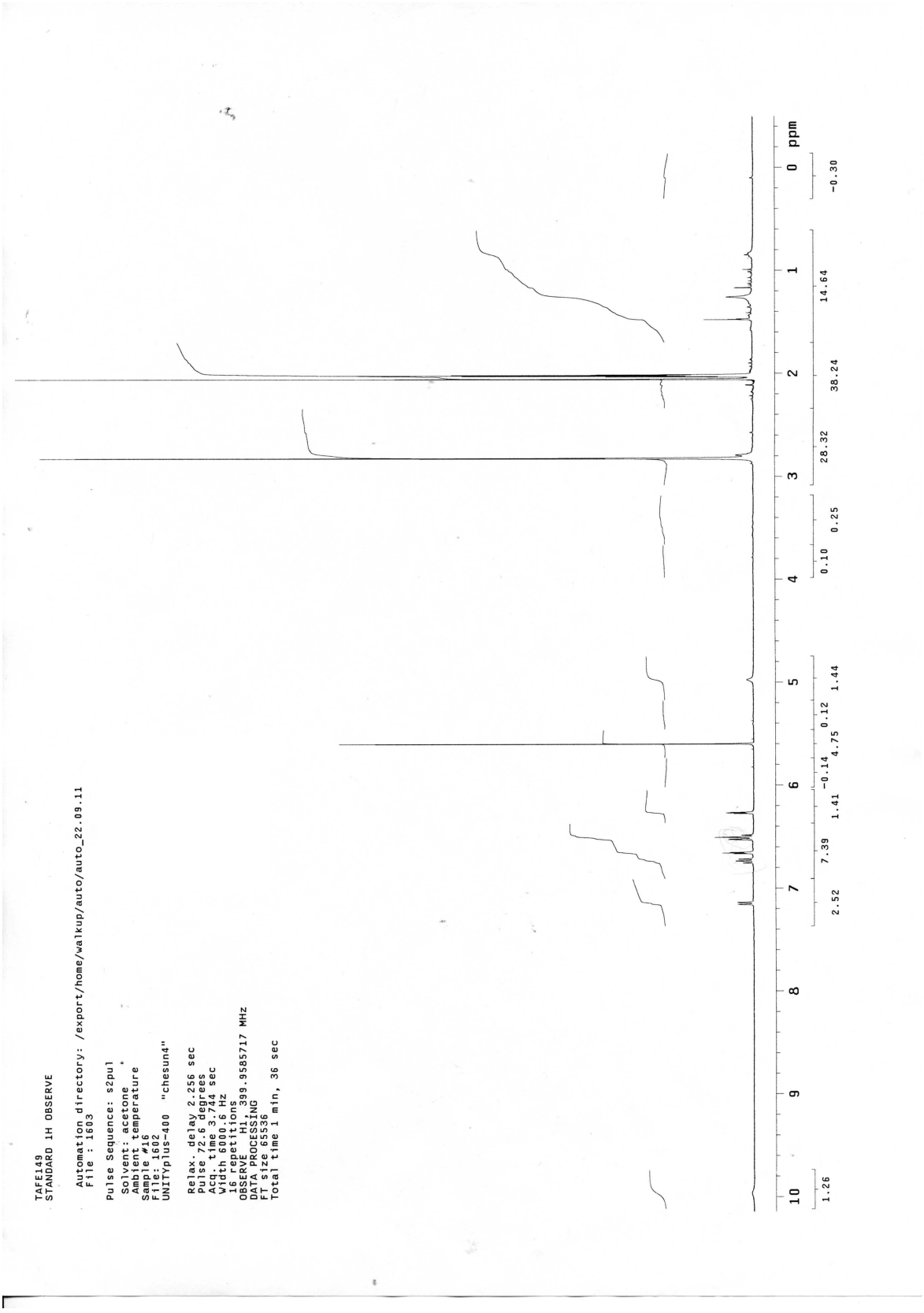
^{1}H NMR of compound 39 4-(2-nitrophenyl)-1*H*-pyrrole:



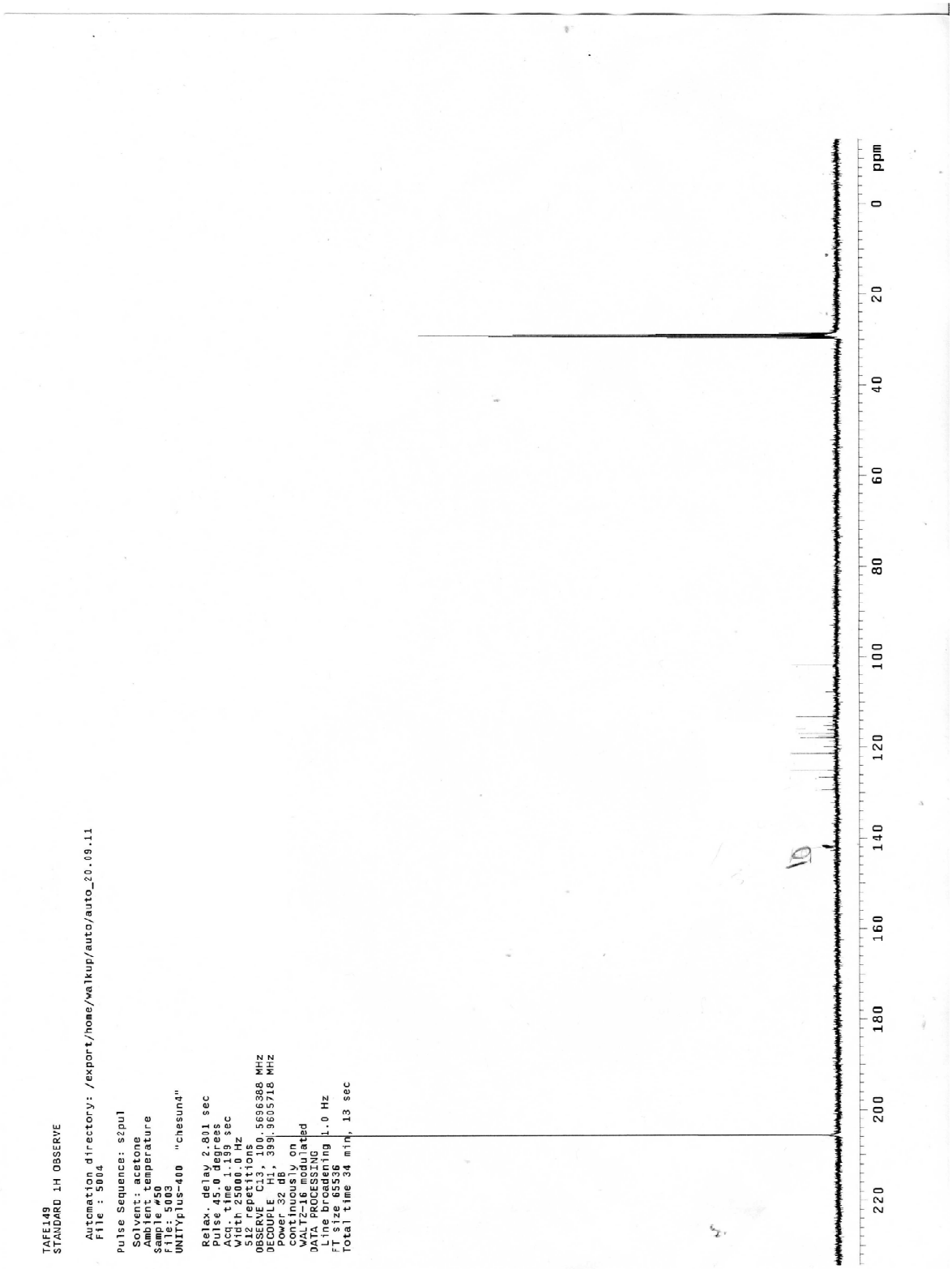
^{13}C NMR of compound 39 4-(2-nitrophenyl)-1*H*-pyrrole:



^{1}H NMR of compound 38 2-(1*H*-pyrrol-3-yl)aniline



^{13}C NMR of compound 38 2-(1*H*-pyrrol-3-yl)aniline



References

1. K. M. Koeller and C. H. Wong. (2001) Enzymes for chemical synthesis. *Nature*. **409**, 232-240.
2. A. Zaks and A. M. Klibanov. (1985) Enzyme-catalyzed processes in organic solvents. *Proc Natl Acad Sci U S A*. **82**, 3192-3196.
3. T. Hudlicky and J. W. Reed. (2009) Applications of biotransformations and biocatalysis to complexity generation in organic synthesis. *Chem Soc Rev*. **38**, 3117-3132.
4. A. Schmid, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts and B. Witholt. (2001) Industrial biocatalysis today and tomorrow. *Nature*. **409**, 258-268.
5. N. J. Turner. (2009) Directed evolution drives the next generation of biocatalysts. *Nat Chem Biol*. **5**, 567-573.
6. H. E. Schoemaker, D. Mink and M. G. Wubbolts. (2003) Dispelling the Myths--Biocatalysis in Industrial Synthesis. *Science*. **299**, 1694-1697.
7. C. Mateo, J. M. Palomo, G. Fernandez-Lorente, J. M. Guisan and R. Fernandez-Lafuente. (2007) Improvement of enzyme activity, stability and selectivity via immobilization techniques. *Enzyme Microb Tech*. **40**, 1451-1463.
8. K. Hernandez and R. Fernandez-Lafuente. (2011) Control of protein immobilization: Coupling immobilization and site-directed mutagenesis to improve biocatalyst or biosensor performance. *Enzyme Microb Tech*. **48**, 107-122.
9. G. N. Somero. (1995) Proteins and temperature. *Annu Rev Physiol*. **57**, 43-68.
10. F. H. Arnold, P. L. Wintrode, K. Miyazaki and A. Gershenson. (2001) How enzymes adapt: lessons from directed evolution. *Trends Biochem Sci*. **26**, 100-106.
11. S. S. Branda, A. Vik, L. Friedman and R. Kolter. (2005) Biofilms: the matrix revisited. *Trends Microbiol*. **13**, 20-26.
12. I. W. Sutherland. (2001) The biofilm matrix an immobilized but dynamic microbial environment. *Trends Microbiol*. **9**, 222-227.
13. C. Beloin, A. Roux and J. M. Ghigo. (2008) Escherichia coli biofilms. *Curr Top Microbiol Immunol*. **322**, 249-289.
14. T. Romeo, ed., *Bacterial Biofilms, Current Topics in Microbiology and Immunology 322*, Springer-Verlag Berlin Heidelberg, 2008.
15. P. S. Stewart and M. J. Franklin. (2008) Physiological heterogeneity in biofilms. *Nat Rev Micro*. **6**, 199-210.
16. D. Lopez, H. Vlamakis and R. Kolter. (2010) Biofilms. *Cold Spring Harb Perspect Biol*. **2**, a000398.
17. K. D. Xu, P. S. Stewart, F. Xia, C. T. Huang and G. A. McFeters. (1998) Spatial physiological heterogeneity in Pseudomonas aeruginosa biofilm is determined by oxygen availability. *Appl Environ Microbiol*. **64**, 4035-4039.
18. P. S. Stewart. (2003) Diffusion in Biofilms. *J Bacteriol*. **185**, 1485-1491.
19. S. A. Rani, B. Pitts, H. Beyenal, R. A. Veluchamy, Z. Lewandowski, W. M. Davison, K. Buckingham-Meyer and P. S. Stewart. (2007) Spatial patterns of DNA replication, protein synthesis, and oxygen concentration within bacterial biofilms reveal diverse physiological states. *J Bacteriol*. **189**, 4223-4233.

20. F. Götz. (2002) Staphylococcus and biofilms. *Mol Microbiol*. **43**, 1367-1378.

21. C. Latasa, C. Solano, J. R. Penadés and I. Lasa. (2006) Biofilm-associated proteins. *Comptes Rendus Biologies*. **329**, 849-857.

22. K. P. Lemon, A. M. Earl, H. C. Vlamakis, C. Aguilar and R. Kolter. (2008) Biofilm development with an emphasis on Bacillus subtilis. *Curr Top Microbiol Immunol*. **322**, 1-16.

23. K. P. Lemon, D. E. Higgins and R. Kolter. (2007) Flagellar motility is critical for *Listeria monocytogenes* biofilm formation. *J. Bacteriol*. **189**, 4418.

24. C. Picioreanu, J. U. Kreft, M. Klausen, J. A. Haagensen, T. Tolker-Nielsen and S. Molin. (2007) Microbial motility involvement in biofilm structure formation-a 3D modelling study. *Water Sci Technol*. **55**, 337-343.

25. J. Lee, A. Jayaraman and T. K. Wood. (2007) Indole is an inter-species biofilm signal mediated by SdiA. *BMC Microbiol*. **7**, 42.

26. J. S. Dickschat. (2008) Quorum sensing and bacterial biofilms. *Nat Prod Rep*. **27**, 343-369.

27. C. M. Waters and B. L. Bassler. (2005) Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol*. **21**, 319-346.

28. K. H. Nealson and J. W. Hastings. (1979) Bacterial bioluminescence: its control and ecological significance. *Microbiol Rev*. **43**, 496-518.

29. A. Vannini, C. Volpari, C. Gargioli, E. Muraglia, R. Cortese, R. De Francesco, P. Neddermann and S. Di Marco. (2002) The crystal structure of the quorum sensing protein TraR bound to its autoinducer and target DNA. *EMBO J*. **21**, 4393-4401.

30. W. T. Watson, T. D. Minogue, D. L. Val, S. B. von Bodman and M. E. Churchill. (2002) Structural basis and specificity of acyl-homoserine lactone signal production in bacterial quorum sensing. *Mol Cell*. **9**, 685-694.

31. M. Schuster, C. P. Lostroh, T. Ogi and E. P. Greenberg. (2003) Identification, Timing, and Signal Specificity of Pseudomonas aeruginosa Quorum-Controlled Genes: a Transcriptome Analysis. *J Bacteriol*. **185**, 2066-2079.

32. D. G. Davies, M. R. Parsek, J. P. Pearson, B. H. Iglewski, J. W. Costerton and E. P. Greenberg. (1998) The involvement of cell-to-cell signals in the development of a bacterial biofilm. *Science*. **280**, 295-298.

33. T. Bjarnsholt, P. O. Jensen, M. Burmolle, M. Hentzer, J. A. J. Haagensen, H. P. Hougen, H. Calum, K. G. Madsen, C. Moser, S. Molin, N. Hoiby and M. Givskov. (2005) Pseudomonas aeruginosa tolerance to tobramycin, hydrogen peroxide and polymorphonuclear leukocytes is quorum-sensing dependent. *Microbiology*. **151**, 373-383.

34. M. J. Lynch, S. Swift, D. F. Kirke, C. W. Keevil, C. E. R. Dodd and P. Williams. (2002) The regulation of biofilm development by quorum sensing in Aeromonas hydrophila. *Environ Microbiol*. **4**, 18-28.

35. B. Huber, K. Riedel, M. Hentzer, A. Heydorn, A. Gotschlich, M. Givskov, S. r. Molin and L. Eberl. (2001) The cep quorum-sensing system of Burkholderia cepacia H111 controls biofilm formation and swarming motility. *Microbiology*. **147**, 2517-2528.

36. K. L. Tomlin, R. J. Malott, G. Ramage, D. G. Storey, P. A. Sokol and H. Ceri. (2005) Quorum-sensing mutations affect attachment and stability of *Burkholderia cenocepacia* biofilms. *Appl Environ Microb*. **71**, 5208-5218.

37. A. Steidle, M. Allesen-Holm, K. Riedel, G. Berg, M. Givskov, S. r. Molin and L. Eberl. (2002) Identification and characterization of an N-acylhomoserine lactone-dependent quorum-sensing system in *Pseudomonas putida* strain IsoF. *Appl Environ Microb*. **68**, 6371-6382.

38. M. Labbate, S. Y. Queck, K. S. Koh, S. A. Rice, M. Givskov and S. Kjelleberg. (2004) Quorum sensing-controlled biofilm development in *Serratia liquefaciens* MG1. *J Bacteriol*. **186**, 692-698.

39. S. A. Rice, K. S. Koh, S. Y. Queck, M. Labbate, K. W. Lam and S. Kjelleberg. (2005) Biofilm formation and sloughing in *Serratia marcescens* are controlled by quorum sensing and nutrient cues. *J Bacteriol*. **187**, 3477-3485.

40. Y. Irie and M. R. Parsek. (2008) Quorum sensing and microbial biofilms. *Curr Top Microbiol Immunol*. **322**, 67-84.

41. R. P. Novick, H. F. Ross, S. J. Projan, J. Kornblum, B. Kreiswirth and S. Moghazeh. (1993) Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J*. **12**, 3967-3975.

42. O. Vidal, R. Longin, C. Prigent-Combaret, C. Dorel, M. Hooreman and P. Lejeune. (1998) Isolation of an *Escherichia coli* K-12 mutant strain able to form biofilms on inert surfaces: involvement of a new ompR allele that increases curli expression. *J Bacteriol*. **180**, 2442-2449.

43. A. Olsen, A. Jonsson and S. Normark. (1989) Fibronectin binding mediated by a novel class of surface organelles on Escherichia coll. *Nature*. **338**, 652-655.

44. C. Prigent-Combaret, E. Brombacher, O. Vidal, A. Ambert, P. Lejeune, P. Landini and C. Dorel. (2001) Complex regulatory network controls inital adhesion and biofilm formation in *Escherichia coli* via regulation of the *csgD* gene. *J Bacteriol*. **183**, 7213-7223.

45. S. Reshamwala and S. Noronha. (2011) Biofilm formation in Escherichia coli cra mutants is impaired due to down-regulation of curli biosynthesis. *Archives of Microbiology*. **193**, 711-722.

46. A. Reisner, J. A. J. Haagensen, M. A. Schembri, E. L. Zechner and S. Molin. (2003) Development and maturation of Escherichia coli K-12 biofilms. *Mol Microbiol*. **48**, 933-946.

47. K. Suzuki, X. Wang, T. Weilbacher, A.-K. Pernestig, O. Melefors, D. Georgellis, P. Babitzke and T. Romeo. (2002) Regulatory circuitry of the CsrA/CsrB and BarA/UvrY systems of *Escherichia coli*. *J Bacteriol*. **184**, 5130-5140.

48. E. Woehl and M. F. Dunn. (1999) Mechanisms of monovalent cation action in enzyme catalysis: the first stage of the tryptophan synthase beta-reaction. *Biochemistry*. **38**, 7118-7130.

49. W. A. Newton and E. E. Snell. (1964) Catalytic properties of tryptophanase, a multifunctional pridoxal phosphate enzyme. *Proc Natl Acad Sci U S A*. **51**, 382-389.

50. D. Wang, X. Ding and P. N. Rather. (2001) Indole can act as an extracellular signal in *Escherichia coli*. *J Bacteriol*. **183**, 4210-4216.

51. P. Di Martino, R. Fursy, L. Bret, B. Sundararaju and R. S. Phillips. (2003) Indole can act as an extracellular signal to regulate biofilm formation of Escherichia coli and other indole-producing bacteria. *Can J Microbiol*. **49**, 443-449.

52. M. Morikawa. (2006) Beneficial biofilm formation by industrial bacteria Bacillus subtilis and related species. *J Biosci Bioeng*. **101**, 1-8.

53. L. Hall-Stoodley and P. Stoodley. (2009) Evolving concepts in biofilm infections. *Cell Microbiol*. **11**, 1034-1043.

54. G. G. Anderson and G. A. O'Toole. (2008) Innate and induced resistance mechanisms of bacterial biofilms. *Curr Top Microbiol Immunol*. **322**, 85-105.

55. C. Chan, L. L. Burrows and C. M. Deber. (2005) Alginate as an auxiliary bacterial membrane: binding of membrane-active peptides by polysaccharides*. *J Pept Res*. **65**, 343-351.

56. M. C. Walters, F. Roe, A. Bugnicourt, M. J. Franklin and P. S. Stewart. (2003) Contributions of Antibiotic Penetration, Oxygen Limitation, and Low Metabolic Activity to Tolerance of Pseudomonas aeruginosa Biofilms to Ciprofloxacin and Tobramycin. *Antimicrob Agents Ch*. **47**, 317-323.

57. D. Kaplan, D. Christiaen and S. Arad. (1987) Chelating Properties of Extracellular Polysaccharides from Chlorella spp. *Appl Environ Microbiol*. **53**, 2953-2956.

58. N. Bagge, M. Schuster, M. Hentzer, O. Ciofu, M. Givskov, E. P. Greenberg and N. HÃ¸iby. (2004) Pseudomonas aeruginosa biofilms exposed to Imipenem exhibit changes in global gene expression and beta-lactamase and alginate production. *Antimicrob Agents Ch*. **48**, 1175-1187.

59. M. R. W. Brown, D. G. Allison and P. Gilbert. (1988) Resistance of bacterial biofilms to antibiotics a growth-rate related effect? *Journal Antimicrob Chemoth*. **22**, 777-780.

60. N. Balaban. (2011) Persistence: mechanisms for triggering and enhancing phenotypic variability. *Curr Opin Genet Dev*. **21**, 768-775.

61. T. Dorr, M. Vulic and K. Lewis. (2010) Ciprofloxacin causes persister formation by inducing the TisB toxin in *Escherichia coli*. *PLoS Biol*. **8**, e1000317.

62. M. Kvist, V. Hancock and P. Klemm. (2008) Inactivation of efflux pumps abolishes bacterial biofilm formation. *Appl Environ Microbiol*. **74**, 7376-7382.

63. K. Matsumura, S. Furukawa, H. Ogihara and Y. Morinaga. (2011) Roles of Multidrug Efflux Pumps on the Biofilm Formation of Escherichia coli K-12. *Biocontrol Science*. **16**, 69-72.

64. X. Z. Li, J. S. Webb, S. Kjelleberg and B. Rosche. (2006) Enhanced benzaldehyde tolerance in Zymomonas mobilis biofilms and the potential of biofilm applications in fine-chemical production. *Appl. Environ. Microbiol.* **72**, 1639-1644.

65. B. Halan, A. Schmid and K. Buehler. (2010) Real-time solvent tolerance analysis of *Pseudomonas sp*. Strain VLB120 catalytic biofilms. *Appl Environ Microbiol*. **77**, 1563-1571.

66. J. L. Adams and R. J. McLean. (1999) Impact of rpoS deletion on Escherichia coli biofilms. *Appl Environ Microbiol*. **65**, 4285-4287.

67. B. Rosche, X. Z. Li, B. Hauer, A. Schmid and K. Buehler. (2009) Microbial biofilms: a concept for industrial catalysis? *Trends biotechnol*. **27**, 636-643.

68. S. G. Cull, J. D. Holbrey, V. Vargas-Mora, K. R. Seddon and G. J. Lye. (2000) Room-temperature ionic liquids as replacements for organic solvents in multiphase bioprocess operations. *Biotechnol Bioeng*. **69**, 227-233.

69. H. Zhao. (2010) Methods for stabilizing and activating enzymes in ionic liquids - A review. *J Chem Technol Biot*. **85**, 891-907.

70. N. Qureshi, B. A. Annous, T. C. Ezeji, P. Karcher and I. S. Maddox. (2005) Biofilm reactors for industrial bioconversion processes: Employing potential of enhanced reaction rates. *Microb. Cell Fact.* **4**.

71. E. C. Hann, A. Eisenberg, S. K. Fager, N. E. Perkins, F. G. Gallagher, S. M. Cooper, J. E. Gavagan, B. Stieglitz, S. M. Hennessey and R. DiCosimo. (1999) 5-Cyanovaleramide production using immobilized pseudomonas chlororaphis B23. *Bioorgan Med Chem*. **7**, 2239-2245.

72. S. R. Carter and W. J. Jewell. (1993) Biotransformation of tetrachloroethylene by anaerobic attached-films at low temperatures. *Water Res.* **27**, 607-615.

73. S. I. Yang, J. R. Lawrence, G. D. W. Swerhone and I. J. Pickering. (2011) Biotransformation of selenium and arsenic in multi-species biofilm. *Environ Chem*. **8**, 543-551.

74. J.-P. Arcangeli and E. Arvin. (1995) Cometabolic transformation of o-xylene in a biofilm system under nitrate reducing conditions. *Biodegradation*. **6**, 19-27.

75. A. K. Luke and S. G. Burton. (2001) A novel application for *Neurospora crassa*: Progress from batch culture to a membrane bioreactor for the bioremediation of phenols. *Enzyme Microb Tech*. **29**, 348-356.

76. G. W. Nofsinger and R. J. Bothast. (1981) Ethanol production by Zymomonas mobilis and Saccharomyces uvarum on aflatoxin-contaminated and ammonia-detoxified corn. *Can J Microbiol*. **27**, 162-167.

77. R. R. Bland, H. C. Chen, W. J. Jewell, W. D. Bellamy and R. R. Zall. (1982) Continuous high rate production of ethanol by Zymomonas mobilis in an attached film expanded bed fermentor. *Biotechnol Lett*. **4**, 323-328.

78. J.-Y. Xin, J.-R. Cui, J.-B. Chen, S.-B. Li, C.-G. Xia and L.-M. Zhu. (2003) Continuous biocatalytic synthesis of epoxypropane using a biofilm reactor. *Process Biochem*. **38**, 1739-1746.

79. J. Swaving and J. A. M. de Bont. (1998) Microbial transformation of epoxides. *Enzyme Microb Technol*. **22**, 19-26.

80. X. Z. Li, B. Hauer and B. Rosche. (2007) Single-species microbial biofilm screening for industrial applications. *Appl. Microbiol. Biotechnol.* **76**, 1255-1262.

81. R. Gross, B. Hauer, K. Otto and A. Schmid. (2007) Microbial biofilms: new catalysts for maximizing productivity of long-term biotransformations. *Biotechnol. Bioeng.* **98**, 1123-1134.

82. B. Zhou, G. J. O. Martin and N. B. Pamment. (2008) Increased phenotypic stability and ethanol tolerance of recombinant Escherichia coli KO11 when immobilized in continuous fluidized bed culture. *Biotechnol Bioeng*. **100**, 627-633.

83. M. I. Setyawati, L.-J. Chien and C.-K. Lee. (2009) Self-immobilized recombinant Acetobacter xylinum for biotransformation. *Biochem. Eng. J.* **43**, 78-84.

84. M. S. Butler. (2005) Natural products to drugs: natural product derived compounds in clinical trials. *Nat Prod Rep*. **22**, 162-195.

85. I. Paterson and E. A. Anderson. (2005) The Renaissance of Natural Products as Drug Candidates. *Science*. **310**, 451-453.

86. M. D. Burke and S. L. Schreiber. (2004) A planning strategy for diversity-oriented synthesis. *Angew Chem Int Edit*. **43**, 46-58.

87. F. E. Koehn and G. T. Carter. (2005) The evolving role of natural products in drug discovery. *Nat Rev Drug Discov*. **4**, 206-220.

88. S. Grüschow, E. J. Rackham, B. Elkins, P. L. A. Newill, L. M. Hill and R. J. M. Goss. (2009) New pacidamycin antibiotics through precursor-directed biosynthesis. *ChemBioChem*. **10**, 355-360.

89. M. Winn, R. J. Goss, K. Kimura and T. D. Bugg. (2010) Antimicrobial nucleoside antibiotics targeting cell wall assembly: recent advances in structure-function studies and nucleoside biosynthesis. *Nat Prod Rep*. **27**, 279-304.

90. H. N. Rydon and J. C. Tweddle. (1955) Experiments on the synthesis of Bz-substituted indoles and tryptophans. Part III. The synthesis of the four Bz-chloro-indoles and -tryptophans. *J Chem Soc (Resumed)*. 3499-3503.

91. C. Ma, X. Liu, X. Li, J. Flippen-Anderson, S. Yu and J. M. Cook. (2001) Efficient asymmetric synthesis of biologically important tryptophan analogues via a palladium-mediated heteroannulation reaction. *J Org Chem*. **66**, 4525-4542.

92. C. W. Perry, A. Brossi, K. H. Deitcher, W. Tautz and S. Teitel. (1977) Preparation of R(+)- and S(-)-α-methyl-p-nitrobenzylamines and their use as resolving agents *Synthesis*. **1977**, 492-494.

93. Y. Konda-Yamada, C. Okada, K. Yoshida, Y. Umeda, S. Arima, N. Sato, T. Kai, H. Takayanagi and Y. Harigaya. (2002) Convenient synthesis of 7' and 6'-bromo--tryptophan and their derivatives by enzymatic optical resolution using -aminoacylase. *Tetrahedron*. **58**, 7851-7861.

94. G. Xie, C. Forst, C. Bonner and R. A. Jensen. (2002) Significance of two distinct types of tryptophan synthase beta chain in bacteria, archaea and higher plants. *Genome Biol*. **3**.

95. C. C. Hyde, S. A. Ahmed, E. A. Padlan, E. W. Miles and D. R. Davies. (1988) Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from Salmonella typhimurium. *J Biol Chem*. **263**, 17857-17871.

96. J. K. Rose and C. Yanofsky. (1974) Interaction of the operator of the tryptophan operon with repressor. *Proc Natl Acad Sci U S A*. **71**, 3134-3138.

97. F. Lee and C. Yanofsky. (1977) Transcription termination at the trp operon attenuators of Escherichia coli and Salmonella typhimurium: RNA secondary structure and regulation of termination. *Proc Natl Acad Sci U S A*. **74**, 4365-4369.

98. S. Nagata, C. C. Hyde and E. W. Miles. (1989) The alpha subunit of tryptophan synthase. Evidence that aspartic acid 60 is a catalytic residue and that the double alteration of residues 175 and 211 in a second-site revertant restores the proper geometry of the substrate binding site. *J Biol Chem*. **264**, 6288-6296.

99. S. Raboni, S. Bettati and A. Mozzarelli. (2009) Tryptophan synthase: a mine for enzymologists. *Cell Mol Life Sci*. **66**, 2391-2403.

100. A. Peracchi, S. Bettati, A. Mozzarelli, G. L. Rossi, E. W. Miles and M. F. Dunn. (1996) Allosteric regulation of tryptophan synthase: effects of pH, temperature, and subunit ligands on the equilibrium distribution of pyridoxal 5-phosphate l-serine intermediates *Biochemistry*. **35**, 1872-1880.

101. R. S. Phillips, E. W. Miles, G. Holtermann and R. S. Goody. (2005) Hydrostatic pressure affects the conformational equilibrium of salmonella typhimurium tryptophan synthase. *Biochemistry*. **44**, 7921-7928.

102. A. Marabotti, P. Cozzini and A. Mozzarelli. (2000) Novel allosteric effectors of the tryptophan synthase alpha(2)beta(2) complex identified by computer-assisted molecular modeling. *BBA-Protein Struct M*. **1476**, 287-299.

103. R. S. Phillips. (2004) Synthetic applications of tryptophan synthase. *Tetrahedron: Asymmetr*. **15**, 2787-2792.

104. T. R. M. Barends, T. Domratcheva, V. Kulik, L. Blumenstein, D. Niks, M. F. Dunn and I. Schlichting. (2008) Structure and mechanistic implications of a tryptophan synthase quinonoid intermediate. *ChemBioChem*. **9**, 1024-1028.

105. E. Schleicher, K. Mascaro, R. Potts, D. R. Mann and H. G. Floss. (1976) Stereochemistry and mechanism of reactions catalyzed by tryptophanase and tryptophan synthetase. *J Am Chem Soc*. **98**, 1043-1044.

106. R. S. Phillips, E. W. Miles and L. A. Cohen. (1985) Differential inhibition of tryptophan synthase and of tryptophanase by the two diastereoisomers of 2,3-dihydro-L-tryptophan. Implications for the stereochemistry of the reaction intermediates. *J Biol Chem*. **260**, 14665-14670.

107. E. W. Miles. (1979) Tryptophan synthase: structure, function, and subunit interaction. *Adv Enzymol Ramb*. **49**, 127-186.

108. K. Nishio, K. Ogasahara, Y. Morimoto, T. Tsukihara, S. J. Lee and K. Yutani. (2010) Large conformational changes in the *Escherichia coli* tryptophan synthase β2 subunit upon pyridoxal 5′-phosphate binding. *FEBS Journal*. **277**, 2157-2170.

109. M. Q. Fatmi and C.-e. A. Chang. (2010) The role of oligomerization and cooperative regulation in protein function: the case of tryptophan synthase. *PLoS Comput Biol*. **6**.

110. M. F. Dunn, V. Aguilar, P. Brzovic, W. F. Drewe, K. F. Houben, C. A. Leja and M. Roy. (1990) The tryptophan synthase bienzyme complex transfers indole between the alpha and beta sites via a 25-30A long tunnel. *Biochemistry*. **29**, 8598-8607.

111. P. S. Brzovic, K. Ngo and M. F. Dunn. (1992) Allosteric interactions coordinate catalytic activity between successive metabolic enzymes in the tryptophan synthase bienzyme complex. *Biochemistry*. **31**, 3831-3839.

112. T. R. Schneider, E. Gerhardt, M. Lee, P.-H. Liang, K. S. Anderson and I. Schlichting. (1998) Loop closure and intersubunit communication in tryptophan synthase. *Biochemistry*. **37**, 5394-5406.

113. H. Kawasaki, R. Bauerle, G. Zon, S. A. Ahmed and E. W. Miles. (1987) Site-specific mutagenesis of the alpha subunit of tryptophan synthase from Salmonella typhimurium. Changing arginine 179 to leucine alters the reciprocal transmission of substrate-induced conformational changes between the alpha and beta 2 subunits. *J Biol Chem*. **262**, 10678-10683.

114. K. S. Anderson, E. W. Miles and K. A. Johnson. (1991) Serine modulates substrate channeling in tryptophan synthase. A novel intersubunit triggering mechanism. *J Biol Chem*. **266**, 8020-8033.

115. M. Lee and R. S. Phillips. (1992) Enzymatic synthesis of chloro-L-tryptophans. *Bioorg Med Chem Lett*. **2**, 1563-1564.

116. R. J. M. Goss and P. L. A. Newill. (2006) A convenient enzymatic synthesis of l-halotryptophans. *Chem Commun*. 4924-4925.

117. M. Winn, A. D. Roy, S. Grüschow, R. S. Parameswaran and R. J. M. Goss. (2008) A convenient one-step synthesis of l-aminotryptophans and improved synthesis of 5-fluorotryptophan. *Bioorg Med Chem Lett*. **18**, 4508-4510.

118. A. J. Oelke, D. J. France, T. Hofmann, G. Wuitschik and S. V. Ley. (2010) Total synthesis of chloptosin. *Angew Chem Int Edit*. **49**, 6139-6142.

119. R. M. van Hardeveld, P. L. J. Gunter, L. J. van Ijzendoorn, W. Wieldraaijer, E. W. Kuipers and J. W. Niemantsverdriet. (1995) Deposition of inorganic salts from solution on flat substrates by

spin-coating: theory, quantification and application to model catalysts. *Appl Sur Sci*. **84**, 339-346.

120. S. E. Cowan, D. Liepmann and J. D. Keasling. (2001) Development of engineered biofilms on poly- L-lysine patterned surfaces. *Biotechnol Lett*. **23**, 1235-1241.

121. R. W. Robinson, D. E. Akin, R. A. Nordstedt, M. V. Thomas and H. C. Aldrich. (1984) Light and electron microscopic examinations of methane-producing biofilms from anaerobic fixed-bed reactors. *Appl Environ Microbiol*. **48**, 127-136.

122. R. Lal and S. A. John. (1994) Biological applications of atomic force microscopy. *Am J Physiol*. **266**, C1-21.

123. A. V. Bolshakova, O. I. Kiselyova, A. S. Filonov, O. Y. Frolova, Y. L. Lyubchenko and I. V. Yaminsky. (2001) Comparative studies of bacteria with an atomic force microscopy operating in different modes. *Ultramicroscopy*. **86**, 121-128.

124. F. o. Ahimou, M. Paquot, P. Jacques, P. Thonart and P. G. Rouxhet. (2001) Influence of electrical properties on the evaluation of the surface hydrophobicity of *Bacillus subtilis*. *J Microbiol Meth*. **45**, 119-126.

125. M. Benoit, D. Gabriel, G. Gerisch and H. E. Gaub. (2000) Discrete interactions in cell adhesion measured by single-molecule force spectroscopy. *Nat Cell Biol*. **2**, 313-317.

126. M. Gad, A. Itoh and A. Ikai. (1997) Mapping cell wall polysaccharides of living microbial cells using atomic force microscopy. *Cell Biol Int*. **21**, 697-706.

127. C. B. Volle, M. A. Ferguson, K. E. Aidala, E. M. Spain and M. E. Nunez. (2008) Spring constants and adhesive properties of native bacterial biofilm cells measured by atomic force microscopy. *Colloids Surf B: Biointerfaces*. **67**, 32-40.

128. J. W. Costerton, P. S. Stewart and E. P. Greenberg. (1999) Bacterial biofilms: a common cause of persistent infections. *Science*. **284**, 1318-1322.

129. T. Iwamoto and M. Nasu. (2001) Current bioremediation practice and perspective. *J. Biosci. Bioeng.* **92**, 1-8.

130. R. A. Sheldon. (2007) Enzyme immobilization: the quest for optimum performance. *ChemInform*. **38**, no-no.

131. F. A. Quiocho and F. M. Richards. (1966) The enzymic behavior of carboxypeptidase-A in the solid state. *Biochemistry*. **5**, 4062-4076.

132. P. López-Serrano, L. Cao, F. van Rantwijk and R. A. Sheldon. (2002) Cross-linked enzyme aggregates with enhanced activity: application to lipases. *Biotechnol Lett*. **24**, 1379-1383.

133. B. Ozbek and K. O. Ulgen. (2000) The stability of enzymes after sonication. *Process Biochemistry*. **35**, 1037-1043.

134. V. Lazarova and J. Manem. (1995) Biofilm characterization and activity analysis in water and wastewater treatment. *Water Res*. **29**, 2227-2245.

135. A. Di Lorenzo, M. Varcamonti, P. Parascandola, R. Vignola, A. Bernardi, P. Sacceddu, R. Sisto and E. de Alteriis. (2005) Characterization and performance of a toluene-degrading biofilm developed on pumice stones. *Microb Cell Fact*. **4**, 4.

136. *Ni-NTA purification system user manual*, Invitrogen.

137. B. Ku, J. Cha, A. Srinivasan, S. J. Kwon, J. C. Jeong, D. H. Sherman and J. S. Dordick. (2006) Chip-based polyketide biosynthesis and functionalization. *Biotechnol Prog*. **22**, 1102-1107.

138. S. Matosevic, G. J. Lye and F. Baganz. (2011) Immobilised enzyme microreactor for screening of multi-step bioconversions: Characterisation of a de novo transketolase-w-transaminase pathway to synthesise chiral amino alcohols. *J Biotechnol*. **155**, 320-329.

139. Y. Ni and R. R. Chen. (2004) Accelerating whole-cell biocatalysis by reducing outer membrane permeability barrier. *Biotechnol Bioeng*. **87**, 804-811.

140. Y. Ni, Z. Mao and R. R. Chen. (2006) Outer membrane mutation effects on UDP-glucose permeability and whole-cell catalysis rate. *Appl. Microbiol. Biotechnol.* **73**, 384-393.

141. A. Brooun, S. Liu and K. Lewis. (2000) A dose-response study of antibiotic resistance in *Pseudomonas aeruginosa* biofilms. *Antimicrob Agents Chemother*. **44**, 640-646.

142. H. Hirakawa, T. Kodama, A. Takumi-Kobayashi, T. Honda and A. Yamaguchi. (2009) Secreted indole serves as a signal for expression of type III secretion system translocators in enterohaemorrhagic *Escherichia coli* O157:H7. *Microbiology*. **155**, 541-550.

143. S. Pinero-Fernandez, C. Chimerel, U. F. Keyser and D. K. Summers. (2011) Indole transport across *Escherichia coli* membranes. *J Bacteriol*. **193**, 1793-1798.

144. J. R. Piperno and D. L. Oxender. (1968) Amino acid transport systems in Escherichia coli K-12. *J Biol Chem*. **243**, 5914-5920.

145. K. D. Brown. (1970) Formation of aromatic amino acid pools in Escherichia coli K-12. *J Bacteriol*. **104**, 177-188.

146. C. A. Santiviago, J. A. Fuentes, S. M. Bueno, A. N. Trombert, A. A. Hildago, L. T. Socias, P. Youderian and G. C. Mora. (2002) The Salmonella enterica sv. Typhimurium smvA, yddG and ompD (porin) genes are required for the efficient efflux of methyl viologen. *Mol Microbiol*. **46**, 687-698.

147. V. Doroshenko, L. Airich, M. Vitushkina, A. Kolokolova, V. Livshits and S. Mashko. (2007) YddG from Escherichia coli promotes export of aromatic amino acids. *FEMS Microbiol Lett*. **275**, 312-318.

148. G. Gribble, in *Natural production of organohalogen compounds*, Springer Berlin / Heidelberg, Editon edn., 2003, vol. 3P, pp. 56-56.

149. C. Wagner, M. El Omari and G. M. Konig. (2009) Biohalogenation: natures way to synthesize halogenated metabolites *J Nat Prod*. **72**, 540-553.

150. C. S. Neumann, D. G. Fujimori and C. T. Walsh. (2008) Halogenation strategies in natural product biosynthesis. *Chemistry&Biology*. **15**, 99-109.

151. F. H. Vaillancourt, E. Yeh, D. A. Vosburg, S. Garneau-Tsodikova and C. T. Walsh. (2006) Nature's inventory of halogenation catalysts: oxidative strategies predominate. *Chem Rev*. **106**, 3364-3378.

152. J. A. Bush, B. H. Long, J. J. Catino, W. T. Bradner and K. Tomita. (1987) Production and biological activity of rebeccamycin, a novel antitumor agent. *J Antibiot (Tokyo)*. **40**, 668-678.

153. Z. Hui-ping, K. Yoshiaki, K. Haruhisa, I. Hideji, P. G. R and H. C. L. (1994) Convolutamines A-E, novel phenylethylamine alkaloids from marine *Bryozoan Amathia convoluta*. *Chemistry letters*. **1994**, 2271-2274.

154. M. D. Lee, J. K. Manning, D. R. Williams, N. A. Kuck, R. T. Testa and D. B. Borders. (1989) Calicheamicins, a novel family of antitumor antibiotics. 3. Isolation, purification and characterization of calicheamicins beta 1Br, gamma 1Br, alpha 2I, alpha 3I, beta 1I, gamma 1I and delta 1I. *J Antibiot (Tokyo)*. **42**, 1070-1087.

155. M. Morrison and G. R. Schonbaum. (1976) Peroxidase-catalyzed halogenation. *Annu. Rev. Biochem.* **45**, 861-888.

156. D. O'Hagan and B. Harper. (1999) Fluorine-containing natural products. *J Fluorine Chem*. **100**, 127-133.

157. A. S. Eustaquio, B. Gust, T. Luft, S.-M. Li, K. F. Chater and L. Heide. (2003) Clorobiocin biosynthesis in Streptomyces: identification of the halogenase and generation of structural analogs. *Chemistry & Biology*. **10**, 279-288.

158. M. Groll, R. Huber and B. C. M. Potts. (2006) Crystal structures of salinosporamide A (NPI-0052) and B (NPI-0047) in complex with the 20S proteasome reveal important consequences of lactone ring opening and a mechanism for irreversible binding. *J Am Chem Soc*. **128**, 5136-5141.

159. A. S. Eustáquio and B. S. Moore. (2008) Mutasynthesis of fluorosalinosporamide, a potent and reversible inhibitor of the proteasome. *Angew Chem Int Edit*. **47**, 3936-3938.

160. E. R. Pereira, L. Belin, M. Sancelme, M. Prudhomme, M. Ollier, M. Rapp, D. Severe, J. F. Riou, D. Fabbro and T. Meyer. (1996) Structure-activity relationships in a series of substituted indolocarbazoles: topoisomerase I and protein kinase C inhibition and antitumoral and antimicrobial properties. *J Med Chem*. **39**, 4471-4477.

161. J. M. Winter, M. C. Moffitt, E. Zazopoulos, J. B. McAlpine, P. C. Dorrestein and B. S. Moore. (2007) Molecular basis for chloronium-mediated meroterpene cyclization. *J. Biol Chem*. **282**, 16362-16368.

162. A. Butler and M. Sandy. (2009) Mechanistic considerations of halogenating enzymes. *Nature*. **460**, 848-854.

163. T. Dairi, T. Nakano, T. Mizukami, K. Aisaka, M. Hasegawa and R. Katsumata. (1995) Conserved organization of genes for biosynthesis of chlortetracycline in Streptomyces strains. *Biosci Biotechnol Biochem*. **59**, 1360-1361.

164. J. Hartung. (1999) The biosynthesis of barbamide—a radical pathway for biohalogenation? *Angew Chem Int Edit*. **38**, 1209-1211.

165. F. H. Vaillancourt, J. Yin and C. T. Walsh. (2005) SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O2-dependent halogenase. *Proc Natl Acad Sci U S A*. **102**, 10111-10116.

166. F. H. Vaillancourt, D. A. Vosburg and C. T. Walsh. (2006) Dichlorination and bromination of a threonyl-S-carrier protein by the non-heme Fe(II) halogenase SyrB2. *Chembiochem*. **7**, 748-752.

167. L. C. Blasiak, F. d. r. H. Vaillancourt, C. T. Walsh and C. L. Drennan. (2006) Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature*. **440**, 368-371.

168. Z. Chang, P. Flatt, W. H. Gerwick, V.-A. Nguyen, C. L. Willis and D. H. Sherman. (2002) The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene*. **296**, 235-247.

169. D. J. Edwards, B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts and W. H. Gerwick. (2004) Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. *Chem Biol*. **11**, 817-833.

170. P. E. Hammer, D. S. Hill, S. T. Lam, K. H. Van Pee and J. M. Ligon. (1997) Four genes from Pseudomonas fluorescens that encode the biosynthesis of pyrrolnitrin. *Appl Environ Microbiol*. **63**, 2147-2154.

171. S. Kirner, P. E. Hammer, D. S. Hill, A. Altmann, I. Fischer, L. J. Weislo, M. Lanahan, K. H. van Pee and J. M. Ligon. (1998) Functions encoded by pyrrolnitrin biosynthetic genes from Pseudomonas fluorescens. *J Bacteriol*. **180**, 1939-1943.

172. S. Keller, T. Wage, K. Hohaus, M. Holzer, E. Eichhorn and K. H. van Pee. (2000) Purification and partial characterization of tryptophan 7-halogenase (PrnA) from *Pseudomonas fluorescens*. *Angew Chem Int Ed Engl*. **39**, 2300-2302.

173. S. Zehner, A. Kotzsch, B. Bister, R. D. Sussmuth, C. Mendez, J. A. Salas and K. H. van Pee. (2005) A regioselective tryptophan 5-halogenase is involved in pyrroindomycin biosynthesis in Streptomyces rugosporus LL-42D005. *Chem Biol*. **12**, 445-452.

174. C. Seibold, H. Schnerr, J. Rumpf, A. Kunzendorf, C. Hatscher, T. Wage, A. J. Ernyei, C. Dong, J. H. Naismith and K. H. Van Pee. (2006) A flavin-dependent tryptophan 6-halogenase and its use in modification of pyrrolnitrin biosynthesis. *Biocatal Biotransfor*. **24**, 401.

175. J. Zeng and J. Zhan. (2011) Characterization of a tryptophan 6-halogenase from Streptomyces toxytricini. *Biotechnol Lett*. **33**, 1607-1613.

176. S. Unversucht, F. Hollmann, A. Schmid and K.-H. van Pée. (2005) FADH2-Dependence of Tryptophan 7-Halogenase. *Adv Synth Catal*. **347**, 1163-1167.

177. J. K. Lee and H. Zhao. (2007) Identification and characterization of the flavin:NADH reductase (PrnF) involved in a novel two-component arylamine oxygenase. *J Bacteriol*. **189**, 8556-8563.

178. C. Dong, S. Flecks, S. Unversucht, C. Haupt, K. H. van Pee and J. H. Naismith. (2005) Tryptophan 7-halogenase (PrnA) structure suggests a mechanism for regioselective chlorination. *Science*. **309**, 2216-2219.

179. S. Flecks, E. P. Patallo, X. Zhu, A. J. Ernyei, G. Seifert, A. Schneider, C. Dong, J. H. Naismith and K. H. van Pee. (2008) New insights into the mechanism of enzymatic chlorination of tryptophan. *Angew Chem Int Ed Engl*. **47**, 9533-9536.

180. E. Yeh, L. J. Cole, E. W. Barr, J. M. Bollinger, Jr., D. P. Ballou and C. T. Walsh. (2006) Flavin redox chemistry precedes substrate chlorination during the reaction of the flavin-dependent halogenase RebH. *Biochemistry*. **45**, 7904-7912.

181. E. Yeh, S. Garneau and C. T. Walsh. (2005) Robust in vitro activity of RebF and RebH, a two-component reductase/halogenase, generating 7-chlorotryptophan during rebeccamycin biosynthesis. *Proc Natl Acad Sci U S A*. **102**, 3960-3965.

182. E. Yeh, L. C. Blasiak, A. Koglin, C. L. Drennan and C. T. Walsh. (2007) Chlorination by a long-lived intermediate in the mechanism of flavin-dependent halogenases. *Biochemistry*. **46**, 1284-1292.

183. P. C. Dorrestein, E. Yeh, S. Garneau-Tsodikova, N. L. Kelleher and C. T. Walsh. (2005) Dichlorination of a pyrrolyl-S-carrier protein by FADH2-dependent halogenase PltA during pyoluteorin biosynthesis. *Proc Natl Acad Sci U S A*. **102**, 13843-13848.

184. C. S. Neumann, C. T. Walsh and R. R. Kay. (2010) A flavin-dependent halogenase catalyzes the chlorination step in the biosynthesis of Dictyostelium differentiation-inducing factor 1. *Proc Natl Acad Sci U S A*.

185. C. Dong, F. Huang, H. Deng, C. Schaffrath, J. B. Spencer, D. O'Hagan and J. H. Naismith. (2004) Crystal structure and mechanism of a bacterial fluorinating enzyme. *Nature*. **427**, 561-565.

186. M. A. Vincent and I. H. Hillier. (2005) The solvated fluoride anion can be a good nucleophile. *Chem Comm*. 5902-5903.

187. C. D. Cadicamo, J. Courtieu, H. Deng, A. Meddour and D. O'Hagan. (2004) Enzymatic fluorination in *Streptomyces cattleya* takes place with an inversion of configuration consistent with an SN2 reaction mechanism. *ChemBioChem*. **5**, 685-690.

188. H. M. Senn, D. O'Hagan and W. Thiel. (2005) Insight into enzymatic C-F bond formation from QM and QM/MM calculations. *J Am Chem Soc*. **127**, 13643-13655.

189. D. O'Hagan, C. Schaffrath, S. L. Cobb, J. T. G. Hamilton and C. D. Murphy. (2002) Biochemistry: Biosynthesis of an organofluorine molecule. *Nature*. **416**, 279-279.

190. M. Onega, M. Winkler and D. O'Hagan. (2009) Fluorinase: a tool for the synthesis of 18F-labeled sugars and nucleosides for PET. *Future Med Chem*. **1**, 865-873.

191. H. Deng, S. L. Cobb, A. D. Gee, A. Lockhart, L. Martarello, R. P. McGlinchey, D. O'Hagan and M. Onega. (2006) Fluorinase mediated C-(18)F bond formation, an enzymatic tool for PET labelling. *Chem Commun (Camb)*. 652-654.

192. K. Arima, H. Imanaka, M. Kousaka, A. Fukuta and G. Tamura. (1964) Pyrrolnitrin, a new antibiotic substance produced by pseudomonas. *Agric Biol Chem*. **28**, 575-576.

193. K.-H. van Pee and J. M. Ligon. (2000) Biosynthesis of pyrrolnitrin and other phenylpyrrole derivatives by bacteria. *Nat Prod Rep*. **17**, 157-164.

194. X. Zhu, K. H. van Pee and J. H. Naismith. (2010) The ternary complex of PrnB (the second enzyme in the pyrrolnitrin biosynthesis pathway), tryptophan, and cyanide yields new mechanistic insights into the indolamine dioxygenase superfamily. *J Biol Chem*. **285**, 21126-21133.

195. J. Lee, M. Simurdiak and H. Zhao. (2005) Reconstitution and characterization of aminopyrrolnitrin oxygenase, a Rieske N-oxygenase that catalyzes unusual arylamine oxidation. *J Biol Chem*. **280**, 36719-36727.

196. I. Wynands and K. H. van Pee. (2004) A novel halogenase gene from the pentachloropseudilin producer Actinoplanes sp. ATCC 33002 and detection of in vitro halogenase activity. *FEMS Microbiol Lett*. **237**, 363-367.

197. L. Heide, L. Westrich, C. Anderle, B. Gust, B. Kammerer and J. Piel. (2008) Use of a halogenase of hormaomycin biosynthesis for formation of new clorobiocin analogues with 5-chloropyrrole moieties. *Chembiochem*. **9**, 1992-1999.

198. X. Zhang and R. J. Parry. (2007) Cloning and Characterization of the Pyrrolomycin Biosynthetic Gene Clusters from Actinosporangium vitaminophilum ATCC 31673 and Streptomyces sp. Strain UC 11065. *Antimicrob Agents Ch*. **51**, 946-957.

199. W. Runguphan, X. Qu and S. E. O'Connor. (2010) Integrating carbon-halogen bond formation into medicinal plant metabolism. *Nature*. **468**, 461-464.

200. A. Deb Roy, S. Gruschow, N. Cairns and R. J. Goss. (2010) Gene expression enabling synthetic diversification of natural products: chemogenetic generation of pacidamycin analogs. *J Am Chem Soc*. **132**, 12243-12245.

201. A. D. Roy, R. J. Goss, G. K. Wagner and M. Winn. (2008) Development of fluorescent aryltryptophans by Pd mediated cross-coupling of unprotected halotryptophans in water. *Chem Commun (Camb)*. 4831-4833.

202. M. Hölzer, W. Burd, H. U. Reißig and K. H. Van Pée. (2001) Substrate specificity and regioselectivity of tryptophan 7-halogenase from *Pseudomonas fluorescens* BL915. *Adv Synth Catal*. **343**, 591-595.

203. X. Zhu, W. De Laurentis, K. Leang, J. Herrmann, K. Ihlefeld, K.-H. van Pée and J. H. Naismith. (2009) Structural insights into regioselectivity in the enzymatic chlorination of tryptophan. *J Mol Biol*. **391**, 74-85.

204. A. Lang, S. Polnick, T. Nicke, P. William, E. P. Patallo, J. H. Naismith and K. H. van Pee. (2011) Changing the regioselectivity of the tryptophan 7-halogenase PrnA by site-directed mutagenesis. *Angew Chem Int Ed Engl*. **50**, 2951-2953.

205. J. Zeng and J. Zhan. (2010) A novel fungal flavin-dependent halogenase for natural product biosynthesis. *ChemBioChem*. **11**, 2119-2123.

206. A. L. Fink. (1999) Chaperone-mediated protein folding. *Physiol Rev*. **79**, 425-449.

207. F. W. Studier and B. A. Moffatt. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*. **189**, 113-130.

208. Novagen, *pET system manual*, 11th edn., EMD Biosciences, (2005).

209. F. William Studier, A. H. Rosenberg, J. J. Dunn, J. W. Dubendorff and V. G. David, in *Methods in Enzymology*, Academic Press, Editon edn., 1990, vol. Volume 185, pp. 60-89.

210. R. Schleif. (2000) Regulation of the l-arabinose operon of Escherichia coli. *Trends Genet*. **16**, 559-565.

211. X. Chen and K. H. Van Pée. (2008) Catalytic mechanisms, basic roles, and biotechnological and environmental significance of halogenating enzymes. *Acta Bioch Bioph Sin*. **40**, 183-193.

212. D. W. Buchan, S. M. Ward, A. E. Lobley, T. C. Nugent, K. Bryson and D. T. Jones. (2010) Protein annotation and modelling servers at University College London. *Nucl. Acids Res.* **38 Suppl**, W563-W568.

213. D. T. Jones. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195-202.

214. A. N. Fedorov and T. O. Baldwin. (1997) Cotranslational protein folding. *J Biol Chem*. **272**, 32715-32718.

215. K. Varadaraj and D. M. Skinner. (1994) Denaturants or cosolvents improve the specificity of PCR amplification of a G + C-rich DNA using genetically engineered DNA polymerases. *Gene*. **140**, 1-5.

216. G. Sarkar, S. Kapelner and S. S. Sommer. (1990) Formamide can dramatically improve the specificity of PCR. *Nucleic Acids Res*. **18**, 7465.

217. R. Chakrabarti and C. E. Schutt. (2001) The enhancement of PCR amplification by low molecular-weight sulfones. *Gene*. **274**, 293-298.

218. *pBAD Manual*, Invitrogen life technologies, (2001).

219. *QuickChange site-directed mutagenesis kit: Instruction Manual*, Stratagene, (2006).

220. www.takara-bio.com.

221. K. Nishihara, M. Kanemori, M. Kitagawa, H. Yanagi and T. Yura. (1998) Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli. *Appl Environ Microbiol*. **64**, 1694-1699.

222. K. Nishihara, M. Kanemori, H. Yanagi and T. Yura. (2000) Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli. *Appl Environ Microbiol*. **66**, 884-889.

223. C. Aslanidis and P. J. de Jong. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res*. **18**, 6069-6074.

224. P. Bernard and M. Couturier. (1992) Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol*. **226**, 735-745.

225. M. D. Morrison, J. J. Hanthorn and D. A. Pratt. (2009) Synthesis of pyrrolnitrin and related halogenated phenylpyrroles. *Org. Lett.* **11**, 1051-1054.

226. J. Sambrook and D. Russell, *Molecular cloning: a laboratory manual*, 3rd edn., Cold Spring Harbour Laboratory Press: New York, (2001).

227. J. Bozzola, J and L. D. Russell, *Electron Microscopy*, 2nd edn., Jones & Bartlett Learning, (1992).

List of Publications Obtained During PhD

1. M. Winn , J. M. Foulkes, S. Peroni, M. H. Simmons, T. W. Overton, R. J. M. Goss. (2012) Biofilms and their engineered counterparts: A new generation of immobilised biocatalysts. *Catalysis Science and Techology*.
2. A. N. Tsoligkas, J. Bowen, M. Winn, R. J. M. Goss, T. W. Overton and M. J. H. Simmons. (2012) Characterisation of spin coated engineered Escherichia coli biofilms using atomic force microscopy. *Colloids and Surfaces B: Biointerfaces*. **89**, 152-160.
3. A. N. Tsoligkas, M. Winn, J. Bowen, T. W. Overton, M. J. H. Simmons and R. J. M. Goss. (2011) Engineering Biofilms for Biocatalysis. *ChemBioChem*. **12**, 1391-1395.
4. M. Winn, R. J. Goss, K. Kimura and T. D. Bugg. (2010) Antimicrobial nucleoside antibiotics targeting cell wall assembly: recent advances in structure-function studies and nucleoside biosynthesis. *Nat Prod Rep*. **27**, 279-304.