Heterologous production of recombinant peptidylglycine α-amidating monooxygenase for the production of biosimilar α-amidated peptides.

By

David Graham Morrison

MRRDAV014

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN

Thesis presented in fulfilment of the requirements for the degree of

Doctor of Philosophy

in Medical Biochemistry,

Department of Integrative Biomedical Sciences,

Faculty of Health Sciences,

UNIVERSITY OF CAPE TOWN

8th September 2021

Supervisors: Prof. E.D. Sturrock; Dr. L. Steenkamp

DECLARATION

I, David Graham Morrison, hereby declare that the work on which this thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university. I authorise the University to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: ……

Date: 08/09/2021

Acknowledgements

I thank the almighty God for his blessing and help throughout this project. Thanks to my parents, Graham and Judy Morrison, for their constant belief and support throughout this time and education. It is on their shoulders I have been able to ascend this mountain. Thank you to my supervisors Dr Lucia Steenkamp, and Prof. Edward Sturrock, for their support, knowledge, patience, and guidance. My gratitude to Prof. Paul Steenkamp for his essential work with UPLC-MS, that proved critical to proving enzyme activity and troubleshooting enzyme assays. Thank you to Dr. Chris Van der Westhuyzen for his vital insight and work to detail the AGG chromophore pathway and securing the plethora of chemicals required over many years. Thank you to Dr. Tsekoa, Dr. Roth, Dr. Van Zyl for their support, advice and instructions. Thanks to my brother for his advice and support through this difficult process and for listening to my complaints. I give thanks to all my friends and colleague for their help and advice through the many years. This work was made possible with financial grants and assistance from the National Research Foundation, the Council of Scientific and Industrial Research, and the Morrison family.

"You cannot be innovative and not have failures. Being paralysed at the prospect of a potential failure is the greatest failure of all."

- Wayne Hale

"At that time Jesus said, "I praise you, Father, Lord of heaven and earth, because you have hidden these things from the wise and learned, and revealed them to little children."

- Matthew 11:25

Table of Contents

Chapter 4: Protein expression, purification, and identification

Appendix II: Primer sequences, properties, and PCR data.

Appendix III: Record of all plasmid DNA sequences and key features.

Appendix IV: Record of all reagents and consumables

Appendix V: Protein concentration standard curve

List of Figures

List of Tables

Abbreviations

2AB	2-Aminobenzaldehyde
αC	Alpha-carbon
αHA	Alpha-hydroxyhippuric acid
AA	Amino acid
acYVG	Acetyl-L-tyrosyl-L-valyl-glycine
AGE	Agarose gel electrophoresis
AGG	2-Aminobenzaldehyde-glycine-glyoxylate
AMP	Anti-microbial peptides
API	Active pharmaceutical ingredient
APS	Ammonium persulphate
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
C-terminal	Carboxy-terminal
CASP	Critical assessment of protein structure prediction
cc	Catalytic core
CHO	Chinese hamster ovary
COACH	Consensus approach modelling
CPS	Chemical peptide synthesis
DBM	Dopamine beta-monooxygenase
DMPD	N,N-Dimethyl-1,4-phenylenediamine
DNA	Deoxyribonucleic acid
DOMON	Dopamine-monooxygenase N-terminal
E-factor	Environmental-factor
EB	Elution buffer
EBI	European Bioinformatics Institute
EC	Enzyme classification number
ER	Endoplasmic reticulum
ESI	Electro-spray ionisation
FDA	Food and drug administration
GC-content	Guanine and cytosine content
HA	Hippuric acid
HIV	Human immunodeficiency virus

hp4d	Hybrid promoter four direct
HPLC	High-performance liquid chromatography
IDR	Innate defence regulator
IMAC	Immobilised metal affinity chromatography
IV	Intravenous
K_{CAT}	Catalytic constant
K_M	Michaelis–Menten constant
LB	Luria broth
MMS	Magnetic microspheres
MS	Mass spectrometer
MSA	Multiple sequence alignments
MWCO	Molecular weight cut-off
NHL	Protein structure found in NCL-1, HT2A, and LIN-41 proteins
NMR	Nuclear magnetic resonance
NTA	Nitrilotriacetic acid
PAL	Peptidyl-α-hydroxyglycine α-amidating lyase protein
PAL	Peptidyl-α-hydroxyglycine α-amidating lyase DNA sequence
PAM	Peptidylglycine α-amidating monooxygenase protein
PAM	Peptidylglycine α-amidating monooxygenase DNA sequence
PCR	Polymerase chain reaction
PDB	Protein data bank
PHAB	Pyridine, hippuric acid, benzenesulphonyl chloride
PHM	Peptidylglycine α -hydroxylating monooxygenase protein
PHM	Peptidylglycine α -hydroxylating monooxygenase DNA sequence
PNGase	Peptide-N-glycosidase F
PPB	Protein production broth
PSS	Protein storage solution
PTM	Post-translational modification
R&D	Research and development
RASSF9	Ras association domain family member 9
RMSD	Root mean square deviation
RT-PCR	Reverse-transcriptase polymerase chain reaction
SD	Standard deviation
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis

SPS	Solution peptide synthesis
SPPS	Solid-phase peptide synthesis
TAE	Tris, acetic acid, EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
TFF	Tangential flow filtration
TGN	Trans-Golgi network
UniProt	Universal protein resource
USA	United States of America
USD$	United States of America dollar
UV	Ultraviolet
UPLC	Ultra performance liquid chromatography
WEB	Wash, equilibration, binding buffer
WT	Wild type
V_{MAX}	Maximum reaction rate
YPD	Yeast peptone dextrose

Abstract

A biological method for peptide synthesis provides increased production capacity of inexpensive peptide pharmaceuticals with environmentally safe procedures relative to current chemical peptide synthesis. Most precursor peptides are readily produced from yeast and bacterial systems using recombinant DNA technologies but require C-terminal amidation for maximum biological activity. Peptidylglycine α-amidating monooxygenase (PAM) is the only enzyme that catalyses the C-terminal amidation of peptides *in vivo* through its two catalytic cores, peptidylglycine α-hydroxylating monooxygenase (PHM) and peptidylglycine α-amidating lyase (PAL). The cost and limited quantities of the commercial PAM variants have necessitated research into low cost, scalable quantities of PAM and peptide amidation to enable inexpensive biological peptide production.

In the present study, an assay for measuring the product of PAM activity, glyoxylate, was developed based on a 2-aminobenzaldehyde-glycine-glyoxylate (AGG) absorbance assay. The AGG chromophore synthesised was identified with ultra-performance liquid chromatography mass-spectroscopy (UPLC-MS). PAM activity was measurable with glyoxylate between 25 µM and 1600 µM with the AGG assay. Furthermore, the activity of PHM alone was measured by the inclusion of an alkaline hydrolysis step to lyse glyoxylate as a substitute for PAL catalytic activity.

Multiple candidate proteins and DNA sequences for PAM were identified by genetic sequence searches and a novel fungal PHM modelled *in silico*. Fourteen PAM, PHM, PAL and truncated constructs were expressed in the non-conventional yeast host, *Yarrowia lipolytica*. The novel fungal PHM's nutrient, temperature, and pH conditions were optimised to maximise protein expression. Enzyme purification was optimised with scalable industrial appropriate methods to purify milligram amounts of fungal PHM.

The AGG assay was validated with a commercially obtained PAM, demonstrating a simple medium-throughput method to measure PAM activity. The novel fungal PHM was characterised with a pH optimum of 4.0 and maximum enzymatic activity at 45°C. Deglycosylation of fungal PHM enhanced enzyme activity by 1.83 fold but lowered the temperature optimum to 37°C. The novel PHM and alkaline hydrolysis catalysed the conversion of the peptide pharmaceutical precursor for exenatide into its final bioactive form.

Chapter 1: Literature review

1.1 Introduction

Bio-active peptides are hormones, neurotransmitters, and growth factors that represent a major class of intercellular signalling agents in biology and a fast-growing class of pharmaceuticals (Murthy *et al.*, 1986; Vlieghe *et al.*, 2010; Cao *et al.*, 2011). Many of these bioactive peptides undergo significant post-translational modifications (PTM), and for the majority of peptide hormones, α-amidation of the carboxy-terminal (C-terminal) of the polypeptide is required for full biological activity (Eipper and Mains, 1988; Merkler, 1994). To date, α-amidated peptides have been extensively used in drug development and the treatment of diseases such as diabetes, cancer and HIV, amongst others (Latham, 1999; Kim and Seong, 2001; Schrader *et al.*, 2014). Peptidylglycine α-amidating monooxygenase (PAM) (EC 1.14.17.3) is the only known enzyme that catalyses the α-amidation of these bio-active peptides.

The distinction between a peptide and a protein has been arbitrarily defined, with a polypeptide described as a protein if it is longer than 100 amino acids (AA) in sequence length (Latham, 1999; Vlieghe *et al.*, 2010). Despite protein and peptides being most commonly synthesised in cells, often as prepropeptides, and found throughout nature, peptide pharmaceuticals are currently synthesised predominantly via chemical methods *in vitro*. These synthesis methods are expensive, environmentally harmful and inefficient as they produce limited quantities of active pharmaceutical ingredient (API) from the starting reagents (Kim and Seong, 2001; Vlieghe *et al.*, 2010).

An inexpensive method of producing these pharmaceutical peptides in large quantities has long been a priority in private and public sector research (Latham, 1999). The biological synthesis and α-amidation of peptides utilising PAM to catalyse the final PTM to activate these peptides could play a key role in this process with minimal environmental risks, increased API production quantities and lower production costs (Guzman *et al.*, 2007; Sheldon, 2007). The biological synthesis method offers avenues towards the use of a single production method scalable from laboratory to industrial volumes (Guzman *et al.*, 2007). Biological peptide synthesis would allow increased lengths of α-amidated peptides sequences to be synthesised beyond currently feasible or practical with present chemical synthesis methods (Merkler, 1994; Li *et al.*, 2013).

1.1.1 Amidation of peptides with PAM

For the PTM of all studied proteins and peptides to date, amidation is the fourth most frequent category of PTM observed (Khoury *et al.,* 2011). PAM activity is essential because it is primarily responsible for this specific PTM on C-terminal glycine precursor peptides (Eipper *et al.,* 1992). Crucially, PAM activity is independent of the penultimate amino acid identity on the peptide backbone compared to other protease activities (Eipper *et al.,* 1993). The only other protease used as a substitute for PAM to achieve C-terminal amidation was carboxypeptidase-Y (Hook *et al.,* 2004). However, limitations relating to substrate specificity and peptide degradation are severe drawbacks of these enzymes for industrial production (Kim and Seong, 2001; Hook *et al.,* 2004). The substrate specificity of carboxypeptidase-Y means it cannot amidate peptides with proline, glutamic acid or aspartic acid at the penultimate C-terminal amino acid position. Moreover, the production methods involving carboxypeptidase-Y utilise harsh reaction conditions, requiring additional reagents and processes, which cause peptide degradation (Kim and Seong, 2001). These limitations have curtailed their industrial use for α-amidated peptide production, resulting in limited commercial products from this method (Kim and Seong, 2001). These problems have made the study of PAM a vital research area since its discovery in 1985 as it offers a viable method for the industrial production of α-amidated peptides (Wand *et al.,* 1985; Handa *et al.,* 2012).

PAM is a bi-functional enzyme capable of catalysing the α-amidation of peptides via the N-oxidative cleavage of the glyoxylate residue of the C-terminal α-Carbon (αC) of glycine terminating peptides (Bradbury *et al.,* 1982; Murthy *et al.,* 1986). PAM consists of two distinct catalytic hetero-domains expressed on a single peptide backbone, which catalyses distinctive reactions. These domains are peptidylglycine α-hydroxylating monooxygenase (PHM) (EC 1.14.17.3) and peptidyl-α-hydroxyglycine α-amidating lyase (PAL), a.k.a peptidylamidoglycolate lyase (EC 4.3.2.5) (Bradbury *et al.,* 1982). Each domain can be expressed separately in various isoforms, but in mammalian systems, they are commonly spliced together into a single mRNA that translates as a single polypeptide (Kato *et al.,* 1990; Eipper *et al.,* 1992). PHM is a copper mono-oxygenase and catalyses the hydroxylation reaction of the α-carbon in the glycine residue of the substrate peptide. This hydroxylation is a critical step for subsequent PAL activity to cleave the peptide bond, thus exposing the C-terminal amine group.

Functionally PHM catalyses the stereospecific hydroxylation of the glycine αC of the C-terminal of the glycine residue on the peptide substrate (Tajima *et al.,* 1990; Bolkenius *et al.,*

1997). The only other enzyme to share a similar protein structure and function to PHM is dopamine beta-monooxygenase (DBM) (EC 1.14.17.1), the study of which has assisted in the understanding of the reaction mechanism of PHM (Li *et al.*, 1994).

The reaction catalysed by PHM involves the oxidation of two copper ions and the consumption of two oxygen molecules to hydroxylate the αC of the C-terminal glycine residue (Murthy *et al.*, 1986; Tajima *et al.*, 1990; Eipper *et al.*, 1993). The regeneration of the Cu ions requires oxidising two ascorbate molecules to semi-dehydroascorbate and releasing two water molecules. Subsequent to this reaction, PAL can then co-ordinate the nucleophilic attack of its zinc prosthetic group to the hydroxylated αC, catalysing the removal of the glyoxylate residue (Figure 1.1).

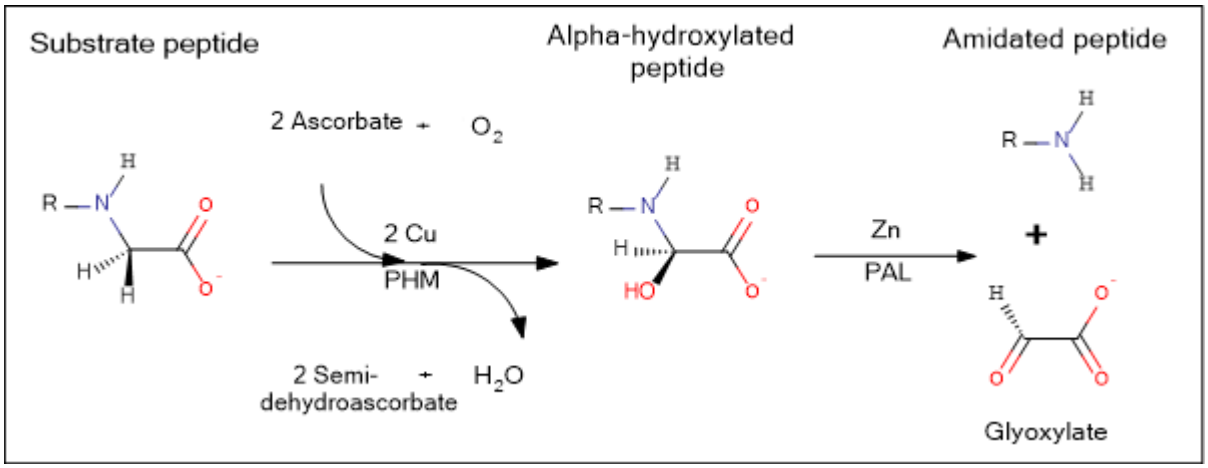


Figure 1.1: A mechanistic diagram of the α-amidating reaction catalysed by PAM to remove the glyoxylate residue from peptidylglycine. R = Rest of peptide substrate. The figure was adapted from Eipper, Stoffers and Mains, 1992.

The hydroxylated αC is an essential prerequisite for the lyase action of PAL to cleave the glyoxylate by-product from the amidated peptide final product (Eipper *et al.*, 1992). Before a further round of reactions can occur, both Cu prosthetic groups in PHM must be reduced to CuI by the oxidation of ascorbate to semi-ascorbate (Kolhekar *et al.*, 1997). This restoration time required to reduce the co-ordinated copper ions into their active valence state is a significant factor in the enzyme activity and forms a rate-limiting step. The highest turnover number recorded for a PHM was from *Rattus norvegicus*, was 37 s^{-1} with 2-(carboxyamino)-acetic acid under ideal *in vitro* conditions (Francisco *et al.*, 2004). By comparison, the maximum turnover rate for PAL from *R. norvegicus* was 380 s^{-1}, under comparable conditions (Kolhekar *et al.*, 2002). The PHM is thus considerably slower in terms of activity than PAL, which has a single Zn ion prosthetic group as its catalyst and requires no co-enzymes for

activity (El Meskini *et al.,* 2003). PHM thus forms the rate-limiting step in the overall peptidylglycine α-amidation reaction.

Compared to typical hydrolytic proteases, the use of two reaction mechanisms, several metal ions and co-enzymes, makes the overall PAM reaction both resource and time-intensive. The slow *in vivo* turnover rate of the enzyme, at 0.5 to 2 s^{-1}, is partly offset by the high concentrations of PAM protein within the endoplasmic reticulum (ER) and Golgi networks (Eipper *et al.,* 1992; Kumar *et al.,* 2016). Additionally, the peptide substrates' high molecular weight relative to the single labile bond on the peptide means a low total activity level could process a large mass of peptides. Thus within the ER, only nM concentrations of PAM are needed to process mM concentrations of peptides. Therefore, the low enzyme-specific activity reported can be presumed to result from the PAM's optimal sub-cellular topological location and minimal evolutionary pressure for higher activity levels (Eipper *et al.,* 1995; El Meskini *et al.,* 2003). Despite these disadvantages of low enzyme activity and the necessity of so many co-enzymes, the amidation of peptides is essential for both peptide stability and activity, thus necessitating this PTM.

The amidation of the C-terminus is a requirement for the majority of biologically active forms of neurotransmitter peptides to bind to their receptors (Eipper, Stoffers and Mains, 1992). In cases where amidation is not required for activation, C-terminal amidation has been found to enhance the activity of a peptide up to 10 000-fold (Edison *et al.,* 1999; Wapinski *et al.,* 2007). It has been theorised that the higher levels of bio-activity observed are related to the C-terminal amide residue rendering the peptide more hydrophobic, increasing its ability to bind to target receptors. Additionally, C-terminal amidation has been found to increase the *in vivo* half-life of peptides from 2-5 minutes to several hours with pharmaceutical peptides (Latham, 1999; Prigge *et al.,* 2000). This ability to activate peptides, increase their activity levels, and prolong their half-lives, is crucial for peptides produced by recombinant methods (Merkler, 1994). The majority of recombinant proteins are produced from yeast and bacteria, which cannot perform the C-terminal amidation PTM. This ability would enable the competition of recombinantly produced peptides with chemically synthesised peptides which currently are the primary industrial method used for the production of α-amidated peptides (Vlieghe *et al.,* 2010).

1.2 Chemical synthesis of pharmaceutical peptides

Peptides represent a fundamental class of biological chemicals synthesised in nature from ribosomal translation as the end product of numerous biological pathways. Despite these intrinsic biological properties, current peptides production is nearly exclusively by chemical methods (Lau and Dunn, 2017; Martin *et al.*, 2020). This paradox results from the historical and technological development of peptide synthesis with chemical processes as first achieved in 1953, predating the development of recombinant protein production methods (Vigneaud *et al.*, 1953). The initial approach for chemical peptide synthesis (CPS) was solution-phase synthesis (SPS), and this was significantly advanced with the invention of solid-phase peptide synthesis (SPPS) in 1963 (Vigneaud *et al.*, 1953; Merrifield, 1963). Consequently, molecular biologists and biochemists had limited participation in the organic chemistry-focused production process for peptides under 100 AA (Agyei and Danquah, 2011). Biochemists thus engaged with pharmaceutical peptides via a focus on the discovery process and the pharmacology of peptides. This production paradigm around organic chemistry and CPS has strengthened over time and limited the involvement of biochemists in peptide production. Reasons for this include strict regulations on API manufacturing around established and proven production methods (Martin *et al.*, 2020). The major hindrance to biologists and biochemists researching this field was the lack of biological and DNA recombinant technologies for the production and PTM of peptides (Guzman *et al.*, 2007; Reichart, 2010).

With current and future recombinant biological methods and production systems, it may be possible to replace CPS with biological synthesis for the majority of current commercial peptide pharmaceuticals. However, CPS is the current production method of choice and allows for unique peptide compositions and synthesis, such as inorganic residues, biologically uncommon elements, D-amino acids and other non-standard AA's (Stevenson, 2009; Vlieghe *et al.*, 2010). These modifications and non-biological chemical groups are either not possible to incorporate with current methods of recombinant protein production and commercial enzymes or have not been used in large scale protein synthesis (Guzman *et al.*, 2007). These unique properties of unnatural peptides can greatly assist in common problems with peptides such as delivery and *in vivo* stability.

The delivery of peptide pharmaceuticals to the patients' system represents a significant problem using oral delivery methods, as the digestive system rapidly degrades peptides. While methods such as intravenous (IV) injections are an effective delivery method, their daily use is

inconvenient and potentially hazardous to patients (Latham, 1999; Vlieghe *et al.*, 2010). Convenient methods such as nasal and pulmonary methods have been developed and are under constant research and development for greater effectiveness (Latham, 1999; Danquah and Agyei, 2012). These convenient methods, though, require larger quantities of peptides for a similar dose as IV infusion. The cost and quantities required are not suitable to the current CPS economic and production paradigms with its incumbent high per unit production costs and capital intensive equipment (Latham, 1999; Guzman *et al.*, 2007).

1.2.1 Synthetic peptide therapeutics market

In the USA, a chemically synthesised peptide has been defined by the Biologics Price Competition and Innovation Act of 2009 as any alpha amino acid polymer synthesised entirely by chemical means and less than 100 amino acids in size (FDA, 2009; Wang and Chow, 2017). In 2009 there were 40 pharmaceutical peptides on the commercial market, with an estimate of over 100 in various stages of development (Zompra *et al.*, 2009). At that stage the peptide drug market was over $15 billion annually in 2012, with the majority (greater than 85%) from synthetic peptides (Glaser, 2013b). The pharmaceutical peptide market was forecast to have growth in sales figures above 10% annually over the period 2013-2018 (Glaser, 2013b). The market trend is due to both the continued growth of established peptide products and the increase in FDA and regulatory approval rates. By 2019 the total had increased to over 100 peptide pharmaceuticals on the commercial market in Europe, U.S., and Japan, with 14 authorised for sale between 2016-2020 (FDA, 2020; Al Musaimi *et al.*, 2021).

In 2010, there were only 17 peptides that are feasible for production with methods that did not require amidation or other PTM's (Vlieghe *et al.*, 2010). The total represents approximately 24% that could be produced cheaply in prokaryote or yeast systems at that time. In addition, only seven of those peptides were longer than 15 AA and could therefore be economically competitive when produced with recombinant expression as compared to CPS (Vlieghe *et al.*, 2010). Subsequently, the average peptide pharmaceutical length has increased from 15 in the 1990s to over 20 in the 2010s, and the percentage share of native homologous peptides has increased as well (Lau and Dunn, 2018).

Despite advances with recombinant peptide expression, mainstream manufacturers have not pursued this process (Kim and Seong, 2001; Lau and Dunn, 2018). The adoption of a new production process is viewed as not commercially viable to fund the research, clinical trials and accreditation of a novel method of peptide production for peptides only representing a

fraction of the market (Guzman *et al.*, 2007; Lau and Dunn, 2018; de la Torre and Albericio, 2019; Puetz and Wurm, 2019). Thus, the majority of peptides are produced by synthetic chemical methods, with the resulting prices 8 to 9 orders of magnitude higher per kg than industrial enzymes made with recombinant production methods (Puetz and Wurm, 2019; Al Musaimi *et al.*, 2021). A thorough understanding of these methods' relative strengths and weaknesses is required to accurately compare them to the potential production of peptide pharmaceuticals by recombinant systems.

1.2.2 Methods of chemical peptide synthesis

Current peptide production systems include CPS, recombinant peptide production and natural product purification (Guzman *et al.*, 2007). Of primary importance are the CPS methods, as these currently form the majority of approved methods for API production (Kim and Seong, 2001; Guzman *et al.*, 2007). CPS consists of three classes based on the techniques used: SPS, SPPS, and hybrid synthesis systems, which combine both SPS and SPPS (Vlieghe *et al.*, 2010; Uhlig *et al.*, 2014). Significant problems with SPS are the lengthy method development processes and the extensive analytical testing required to produce peptides before and during early clinical trial stages (Uhlig *et al.*, 2014). Despite these drawbacks, SPS remains the preferred production method of small to medium (2-10 AA) sized pharmaceutical peptides in sub-kilogram quantities (Glaser, 2013a; Thundimadathil, 2013).

SPPS is the primary method employed to produce pharmaceutical-grade purity peptides in kilogram quantities, but this often encounters complex development processes to switch over from SPS (Uhlig *et al.*, 2014). The amount of peptide produced is vital for its pharmaceutical applications as peptides have problems with delivery and short *in vivo* half-lives that require large doses to counteract this shortcoming compared to small molecule drugs.

1.2.3 Problems and restrictions with chemical peptide synthesis

A significant problem with all CPS methods are decreases in peptide fidelity with increases of marginal peptide length which exacerbates final losses of peptide product (Guzman *et al.*, 2007). Polypeptides greater than 40 AA are challenging in terms of fidelity to the template peptide and limited in terms of quantity (Latham, 1999; Kolb *et al.*, 2001). Production of kilogram quantities of +40 AA peptide has been consequently economically prohibitive. Still, click chemistry with component peptide ligation offers a potential production method to counteract this, but this still incurs more costs, generates additional wastes, and adds more production steps (Kolb *et al.*, 2001; Li *et al.*, 2013).

Fidelity issues stem from the higher incumbent error rate in CPS compared to biological peptide synthesis (Guzman *et al.*, 2007). Compounding this fundamental flaw of CPS are the requirements of side-chain blockers and highly pure reagents during synthesis, which cumulates in slow production times and high costs (Guzman *et al.*, 2007). Finally, the constant purification of the reaction solutions to remove impurities and by-products generated by these processes and reactions and the safe disposal of solvents and resins further economically strains this process relative to biological synthesis (Guzman *et al.*, 2007).

1.2.4 Environmental waste and production efficiency

SPPS costs are increasing with the costly disposal of waste produced during manufacturing and the increasing costs of environmental waste legislative compliance (Sheldon, 2007; Hojo *et al.*, 2011). Typically for synthesising a kg of 20 AA peptide with SPPS, over 1000 litres of solvent waste are generated (Thundimadathil, 2013).

Alternative peptide synthesis techniques are under investigation for reducing the costs associated with conventional peptide synthesis techniques (Hojo *et al.*, 2011; Thundimadathil, 2013). Research has been accelerated as more stringent legislation is enacted relating to the use and disposal of organic solvents and as public pressure increases the movement of industry to greener practices (Hojo *et al.*, 2011). A useful measure of the waste produced by the chemical industry is the E-factor, which is defined as the quantity of waste generated per quantity of final product (Sheldon, 2007). In this respect, the pharmaceutical industry is amongst the worst, with an E-factor of 25 to 100 for tonnage quantity of final product (Sheldon, 2007). Of greater concern is that peptides are at the high end of the E-factor range estimation. The increasing peptide lengths being researched and produced are set to far exceed these levels (Hojo *et al.*, 2011). In stark contrast, the petroleum industry has an E-factor of less than 0.1 while producing 10^8 metric tonnes of final product. The high E-factor is also compounded by the nature of the waste produced, which in the case of chemical peptide synthesis have high remediation, handling, storage and disposal costs (Hojo *et al.*, 2011). These factors are accelerating the movement within the pharmaceutical industry to adopt greener methods with greater sustainability (Sheldon, 2007; Hojo *et al.*, 2011).

An advantage of the fermentation production route of potential therapeutic peptides may be the rapid scale-up through all stages of development. A single recombinant host could produce small scale (mg to g) quantities of peptide needed for R&D to clinical trial level quantities (g to kg) to industrial-scale production (kg to metric tonnes), with limited fermentation and

peptide quality testing at each step. The product testing at each stage of production is required for certification by multiple regulatory authorities to ensure API fidelity (Guzman *et al.*, 2007). The regulatory burden can become prohibitive, as small scale CPS methods, such as SPS, could be incompatible with larger-scale production methods, such as SPPS (Guzman *et al.*, 2007). Problems include prohibitively high costs and technical challenges in the conversion of production between processes, including the inability to create the same peptide. These difficulties have cancelled the development of some potential peptide therapeutics due to a lack of feasible production and unfavourable economics (Guzman *et al.*, 2007). These stalled drugs could be resurrected with the paradigm change to biological synthesis, potentially offering novel production methods and lower costs (Merkler, 1994).

The use of large-scale yeast fermentation to manufacture peptides provides a greener and highly sustainable method for peptide synthesis. The final product yield of the peptide API could still be comparable to current CPS methods and cater to multiple scales and end-users. By contrast, the majority of the waste produced could be sustainably disposed of, as exemplified by modern brewery plants (Simate *et al.*, 2011). This method of peptide production could satisfy all the standards of green chemistry, as non-conventional yeast like *Yarrowia lipolytica* can utilise wastes such as glycerol and yeast extract (Sheldon, 2007). A further advantage is the low production cost with yeast fermentation methods due to standardised bioreactor operations (Roth *et al.*, 2009). Compared to the expensive, highly pure reagents required for mammalian systems, the inexpensive media and operational costs make a compelling argument for yeast-based recombinant peptide systems (Kim *et al.*, 2012).

1.3 Yeast expression of peptides, enzymes, and proteins

Since prehistory, mankind has used microbial fermentation to produce a variety of proteins and fermentation products. With the advent of modern biotechnological techniques, it is possible to understand and control this process to produce novel proteins. The use of recombinant peptides as drugs was established with the use of recombinant human insulin from *Escherichia coli* in 1981 and has grown from there to a multi-billion industry (Keefer *et al.*, 1981). Research into and the use of biologically expressed proteins has continuously grown and is typically performed in several model organisms. A list of the most important common recombinant expression hosts typically includes *Escherichia coli* for bacteria, *Saccharomyces cerevisiae* for yeast, *Aspergillus niger* for fungi, *Drosophila melanogaster* for insects, and finally *R. norvegicus* and *Homo sapiens* for mammals. To date, PAM and PHM and PAL in some studies

have been used in expression studies with these organisms to varying degrees of success (Bauman *et al.,* 2007).

For the production of large quantities of heterologous proteins that require post-translational modification, the predominant systems in use are fungal and, in particular, yeast systems (Porro *et al.,* 2011). A promising system is a non-conventional yeast, *Y.lipolytica*, first identified in the early 1960s (Liu *et al.,* 2015). Its potential for biotechnological applications was quickly recognised due to its ability to grow on lipid-rich substrates, high production levels of inorganic acids and prolific secretion of proteins (Barth and Gaillardin, 1996). Initially, this organism was identified as *Candida lipolytica* in the late 1960s but is today classified in its own genus as *Y. lipolytica* (Barth and Gaillardin, 1996). Since its discovery, it has undergone extensive genetic studies and has a proven ability to both produce and secrete heterologous protein from diverse genetic sources in concentrations as high as g/L (Hofmeyer *et al.,* 2014).

Y. lipolytica has some of the key genetic advantages in comparison to other common yeast systems such as *S. cerevisiae* and *Pichia pastoris* (Madzak *et al.,* 2004). *Y. lipolytica* secretes protein primarily through a co-transcriptional pathway and it has a high secretion capacity with minimal glycosylation potential similar to mammalian glycosylation patterns (Nicaud *et al.,* 2002; Bankar *et al.,* 2009). Additionally, it is a non-pathogenic organism and an obligate aerobe suitable for cell density fermentations (Madzak *et al.,* 2004; Kahari *et al.,* 2007; Bellou *et al.,* 2014). However, *Y. lipolytica* is sensitive to oxygen concentrations, and growth temperatures may not exceed 32°C (Beopoulos *et al.,* 2009; Bellou *et al.,* 2014).

The use of *S. cerevisiae* as a recombinant expression host has resulted in the production of active PAM, but not in large industrially relevant quantities (Bauman et al. 2007; El Meskini et al. 2003). PAM was likely being directed to cytosolic and not secretory pathways, resulting in nanogram production of PAM per litre of culture (El Meskini et al., 2003). Attempts with the PHM catalytic core (cc) resulted in the same intracellular retention, but PHM could only be detected after media concentration and western blot analysis (El Meskini et al., 2003). It was observed that the PHM produced by the wild type yeast had activity levels comparable to that produced by chinese hamster ovary (CHO) cells. These results suggested that neither the inclusion of copper chaperones nor alterations to the copper trafficking pathways in yeast would be necessary for expressing active PHM (El Meskini et al., 2003).

The use of *Y. lipolytica* as a heterologous protein expression host is promising as it is an established system for use with novel sequences of PAM. This factor was important as previous

expression attempts of *R. norvegicus* PAM or PHM in *E. coli* or yeast were unsuccessful in terms of negligible (ng/mL) production of PAM (Kolhekar *et al.*, 1997; El Meskini *et al.*, 2003; Bauman *et al.,* 2007). However, up to 15 mg/day with CHO cells in a highly automated bioreactor was possible, but this is rendered prohibitive for research and industry in terms of expensive running costs and elaborate equipment and problematic due to enzyme degradation and activity losses (Bauman *et al.,* 2007). To overcome these high production costs, the use of non-conventional yeast with a novel PAM or PHM sequence could be a potential solution.

Identifying a PAM or PHM homologue in the fungal kingdom may offer a novel PAM or PHM expression attempt in yeast that overcomes the host protein degradation systems as well as trafficking and secretion issues observed by El Meskini *et al.,* 2003. Potential PHM sequences have been found in species as diverse as the gram-positive soil bacterium *Sorangium cellulosum*, invertebrates such as *Drosophila melanogaster* and the flagellate algae *Chlamydomonas reinhardtii* indicating the early establishment of this enzyme in the evolutionary history of life (Attenborough *et al.*, 2012; Kumar *et al.*, 2019). Despite PAM being conserved across Eukaryotes in multiple kingdoms, no fungal PAM or PHM homologues have been identified to date (Attenborough *et al.*, 2012; Kumar *et al.*, 2016, 2019).

1.4 Peptidylglycine α-amidating monooxygenase domains, structure and mechanism of activity

The key to understanding the function and activity of an enzyme is the protein structure which helps elucidate fundamental enzyme properties such as enzymatic catalysis at the active site. A solved protein structure allows the rapid and logical identification of critical residues and a guide to describing the catalytic mechanism. The canonical reference for a mammalian PAM structure is the PAM-1 isoform from *Rattus norvegicus* (Stoffers *et al.,* 1989). PAM consists of two catalytic domains, PHM and PAL, with several distinct non-catalytic regions, namely, a linker region, a transmembrane and a Ras association domain family member 9 (RASSF9) interaction domains (Figure 1.2) (Kolhekar *et al.*, 1997; Vishwanatha *et al.*, 2014).

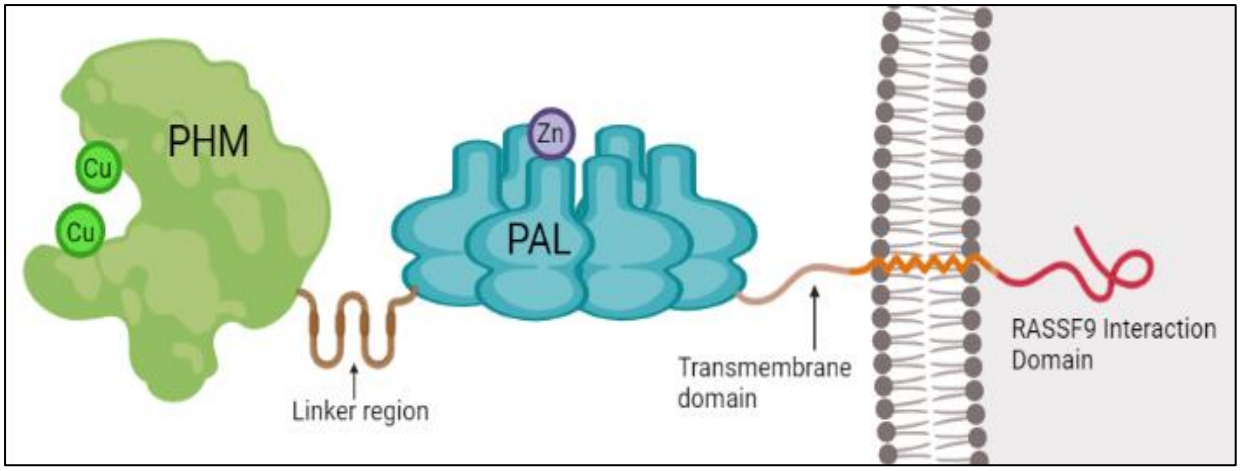


Figure 1.2: A representation protein domain and region of *Rattus norvegicus* PAM-1 isoform protein, created with BioRender.com.

Although the complete PAM structure has not been crystallised or modelled, the general protein structure can be speculated based upon the structurally mapped domains, PAM and PAL, while using sequence homology to predict the remaining regions and domains (Prigge, 1997; Chufán *et al.*, 2009).

1.4.1 Peptidylglycine α-hydroxylating monooxygenase

PHM is one of eleven mammalian copper-dependent enzymes, including tyrosinase, superoxide dismutase and dopamine β-monooxygenase (Prohaska, 2014). As such PAM has been studied for its role in copper metabolism and diet in both humans, animal and fungal models (Kolhekar *et al.*, 1997; El Meskini *et al.*, 2003; Prohaska, 2014). This dietary requirement and unique mechanism of PHM has necessitated the study of the enzyme structure for insights into its unique activity and functions in metabolism.

1.4.1.1 PHM Structure

Several truncated PHM catalytic cores, including examples with enzyme activity, have been described using X-ray crystallography studies to form protein structural models (Figure 1.3) (Prigge, 1997; Rudzka *et al.*, 2013).

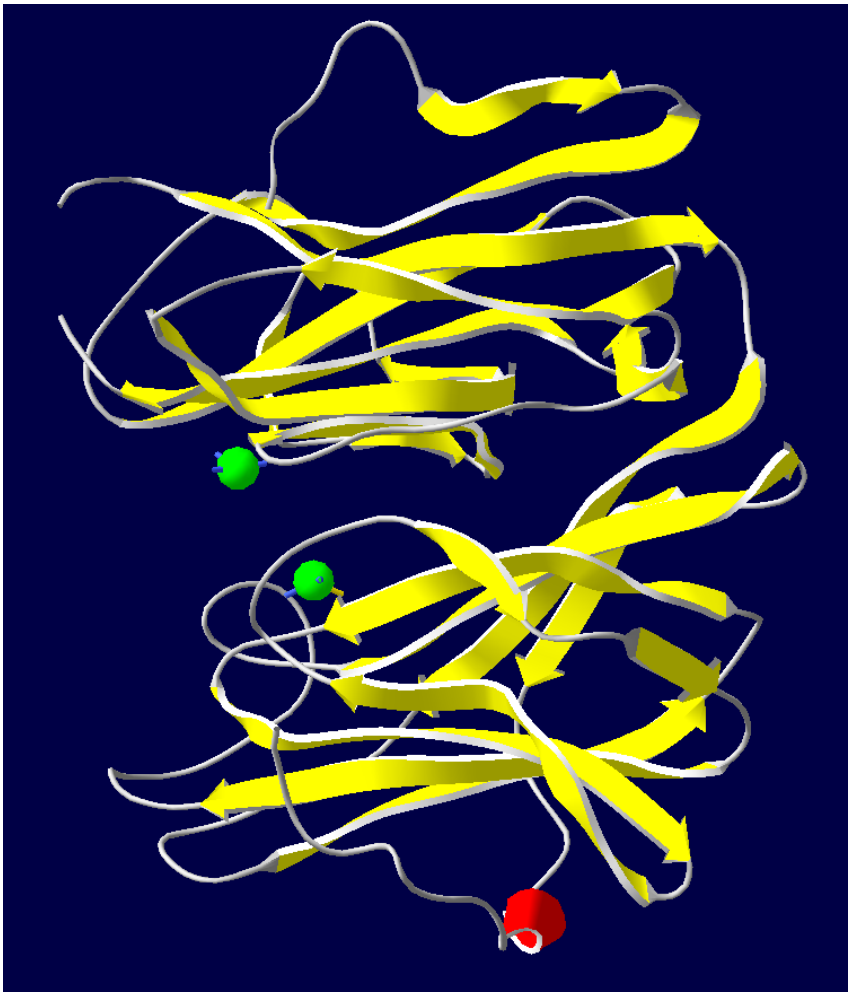


Figure 1.3: A ribbon-style representation of PHM peptide backbone displaying the secondary structure of *Rattus norvegicus* PHM created from 1phm protein data bank file from Prigge *et al.,* 1997. Beta strands as yellow ribbons, alpha-helix in red and the two copper ions in green.

The overall structure presented by PHM is a prolate epsiloid with two topologically similar domains, each binding a Cu ion (Cu_H and Cu_M) (Bradbury and Smyth, 1987; Prigge *et al.,* 2000; Siebert *et al.,* 2005). Each domain contains 9 beta-strands of 150 residues and shares an eight-stranded antiparallel jelly motif (Prigge, 1997). However, when the domains are structurally aligned, they have a sequence identity of less than 5% (Prigge *et al.,* 2000). The domain interiors are highly hydrophobic, creating a solvent accessible interdomain space containing the active site residues and both non-covalently bound copper ions (Prigge, 1997). The Cu_H is bound via His^{107}, His^{108}, and His^{172} in the Domain H, while Cu_M is bound via His^{242}, His^{244} and Met^{314} (Siebert *et al.,* 2005). These residues perform critical structural properties and functional activities in the protein structure.

The Cu_M has a proven role in the overall enzyme structural stability as it anchors a 15-residue loop at positions 299-314 (Siebert *et al.*, 2005). The coordination of Cu_M site with its Cu ion also influences and stabilise the coordination of the Cu_H site, which, despite a separation of 11 Å, implies a structural relationship between Cu binding sites and their ions (Siebert *et al.*, 2005). The proximity may be crucial as it adjusts the reduced coppers' electrochemical potentials, which allows for the favouring of an electron transfer, increasing enzymatic catalysis (Siebert *et al.*, 2005).

The metal-binding residues, in the Cu binding sites, are amongst the most highly conserved AA in PHM, with mutations of the histidine residues resulting in loss of enzyme activity (Prigge *et al.*, 2000). The only exception to this was a mutation of M314Y, which resulted in a sevenfold increase of K_M with no effect on V_{MAX} (Prigge *et al.*, 2000). Interestingly all six of the Cu binding site residues are also conserved in the closely related DBM, as literature and protocols relating to DBM have aided in research into PHM (Prigge, 1997; Wimalasena and Wimalasena, 1991).

Other highly conserved residues in the domain include cysteine residues that form six disulphide bridges helping to anchor solvent-exposed loops that connect domain A with two strands from separate sheets in domain B (Prigge, 1997; Attenborough *et al.*, 2012). The conserved nature of these AA's is crucial as they form the basis for the identification of PHM catalytic cores (cc) when searching genomic or proteomic databases for novel PHM or PAM sequences in general (Attenborough *et al.*, 2012).

1.4.12 Peptidylglycine α -hydroxylating monooxygenase mechanism

The copper monooxygenase activity of PHM has undergone extensive study in the 30 years since its discovery, with research into the exact details of the reaction mechanism ongoing (Abad *et al.*, 2014). The accepted model of the reaction mechanism for PHM is that it requires two ascorbates and one O_2 to hydroxylate the αC on the peptidylglycine residue (Murthy *et al.*, 1986; Prigge, 1997). The reaction mechanism at the PHM active site with a peptidylglycine substrate is displayed in Figure 1.4.

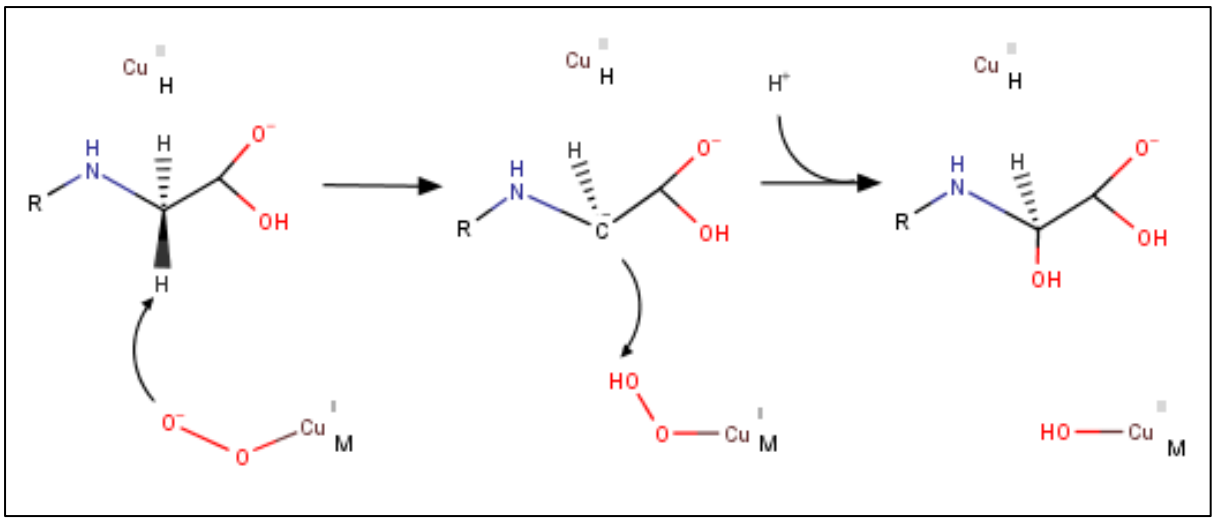


Figure 1.4: A mechanistic diagram of the PHM domain catalysing hydroxylation of the alpha carbon residue of a peptidylglycine substrate (Prigge et al., 2000). R = Rest of peptide.

The hydroxylation reaction of the Cu_M ion is initiated with a nucleophilic attack from the bonded oxygen to the proximal hydrogen of the αC of the terminal glycine residue of the substrate, Figure 1.4 (Prigge *et al.,* 2000; Siebert *et al.,* 2005). This, in turn, disassociates the hydrogen to the Cu_M β-oxygen resulting in the oxygen's nucleophilic attack. The results of which are the formation of a hydroxyl residue and copper in valence state II (Prigge *et al.,* 2000; Rudzka *et al.*, 2013). During this reaction, the Cu_M-O residue is reduced to Cu_M-OH by the incorporation of an H^+ ion. The resultant reduced Cu subsequently requires the oxidation of two ascorbate molecules or related oxidants to restore activity to the PHM enzyme (Rudzka *et al.*, 2013). The restoration step with the ascorbates is the cause of the relatively low turnover rate of the PHM subunits. The overall PAM amidation reaction is thus rate limited by these reaction conditions and requirements for PHM activity. By contrast68, the PAL reaction mechanism requires no restoration or co-factors for activity and has a higher relative turnover rate.

1.4.2 Peptidylglycine α-amidating lyase

With the hydroxylation of the αC by PHM completed, the final reaction of the α-amidation of the peptide substrate is catalysed by the PAL portion of the PAM enzyme (Eipper *et al.*, 1993). The activity of this enzyme is vital to complete the amidation reaction as the spontaneous catalysis of the amide bond is not possible under the acidic conditions of the ER and trans-Golgi network (TGN).

1.4.2.1 PAL Structure

PAL has been expressed as a distinct isoform from PAM, as observed with PHM, and represents a novel structure compared to PHM (Figure 1.5) (Kolhekar *et al.*, 1997, 2002; Chufán *et al.*, 2009).

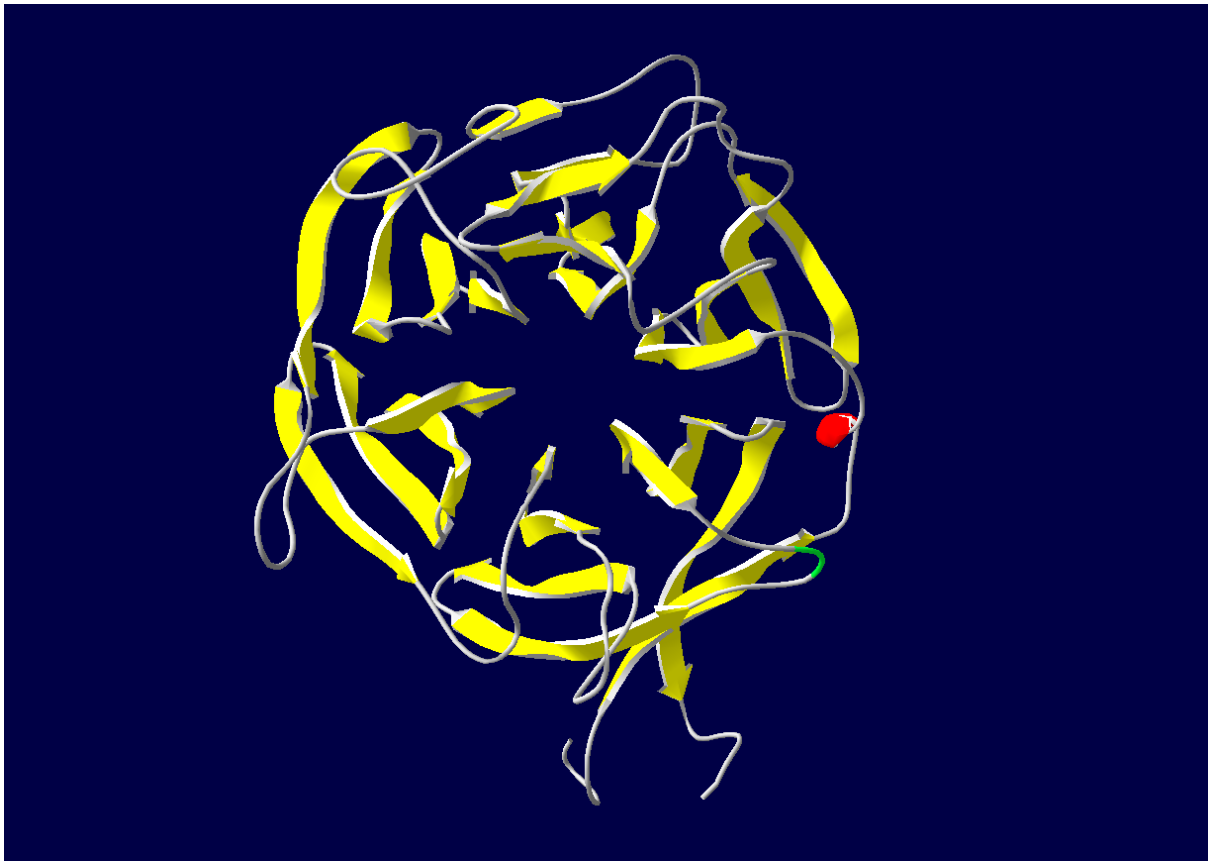


Figure 1.5: A ribbon-style representation of PAL peptide backbone displaying the secondary structure of *Rattus norvegicus* PAL created from 3fvz protein data bank file from Chufán et al., 2009. Beta strands as yellow ribbons, and the alpha-helix in red.

The PAL structure includes four NHL (NCL-1, HT2A and LIN-41) family type repeats that define its tertiary structure (Chufán *et al.*, 2009). These NHL repeats and ions help stabilise PAL in a β-propeller confirmation consisting of 6 blades, all centred on a central cavity. Each blade is composed of 4 antiparallel strands. The first strand of each blade is located adjacent to the central cavity, and thereafter each strand looping back on the previous strand until the fourth and final strand at the protein edge (Chufán *et al.*, 2009). The central pore created by these β-strands contains the Zn and K ions, which respectively help the catalysis and enzyme stability of PAL (Ul-Hasan *et al.*, 2013).

The Zn ion is located at the bottom of the pore and is coordinated by three Nε from three histidine residues (His^{585}, His^{690} and His^{786}) (Chufán *et al.*, 2009). The side chains residues are

not in a strict geometric formation but instead form an intermediate between tetrahedral and octahedral geometries (Chufán *et al.*, 2009). The instability of an intermediate geometry may be negated by the presence of proline (Pro^{584}, Pro^{689} and Pro^{785}), preceding each histidine, and help coordinate the Zn ion (Chufán *et al.*, 2009). The Zn, histidines and prolines together help form the active site of PAL, as evidenced by the presence of $Zn(His)_3$ at the active sites of enzymes such as β-lactamase and carbonic anhydrase (Toney *et al.*, 2001; Krishnamurthy *et al.*, 2008; Chufán *et al.*, 2009). PAL requires no co-factors to catalyse the release of the glyoxylate residue (Chufán *et al.*, 2009).

The presence of these conserved residues and critically the NHL repeat regions are crucial in terms of identification of the novel PAL cc. Problematically, PAL protein structures and sequences can interfere in genomic and proteomic searches for novel PAM sequences. PHM structures and C-terminal amidation activity has been found to occur both in bacteria and algae without an associated PAL domain (Attenborough *et al.*, 2012; Kumar *et al.*, 2019).

1.4.2.2 Peptidyl-α-hydroxyglycine α-amidating lyase mechanism

The PAL is a metalloprotease lyase that requires a hydroxyl residue on the αC of the glycine at the C-terminus of a peptide substrate. To determine PAL activity, the predominant substrate used in literature is alpha-hydroxyhippuric (αHA), the hydroxylated version of the common PAM substrate hippuric acid (HA). The PAL activity with αHA is clearly defined and provides a comprehensive description of the PAL reaction mechanism (Figure 1.6) (Eipper *et al.*, 1992).

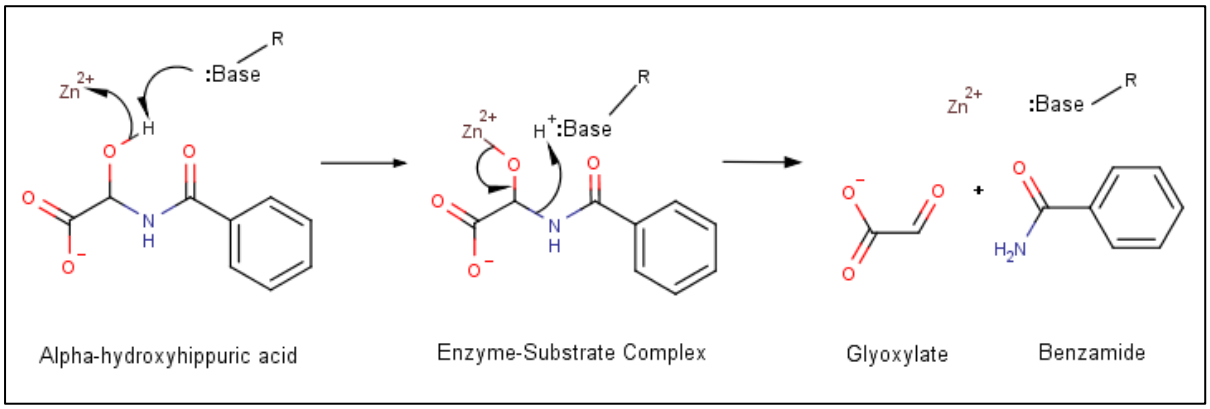


Figure 1.6: A mechanistic diagram of the PAL domain active site catalysing the removal of the glyoxylate residue from the alpha-hydroxyhippuric acid peptide substrate. R = Rest of enzyme.

The reaction is catalysed by the nucleophilic attack of the Zn ion on the α-amidated glycine residue of a substrate peptide. The presence of Zn is essential for activity with activity negated by the removal of Zn with the addition of chelating reagents (Eipper *et al.*, 1992).

The general pH requirement of PAL activity and PAM activity is acidic, between pH 4.5 and 6.5. There are a few exceptions, such as *Bos taurus* PAM (Eipper and Mains, 1988). The acidic requirement of mammalian PHM is likely due to low pH levels found in TGN and the ER. These conditions can inhibit the autocatalysis and removal of the glyoxylate residue. The hydrolysis of the glyoxylate residue from a peptidyl-α-hydroxyglycine residue on substrates is spontaneous at pH levels greater than 7.0 (Katopodis *et al.*, 1991). Thus, it is possible to circumvent the requirement for PAL activity in PAM amidation reactions if a peptide containing solution has the pH altered pH>7.0. The possibility of alkaline hydrolysis for PHM peptide products without PAL offers a potential explanation for the conservation of only the PHM domain in many organisms.

1.4.3.1 PAM linker region

PAM can be expressed or processed as a whole bi-functional enzyme or separate PHM and PAL enzymes due to RNA splicing or PTM (Eipper *et al.*, 1993; Vishwanatha *et al.*, 2014). The canonical *R. norvegicus* isoform PAM-1 is expressed with a linker region of 141 AA, with other isoforms containing a truncated version of 36 AA or existing as separate enzymes (Eipper *et al.*, 1993). The expression of catalytic cores without this linker region indicates neither catalytic nor essential structural function. However, the major splice variant, PAM-1, contains this linker region as it assists in the vesicle trafficking of *de novo* translated PAM (Vishwanatha *et al.*, 2014).

The linker region has not been structurally modelled, but investigation with far UV circular dichroism and NMR studies indicate it is unstructured (Kolhekar *et al.*, 1997; Vishwanatha *et al.*, 2014). This lack of ordered structure may benefit the enzyme in its proposed role as a pH sensor to assist in the intra-vesicle trafficking of PAM isoforms (Vishwanatha *et al.*, 2014). Crystallisation studies of the bifunctional PAM protein, including successful methods used to determine PAL and PHM structures, have not resulted in a crystal structure (Vishwanatha *et al.*, 2014). The use of cryo-electron microscopy may resolve the whole PAM structure to a suitable level (Callaway, 2015).

PAM is the final step in the synthesis of bioactive alpha amidated peptides and thus plays a role in the feedback and trafficking of many intracellular pathways and the endocrine system (Eipper *et al.*, 1992; Alam *et al.*, 2001; Gaier *et al.*, 2014; Vishwanatha *et al.*, 2014). This position as gatekeeper allows PAM to function as a rate-limiting step for bioactive peptide synthesis and signalling between cells with neuropeptides or metabolic hormones. As

transmission speed can be critical with cellular messaging and given that PHM and PAL were expressed as a single peptide, it was proposed that these domains were proximally located to increase activity levels. Studies into the potential for substrate channelling between the catalytic domains of PHM and PAL have not shown substrate channelling, and it is likely to offer only minor improvements in K_{CAT} (Moore and May, 1999).

The presence of isoforms and linker regions implies that the enzyme structure of PHM and PAM is capable of maintaining activity and structural stability despite various attachments, such as linker regions or PAL (Eipper *et al.*, 1995; Vishwanatha *et al.*, 2014). It is likely PAM could be immobilised, as each catalytic core has no membrane-associated elements. PHM is theorised to remain exposed entirely to the solution while remaining proximate to the cell wall due to the attached PAL transmembrane domain when expressed as PAM-1 isoform from *R. norvegicus* (Kim and Hahn, 2012). The discovery and recombinant expression of an extracellular version of PAM or PHM or an enzyme amenable to immobilisation would be highly desirable from an industrial standpoint.

1.5 Enzyme immobilisation

The utility and cost-effectiveness of any potential industrial PAM amidation system would be improved by including elements to increase enzyme reusability and simplify purification. Critical to immobilisation and purification method selection is the care taken to preserve enzyme stability whilst minimising substrate contamination for subsequent purifications (Guzman *et al.*, 2007). Any potential PAM would be used with a multitude of peptide precursors under batch production conditions, and any method to rapidly separate and transfer enzymes between batches would be advantageous. The low turnover rate of PAM means that the processing of g to kg batches of peptide will require the use of mg to g levels of PAM to process peptides under reasonable time frames and economically viable conditions (El Meskini *et al.*, 2003; Francisco *et al.*, 2004).

Possible solutions to this problem include engineering the enzyme to be efficiently purified from the peptide solution(s) for re-use or immobilisation onto a solid support matrix. The immobilisation can minimise contamination of the substrate and reaction solution by the enzyme. These factors enable subsequent re-use while reducing downstream purification costs of the peptide products (Wang *et al.*, 2009). Ideally, immobilisation would increase the enzyme activity half-life, re-use the enzyme between batches of differing peptides, increase pH and thermal stability while minimising enzyme activity losses and costs from purification (Cipolatti

et al., 2014). These improvements have been observed with current enzyme immobilisation and purification techniques.

1.5.1 Immobilisation techniques

Typically immobilisation has included either covalent or non-covalent bonds to attach an enzyme to a support matrix allowing submersion of the enzyme with the substrate (Cipolatti *et al.*, 2014). Common immobilisation examples include the enzyme attachment to nanowires, enzyme aggregation, enrichment of a soluble gel-type matrix, vessel surface coating, immobilisation to a solid support scaffold, and immobilisation onto magnetic microspheres (MMS) (Cipolatti *et al.*, 2014; Garny *et al.*, 2014).

A standard method is glutaraldehyde covalent bonding to glass or paper surfaces with crosslinking. It has been shown to have great potential for medical, chemical and pharmaceutical industries, particularly for proteolytic enzyme immobilisation (Migneault *et al.*, 2004). Immobilisation efficiency with trypsin and glutaraldehyde can reach as high as 95% with various supports and even increase enzyme activity and stability (Migneault *et al.*, 2004). However, disadvantages for this include the permanent nature of the covalent bonds meaning support and enzyme must be replaced eventually, and neither the support nor enzyme are not reusable (Cipolatti *et al.*, 2014). A disadvantage of covalent bonds is the potential reduction and restriction of access to the enzyme active site (Cipolatti *et al.*, 2014). In addition to blocking the active site, non-specific bonds can induce non-active conformational states and even sterically unravel the enzyme (Cipolatti *et al.*, 2014). The testing of multiple immobilisation strategies and materials may bypass these problems.

1.5.2 Magnetic microspheres

A promising new technology is MMS, as this is used for both the purification of enzymes and their use as an enzyme scaffold in the production of chemicals and pharmaceuticals (Zhou *et al.*, 2015). Used in conjunction with Ni-affinity immobilised metal affinity chromatography (IMAC) and multiple histidine purification tag, it allows selective enzyme retention or elution into solution (Garny *et al.*, 2014). This combination has minimal effect on enzyme structural or forms covalent bonds with the principles of His-tag IMAC well understood and easy to use. Additionally, the enzyme access to the substrate is optimal when the enzyme is eluted into solution. The use of MMS and their solution accessible interiors, enzyme elution and purification may not even be necessary, with solution conditions altered to adhere an enzyme and transfer the MMS and enzyme between batches (Garny *et al.*, 2014). When combined with

a PAM amidation system, these attributes could offer a novel and economically efficient system for peptide pharmaceutical production (Zhou *et al.*, 2015).

1.6 Peptide pharmaceuticals

The current pharmaceutical sector is a dynamic, research and development (R&D) intense sector with over USD$ 74 billion spent in 2019 (Puetz and Wurm, 2019). The high costs are partly due to the low success rate of developing pharmaceuticals, which requires high-throughput screening of large libraries and multiple failures of candidate peptides in development (Lau and Dunn, 2017; Martin *et al.*, 2020). The high R&D costs are further compounded by clinical trials and regulatory review costs (International Federation of Pharmaceutical Manufacturers & Associations, 2012). Total costs can amount to over $ 1.4 billion spent for a single new drug in 2019 compared to $ 138 million in 1975, in inflation-adjusted $USD (Puetz and Wurm, 2019).

The R&D cost is recovered with high prices to extract the maximum marginal profit per pharmaceutical product sold, limiting healthcare access. The costs and lack of treatment are amplified in price-sensitive developing countries and often encounter multi-year delays in introducing the market of proven medicines (McCamish and Woollett, 2011). An obvious means to lower marginal costs of the drug in commercial production, and to minimise drug costs in R&D, preclinical, and clinical phases, is to increase economies of scale and ease scalability for peptide APIs with recombinantly produced biological peptides.

A problem with many API produced with current CPS methods is the batch nature of the production systems and lag time in securing upstream chemical reagents before production can begin (Glaser, 2013b). By contrast, the scale-ability of biological systems means that only simple media and bioreactors are required (Roth *et al.*, 2009). Thus provided downstream equipment and reagents are in place and operational, a new batch of API could be prepared in days or weeks based on market demand and urgency of supply, greatly assisting research into novel peptides (Wapinski *et al.*, 2007).

1.6.1 Research areas for peptide pharmaceuticals

Current peptide pharmaceutical research focuses on six broad research areas for novel therapies and new diagnostic kits (Latham, 1999; Al Musaimi *et al.*, 2018). These include cancer, anti-biotics (especially anti-microbial and anti-fungal), viral indications, immune system disorders, cardiovascular disorders, and neurological disorders (Latham, 1999). The current peptide

pharmaceutical market offers treatments for all the above diseases and disorders was estimated at over $15 billion in 2021 (FDA, 2020). The market share is set to grow for peptides due to their high success rate at twice that of conventional small molecule drugs and growth of biosimilars receiving approval (Craik *et al.*, 2013; Walsh, 2014).

The market share and development success has stimulated a growing interest in peptides, especially anti-microbial peptides (AMP), which come from various sources such as arthropods, fungi, and bacteria (Bulet *et al.*, 2004; Huang *et al.*, 2010). The high success rate and fast development times could be decisive in combatting growing antibiotic resistance in bacteria and emerging pandemics. AMP have lower rates of resistance mutations than small-molecule antibiotics (Huang *et al.*, 2010). The present SARS-COVID-19 pandemic has resulted in many peptide products and biologics undergoing screening and development, but these will require a rapid scale-up from µg to multiple tonnes quantities to meet global demand (Callaway, 2020; Zhang *et al.*, 2020; Al Musaimi *et al.*, 2021).

A powerful method of screening potential peptide pharmaceuticals is a phage-based library screen. An immense variety of peptide combinations can be investigated, with the ability to screen $1x10^{14}$ peptide structures representing all the possible amino acids combinations in a peptide 11 AA long (Cwirla *et al.*, 1990; Latham, 1999; Funke and Willbold, 2009). The variety is vital as two of the major problems with peptides pharmaceuticals are *in vivo* stability and delivery, which can be improved by using better peptides selected from a phage generated library based on desired characteristics. Peptides can be designed and selected for resistance to protease degradation, leading to increased half-life times that longer periods between dosages. The selection for peptides with increased target site affinities and better pharmacokinetics allows for lower peptide dosages for an equivalent effect. These are especially important concerning problems with peptide delivery which are heavily influenced by these characteristics. Peptides using unnatural modifications have been shown to have these characteristics and are used in nearly 25% of commercial peptide pharmaceuticals (Vlieghe *et al.*, 2010).

Of growing interest are the use of D-enantiomer amino acid peptides, as these are highly protease-resistant and have good absorption from oral administration (Funke and Willbold, 2009). The protease resistance is due to proteases being unable to process D-enantiomer amino acids, limiting the loss of peptides in digestion and increasing the peptide's *in vivo* half-life (Funke and Willbold, 2009). The CPS with D-enantiomer AA has proven difficult, encouraging

the exploration of biological synthesis methods that can be modified to translate genetic information to synthesis peptides with these unnatural AA (Funke and Willbold, 2009).

The phage peptides produced with phage screening are host generated indicates they are inherently producible from the host bacterial cell system. A biologically produced peptide allows for the rapid expansion in the production of the peptides of interest by transfer of the relevant genetic information into expression focused cell lines of that original screening host (Cwirla *et al.*, 1990; Funke and Willbold, 2009). These peptides can also incorporate novel peptides if the cell produces both unnatural AA and L-amino acids (Funke and Willbold, 2009).

Unnatural peptides are resistant to protease degradation, thereby increasing *in vivo* half-life and can allow for increased receptor binding (Funke and Willbold, 2009). Importantly, D-enantiomer amino acids are already incorporated into several commercial peptides. The production and research of recombinant peptides will need to match the ability to incorporate these unnatural AA to match SPPS (Funke and Willbold, 2009). Thus, a complete set of biological systems for screening and producing novel peptides can rapidly iterate products with short development cycles for recombinant peptide pharmaceutical production. This is enhanced by the fact that the cost of recombinant peptide development decreases as the cost of research decreases with lower unit costs of DNA sequencing and synthesis (Leproust, 2016).

Also, using PAM amidated peptides in conjunction with techniques such as non-ribosomal peptide synthases could offer novel classes of peptides that are being discovered, such as immunomodulatory peptide innate defence regulator (IDR)-1018 (de la Fuente-Núñez *et al.*, 2014). This functions to inhibit bacterial biofilm expression by degrading guanosine phosphate, which is used in a highly conserved bacterial signalling pathway. Naturally produced in humans, this peptide functions to inhibit and eliminate biofilms, which significantly increase bacterial antibiotic resistance and are a key component of chronic unresponsive bacterial infections (de la Fuente-Núñez *et al.*, 2014). These novel peptides are promising, and if made with cheaper recombinant production methods, they will help counter these emerging global threats and ongoing pandemics.

1.6.2 The current peptide drug market

For the previous 50 years, there has been a decrease in the number of novel pharmaceutical compounds developed per billion US dollars spent on research and development (Congressional Budget Office, 2006; Scannell *et al.*, 2012). Currently, it costs over $1.4 billion to bring a compound to market due to increased R&D and clinical costs, regulatory costs and

the need for the compound to be superior in many aspects to both its similar predecessor compounds and any similar compounds in the market (Scannell *et al.*, 2012; Puetz and Wurm, 2019). Compounding these problems is the lack of research and release of trial data, particularly why a specific compound failed in clinical trials (Scannell *et al.*, 2012). The further development and implementation of pharmaceutical peptides, in particular biologically synthesised peptides, could potentially reduce some of these problems.

Recombinantly produced insulin provides an example of established recombinant peptide production at global scales, with over 6 000 metric tonnes produced in 2006 and approximately 16 000 metric tonnes in 2012 (Gebel, 2013). This production represents global sales of $ 16.7 billion in 2011, with demand and supply growing annually. The global pharmaceutical peptide market was estimated at $ 8.5 billion in 2008, with an annual demand approaching 5 000 kg (Guzman *et al.*, 2007; Gebel, 2013). These figures have continued to grow and represent a major challenge in providing sufficient medical care for patients and a significant opportunity if supply can be expanded

The peptide pharmaceutical market was estimated at $ 25 billion in 2018, with a compounded annual growth rate of 8% basis (de la Torre and Albericio, 2019; D'Aloisio *et al.*, 2021). The market share is only set to grow for peptides due to their high success rate at twice that of conventional small molecule drugs (Craik *et al.*, 2013). An estimate for the R&D capital spent on peptides was $ 74 billion in 2018, with annual growth for the preceding decade (Puetz and Wurm, 2019).

There were over 60 pharmaceutical peptides commercially marketed in 2010, with over 100 in several phases of clinical development (Funke and Willbold, 2009; Zompra *et al.*, 2009; Thayer, 2011). Subsequently, over 100 peptide candidates are in the market, and multiple new peptide products receive FDA approval annually (de la Torre and Albericio, 2019; D'Aloisio *et al.*, 2021). The rate at which peptide drugs are entering clinical trials is increasing from one per year in the 1970s to over 16 per year in the 2000s, but the total receiving approval has stabilised on an annual basis in the 2010s (Thundimadathil, 2013; Al Musaimi *et al.*, 2018; Al Musaimi *et al.*, 2021).

In 2010 over 45 peptide drugs are alpha amidated peptides, making this the predominant form of PTM on the pharmaceutical peptide market, with this trend persisting (Vlieghe *et al.*, 2010; Paradís-Bas *et al.*, 2016; de la Torre and Albericio, 2020). The current lack of an inexpensive method to amidate the C-terminal of recombinant peptides has been one of the key barriers to

the entry of recombinant peptides and other biosimilar peptides to the market (Bray, 2003). The ability to amidate peptides would change this economic paradigm as 62 peptides, or 52% of the market could have been produced with recombinant methods and an amidation step (Guzman *et al.*, 2007; Usmani *et al.*, 2017). Moreover, most new peptides released to the market are longer than 20 AA, which may be more suited to recombinant synthesis than current SPPS (Figure 1.7) (Lau and Dunn, 2017).

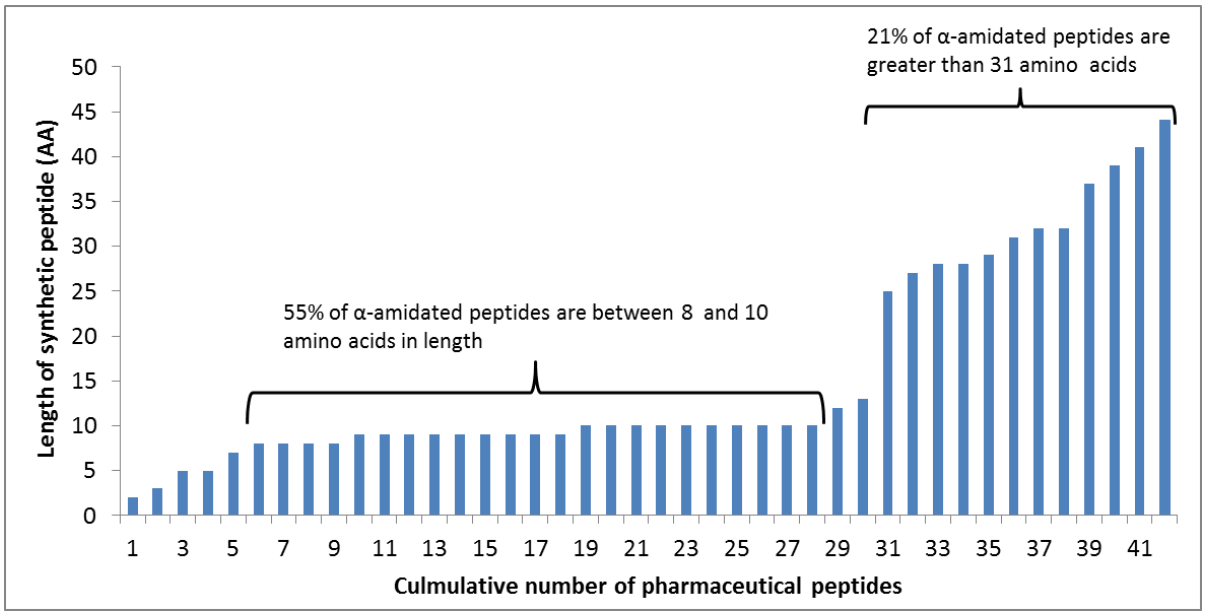


Figure 1.7: A bar graph comparing the cumulative number of commercially available pharmaceutical peptides to their amino acid lengths in 2016.

While the current majority of commercially available peptides are between 8 and 10 AA length, this is due to the historical advantages of CPS favouring these smaller peptides (Kolb *et al.*, 2001). This size limit and preferences meant that new peptides in larger classes were not preferred, and research into them was thereby limited (Kolb *et al.*, 2001; Vlieghe *et al.*, 2010; Glaser, 2013a; Li, Aneja and Chaiken, 2013). The longer classes of peptides could be ideal candidates to test the economic parameters and manufacturing competitiveness with the current CPS versus recombinant production with a biological alpha-amidation system.

The low cost and proofreading systems inherent to biological protein synthesis would result in greater fidelity for longer sequences, as with the addition of each AA, the number of steps in the increases CPS and the total loss of final peptide product increases (Guzman *et al.*, 2007). This is due to the natural error rate in CPS, and this increases the costs in terms of additional reagents, increased production time, lower throughput and expensive downstream activities such as purification and production identification and analyses of each step and the final

product (Guzman *et al.*, 2007). For recombinant biological peptide production, the increased length of a peptide can lead to decreased yield but is counteracted by low costs of media, ability to scale, with standardised methods and uniform production times. Additionally, the low error frequency with yeast protein synthesis and the host's endogenous proofreading systems, such as chaperons and ubiquitin, may reduce final product variability and lower purification and analysis costs (Kurland and Gallant, 1996; Ogle and Ramakrishnan, 2005).

Specifically, with the ability to amidate recombinant peptides, all of the top 5 peptide drugs by sales in 2010 (Glatiramer, Leuprolide, Goserelin, Octreotide, and Exenatide) could be produced recombinantly as opposed to current SPPS production. These top 5 sellers had total sales, in terms of sales revenue in 2010, of $ 9.2 billion (Thundimadathil, 2013). Of interest are all these top sellers patents expired in 2017, which would allow for the potential mass production of low-cost biosimilars without any license fees. The adoption of these cheaper substitutes and generics will expand the related market sizes as more patients could afford these treatments leading to higher demand and expanded production capacity. The sales of longer sequence peptide pharmaceuticals would increase, generating higher profits for recombinant producers of these generics as compared to their traditional CPS competitors.

The sale and use of these recombinantly produced generics has faced regulatory hurdles from both regulatory agencies and current predominant pharmaceutical suppliers in the market (Walsh, 2014). Regulatory authorities accept the use of recombinantly produced peptide therapeutics as it is the only means for producing large proteins (greater than 500 AA) (Vlieghe *et al.*, 2010). A major obstacle in establishing these biosimilars produced with biological amidation will be current producers that have major capital tied up in current conventional CPS methods and would be unwilling to accept increased competition from a novel technology (Puetz and Wurm, 2019). The incumbent pharmaceutical corporations can employ oligarchy tactics to prevent the adoption of recombinantly produced peptides through price-fixing collusion to lower prices to below-market dictated prices or political and legislative efforts to block new market entrants.

However, new technologies that lower costs will be adopted in the long run as companies will always pursue greater profits. The fastest way to bring recombinantly produced peptides with α-amidation technology to market with these conditions will probably be a new market entrant and should be encouraged for this new technology. Technologies for rapidly identifying and producing recombinant peptides exist and are highly suited to the biological production

methods, which would aid new market entrants overcome barriers of entry imposed by competitors utilising CPS.

1.6.5 Future trends

Given the shift in pharmaceutical research to long term diseases such as Alzheimer's, cancer, HIV and diabetes, the use of peptides to treat these conditions, with more potential products in trials, is auspicious (Funke and Willbold, 2009). A biological synthesis method for a peptide used in all clinical trial stages would also overcome the FDA's current biologically identical substitution prohibition. The substitution provision effectively prohibits the use of biologically synthesised peptides in place of identical CPS products.

There are steady and growing demands for longer, increasingly complex peptides at lower costs, especially pharmaceutical peptides (Glaser, 2013b; Fosgerau and Hoffmann, 2015). As previously injectable peptide products are repurposed, the need for greater production capacity for existing and novel peptides will increase. Treatments will be able to use oral, nasal and transdermal delivery systems but will require increased peptide quantities to compensate for lower absorption (Glaser, 2013b; Fosgerau and Hoffmann, 2015). The ability to produce and rapidly scale up peptide drug candidates from pre-clinical research, through trials and into the commercial market using a single synthesis method could satiate these demands (Glaser, 2013b).

1.7.1 Problem statement

The majority of currently α-amidated peptides for research and pharmaceutical use are produced by chemical peptide synthesis, which has high production costs, significant environmental negative externalities and limited peptide yields. Can a recombinant heterologously produced PAM, or PHM, be used to α-amidate biosynthetic peptides and offer an inexpensive and safe method of pharmaceutical peptide synthesis as compared to chemical peptide synthesis methods?

1.7.2 Hypothesis

It is possible to express a PAM, or PHM with Fungi, to α-amidate polypeptides. This could offer a viable, inexpensive and safe method of pharmaceutical peptide synthesis as compared to current chemical methods.

Biological pharmaceutical peptide synthesis could allow for environmentally safe local production of current and potentially novel and longer, α-amidated polypeptides in greater yields and at lower costs.

1.7.3 Aim

To express and purify PAM with *Yarrowia lipolytica* for characterisation and alpha-amidated peptide production.

1.7.4 Objectives

- To construct several sequences of PAM, PHM or PAL from phylogenetically distinct organisms.
- To express and purify PAM, PHM or PAL in *Yarrowia lipolytica*.
- To characterise any PAM, PHM or PAL produced to establish which would be ideal for industrial use.

1.7.5 Research Questions

- Is the production of α-amidated peptides through biosynthesis with recombinant PAM, PHM or PAL possible?
    - Is it economically feasible and/or better than with current chemical synthesis techniques?
- Can a new assay be adapted for the measurement of PAM, PHM or PAL activity?
    - Can this assay be incorporated as part of the production process of α-amidated peptides?
- Can the PAM, PHM or PAL enzyme be recombinantly expressed?
    - Is this possible with the yeast *Yarrowia lipolytica?*
    - What quantities can be produced?
    - What is an optimal method for the expression of PAM, PHM or PAL?
- What are the characteristics of the PAM, PHM or PAL produced?
    - Can the recombinant enzyme catalyse the α-amidation of glycine extended peptide substrate precursor?

1.8 References:

Abad, E., Rommel, J. B. and Kästner, J. (2014) 'Reaction mechanism of the bicopper enzyme peptidylglycine α-hydroxylating monooxygenase', *Journal of Biological Chemistry*, 289(20), pp. 13726–13738. doi: 10.1074/jbc.M114.558494.

Agyei, D. and Danquah, M. K. (2011) 'Industrial-scale manufacturing of pharmaceutical-grade bioactive peptides.', *Biotechnology Advances*, 29(3), pp. 272–7. doi: 10.1016/j.biotechadv.2011.01.001.

Alam, M., Steveson, T., Johnson, R., Bäck, N., Abraham, B., Mains, R. and Eipper B. (2001) 'Signaling mediated by the cytosolic domain of peptidylglycine α-amidating monooxygenase', *Molecular Biology of the Cell*. Edited by H. Riezman, 12(3), pp. 629–644. doi: 10.1091/mbc.12.3.629.

Attenborough, R. M. F., Hayward, D., Kitahara, M., Miller, D. and Ball, E. (2012) 'A "neural" enzyme in nonbilaterian animals and algae: Preneural origins for peptidylglycine α-amidating monooxygenase', *Molecular Biology and Evolution*, 29(10), pp. 3095–3109. doi: 10.1093/molbev/mss114.

Bankar, A. V., Kumar, A. R. and Zinjarde, S. S. (2009) 'Environmental and industrial applications of *Yarrowia lipolytica*', *Applied Microbiology and Biotechnology*, 84(5), pp. 847–865. doi: 10.1007/s00253-009-2156-8.

Barth, G. and Gaillardin, C. (1996) '*Yarrowia lipolytica*', in *Nonconventional Yeasts in Biotechnology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 313–388. doi: 10.1007/978-3-642-79856-6_10.

Bauman, A. T., Ralle, M. and Blackburn, N. J. (2007) 'Large scale production of the copper enzyme peptidylglycine monooxygenase using an automated bioreactor.', *Protein Expression and Purification*, 51(1), pp. 34–8. doi: 10.1016/j.pep.2006.06.016.

Bellou, S., Makrim A., Triantaphyllidou, I., Papanikolaou, S. and Aggelis, G. (2014) 'Morphological and metabolic shifts of *Yarrowia lipolytica* induced by alteration of the dissolved oxygen concentration in the growth environment', *Microbiology*, 160(Pt_4), pp. 807–817. doi: 10.1099/mic.0.074302-0.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J., Molina-Jouve, C. and Nicaud, J. (2009) '*Yarrowia lipolytica* as a model for bio-oil production.', *Progress in Lipid Research*, 48(6), pp. 375–87. doi: 10.1016/j.plipres.2009.08.005.

Bolkenius, F. N., Ganzhorn, A., Chanal, M., and Danzin, C., (1997) 'Selective mechanism-based inactivation of peptidylglycine α-hydroxylating monooxygenase in serum and heart atrium vs. brain', *Biochemical Pharmacology*, 53(11), pp. 1695–1702. doi: 10.1016/S0006-2952(97)00051-8.

Bradbury, A. F., Finnie, M. D. A. and Smyth, D. G. (1982) 'Mechanism of C-terminal amide formation by pituitary enzymes', *Nature*, 298(5875), pp. 686–688. doi: 10.1038/298686a0.

Bradbury, A. F. and Smyth, D. G. (1987) 'Enzyme-catalysed peptide amidation. Isolation of a stable intermediate formed by reaction of the amidating enzyme with an imino acid', *European Journal of Biochemistry*, 169(3), pp. 579–584. doi: 10.1111/j.1432-1033.1987.tb13648.x.

Bray, B. L. (2003) ‘Large-scale manufacture of peptide therapeutics by chemical synthesis.’, *Nature Reviews. Drug Discovery*, 2(7), pp. 587–93. doi: 10.1038/nrd1133.

Bulet, P., Stocklin, R. and Menin, L. (2004) ‘Anti-microbial peptides: From invertebrates to vertebrates’, *Immunological Reviews*, 198(1), pp. 169–184. doi: 10.1111/j.0105-2896.2004.0124.x.

Callaway, E. (2015) ‘The revolution will not be crystallized: A new method sweeps through structural biology’, *Nature*, 525(7568), pp. 172–174. doi: 10.1038/525172a.

Callaway, E. (2020) ‘The race for coronavirus vaccines’, *Nature*, 580, pp. 576–577. Available at: https://media.nature.com/original/magazine-assets/d41586-020-01221-y/d41586-020-01221-y.pdf.

Cao, F., Gamble, A., Kim, H., Onagi, H., Gresser, M., Kerr, J., and Easton, C. (2011) ‘Potent and selective inhibitors of human peptidylglycine α-amidating monooxygenase’, *MedChemComm*, 2(8), p. 760. doi: 10.1039/c1md00079a.

Chufán, E. E., De, M., Eipper, B., Mains, R., and Amzel, L. (2009) ‘Amidation of bioactive peptides: Ahe structure of the lyase domain of the amidating enzyme’, *Structure*, 17(7), pp. 965–973. doi: 10.1016/j.str.2009.05.008.Amidation.

Cipolatti, E. P., Silva, M., Klein, M., Feddern, V., Feltes, M., Oliveira, J., Ninow, J. and De Oliveira, D. (2014) ‘Current status and trends in enzymatic nanoimmobilization’, *Journal of Molecular Catalysis B: Enzymatic*, 99, pp. 56–67. doi: 10.1016/j.molcatb.2013.10.019.

Congressional Budget Office (2006) *Research and development in the pharmaceutical industry*. Available at: https://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/76xx/doc7615/10-02-drugr-d.pdf.

Craik, D. J., Fairlie, D., Liras, S., and Price, D. (2013) ‘The future of peptide-based drugs.’, *Chemical Biology & Drug Design*, 81(1), pp. 136–47. doi: 10.1111/cbdd.12055.

Cwirla, S. E., Peters, E., Barrett, R. and Dower, W. (1990) ‘Peptides on phage: A vast library of peptides for identifying ligands.’, *Proceedings of the National Academy of Sciences of the United States of America*, 87(16), pp. 6378–82.

D’Aloisio, V., Dognini, P., Hutcheon, G. and Coxon, C. (2021) ‘PepTherDia: Database and structural composition analysis of approved peptide therapeutics and diagnostics’, *Drug Discovery Today*, 26(6), pp. 1409–1419. doi: 10.1016/j.drudis.2021.02.019.

Danquah, M. and Agyei, D. (2012) ‘Pharmaceutical applications of bioactive peptides’, *OA Biotechnology*, 1(2), pp. 1–7. doi: 10.13172/2052-0069-1-2-294.

Edison, A. S., Espinoza, E. and Zachariah, C. (1999) ‘Conformational ensembles: The role of neuropeptide structures in receptor binding.’, *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 19(15), pp. 6318–26.

Eipper, B. A., Milgram, S., Husten, E., Yun, H. and Mains, R. (1993) ‘Peptidylglycine alpha-amidating monooxygenase: A multifunctional protein with catalytic, processing, and routing domains.’, *Protein Science : A Publication of the Protein Society*, 2(4), pp. 489–97. doi: 10.1002/pro.5560020401.

Eipper, B. A., Quon, A. S., Mains, R. E., Boswell, J. S. and Blackburn, N. J. (1995) ‘The catalytic core of peptidylglycine alpha-hydroxylating monooxygenase: Investigation by site-

directed mutagenesis, Cu X-ray absorption spectroscopy, and electron paramagnetic resonance.', *Biochemistry*, 34(9), pp. 2857–65.

Eipper, B. A. and Mains, R. E. (1988) 'Peptide α-amidation', *Annual Review of Physiology*, 50(1), pp. 333–344. doi: 10.1146/annurev.ph.50.030188.002001.

Eipper, B. A., Stoffers, D. A. and Mains, R. E. (1992) 'The biosynthesis of neuropeptides: Peptide alpha-amidation', *Annual Review of Neuroscience*, 15(1), pp. 57–85. doi: 10.1146/annurev.ne.15.030192.000421.

FDA (2009) *Guidance for industry on biosimilars: Q & As regarding implementation of the BPCI Act of 2009: questions and answers part II*. Food and Drug Administration. Available at: http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm271790.htm

FDA (2020) *Impact Story : Developing the Tools to Evaluate Complex Drug Products : Peptides*. Food and Drug Administration. Available at: https://www.fda.gov/drugs/regulatory-science-action/impact-story-developing-tools-evaluate-complex-drug-products-peptides.

Fosgerau, K. and Hoffmann, T. (2015) 'Peptide therapeutics: Current status and future directions', *Drug Discovery Today*, 20(1), pp. 122–128. doi: 10.1016/j.drudis.2014.10.003.

Francisco, W. A. Wille, G., Smith, A., Merkler, D. and Klinman, J. (2004) 'Investigation of the pathway for inter-copper electron transfer in peptidylglycine α-amidating monooxygenase', *Journal of the American Chemical Society*, 126(41), pp. 13168–13169. doi: 10.1021/ja046888z.

de la Fuente-Núñez, C., Reffuveille, F., Haney, E., Straus, S.and Hancock, R. (2014) 'Broad-spectrum anti-biofilm peptide that targets a cellular stress response', *PLoS Pathogens*. Edited by M. R. Parsek, 10(5), p. e1004152. doi: 10.1371/journal.ppat.1004152.

Funke, S. A. and Willbold, D. (2009) 'Mirror image phage display—A method to generate D-peptide ligands for use in diagnostic or therapeutical applications', *Molecular BioSystems*, 5(8), p. 783. doi: 10.1039/b904138a.

Gaier, E. D., Kleppinger, A., Ralle, M., Covault, J., Mains, R., Kenny, A. and Eipper B. (2014) 'Genetic determinants of amidating enzyme activity and its relationship with metal cofactors in human serum', *BMC Endocrine Disorders*, 14(1), p. 58. doi: 10.1186/1472-6823-14-58.

Garny, S., Verschoor, J., Gardiner, N. and Jordaan, J. (2014) 'Spectrophotometric activity microassay for pure and recombinant cytochrome P450-type nitric oxide reductase', *Analytical Biochemistry*, 447(1), pp. 23–29. doi: 10.1016/j.ab.2013.11.005.

Gebel, E. (2013) *Making Insulin: Diabetes Forecast®*. Available at: http://www.diabetesforecast.org/2013/jul/making-insulin.html

Glaser, V. (2013a) 'Reducing the cost of peptide synthesis', *Genetic Engineering and Biotechnology News*, 33(13), pp. 32–34. doi: 10.1089/gen.33.13.18.

Glaser, V. (2013b) 'Scaling up peptide drugs', *Genetic Engineering & Biotechnology News*, 33(7). Available at: http://www.genengnews.com/keywordsandtools/print/1/30749/.

Guzman, F., Barberis, S. and Illanes, A. (2007) 'Peptide synthesis: Chemical or enzymatic', *Electronic Journal of Biotechnology*, 10(2), pp. 0–0. doi: 10.2225/vol10-issue2-fulltext-13.

Handa, S., Spradling, T., Dempsey, D. and Merkler, D. (2012) 'Production of the catalytic core of human peptidylglycine α-hydroxylating monooxygenase (hPHMcc) in *Escherichia coli*', *Protein Expression and Purification*, 84(1), pp. 9–13. doi: 10.1016/j.pep.2012.04.012.

Hofmeyer, T., Bulani, S., Grzeschik, J., Krah, S., Glotzbach, B., Uth, C., Avrutina, O., Brecht, M., Göringer, H., Van Zyl, P. and Kolmar, H. (2014) 'Protein production in *Yarrowia lipolytica* via fusion to the secreted lipase Lip2p.', *Molecular Biotechnology*, 56(1), pp. 79–90. doi: 10.1007/s12033-013-9684-2.

Hojo, K., Hara, A., Kitai, H., Onishi, M., Ichikawa, H., Fukumori, Y. and Kawasaki, K. (2011) 'Development of a method for environmentally friendly chemical peptide synthesis in water using water-dispersible amino acid nanoparticles', *Chemistry Central Journal*, 5(1), p. 49. doi: 10.1186/1752-153X-5-49.

Hook, V., Yasothornsrikul, S., Greenbaum, D., Medzihradszky, K., Troutner, K., Toneff, T., Bundey, R., Logrinova, A., Reinheckel, T., Peters, C. and Bogyo, M. (2004) 'Cathepsin L and Arg/Lys aminopeptidase: A distinct prohormone processing pathway for the biosynthesis of peptide neurotransmitters and hormones', *Biological Chemistry*, 385(6), pp. 473–480. doi: 10.1515/BC.2004.055.

Huang, Y., Huang, J. and Chen, Y. (2010) 'Alpha-helical cationic antimicrobial peptides: relationships of structure and function.', *Protein & Cell*, 1(2), pp. 143–52. doi: 10.1007/s13238-010-0004-3.

International Federation of Pharmaceutical Manufacturers & Associations (2012) *The pharmaceutical industry and global health: facts & figures*. Available at: http://www.ifpma.org/fileadmin/content/Publication/2013/IFPMA_-_Facts_And_Figures_2012_LowResSinglePage.pdf.

Kahari, D., Koorsen, G., Stoychev, S., Chipeta, Z., Labuschagne, M., Van Zyl, P., Nthangeni, B. and Zyl, V. (2007) 'Glycoprofiling of N-linked glycans of erythropoietin therapeutic protein expressed in *Yarrowia lipolytica*', in *CSIR innovations conference*, p. 6665. CSIR. Available at http://researchspace.csir.co.za/dspace/bitstream/handle/10204/2866/Kahari_P_2008.pdf;sequence=1.

Kato, I., Yonekura, H., Yamamoto, H. and Okamoto, H. (1990) 'Isolation and functional expression of pituitary peptidylglycine alpha-amidating enzyme mRNA. A variant lacking the transmembrane domain.', *FEBS letters*, 269(2), pp. 319–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/2401356.

Katopodis, A., Ping, D., Smith, C. and May, S. (1991) 'Functional and structural characterization of peptidylamidoglycolate lyase, the enzyme catalyzing the second step in peptide amidation', *Biochemistry*, 30(25), pp. 6189–6194. doi: 10.1021/bi00239a016.

Keefer, L. M., Piron, M. A. and De Meyts, P. (1981) 'Human insulin prepared by recombinant DNA techniques and native human insulin interact identically with insulin receptors.', *Proceedings of the National Academy of Sciences*, 78(3), pp. 1391–1395. doi: 10.1073/pnas.78.3.1391.

Khoury, G. A., Baliban, R. C. and Floudas, C. A. (2011) 'Proteome-wide post-translational modification statistics: Frequency analysis and curation of the swiss-prot database', *Scientific Reports*, 1(1), p. 90. doi: 10.1038/srep00090.

Kim, D. S. and Hahn, Y. (2012) 'Human-specific protein isoforms produced by novel splice sites in the human genome after the human-chimpanzee divergence', *BMC Bioinformatics*,

13(1), p. 299. doi: 10.1186/1471-2105-13-299.

Kim, J. Y., Kim, Y.-G. and Lee, G. M. (2012) 'CHO cells in biotechnology for production of recombinant proteins: Current state and further potential', *Applied Microbiology and Biotechnology*, 93(3), pp. 917–930. doi: 10.1007/s00253-011-3758-5.

Kim, K. H. and Seong, B. L. (2001) 'Peptide amidation: Production of peptide hormones *in vivo* and *in vitro*', *Biotechnology and Bioprocess Engineering*, 6(4), pp. 244–251. doi: 10.1007/BF02931985.

Kolb, H. C., Finn, M. G. and Sharpless, K. B. (2001) 'Click chemistry: Diverse chemical function from a few good reactions', *Angewandte Chemie International Edition*, 40(11), pp. 2004–2021. doi: 10.1002/1521-3773(20010601)40:11<2004::AID-ANIE2004>3.0.CO;2-5.

Kolhekar, A. S., Keutmann, H., Mains, R., Quon, A. and Eipper, B. (1997) 'Peptidylglycine α-hydroxylating monooxygenase: Active site residues, disulfide linkages, and a two-domain model of the catalytic core', *Biochemistry*, 36(36), pp. 10901–10909. doi: 10.1021/bi9708747.

Kolhekar, A. S. Bell, J., Shiozaki, E., Jin, L., Keutmann, H., Hand, T., Mains, R. and Eipper, B. (2002) 'Essential features of the catalytic core of peptidyl-α-hydroxyglycine α-amidating lyase', *Biochemistry*, 41(41), pp. 12384–12394. doi: 10.1021/bi0260280.

Krishnamurthy, V. M. Kaufman, G., Urbach, A., Gitlin, I., Gudiksen, K., Weibel, D. and Whitesides G. (2008) 'Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding', *Chemical Reviews*, 108(3), pp. 946–1051. doi: 10.1021/cr050262p.

Kumar, D., Blaby-Haas, C., Merchant, S., Mains, R., King, S. and Eipper, B. (2016) 'Early eukaryotic origins for cilia-associated bioactive peptide-amidating activity.', *Journal of Cell Science*, 129(5), pp. 943–56. doi: 10.1242/jcs.177410.

Kumar, D., Mains, R., Eipper, B. and King, S. (2019) 'Ciliary and cytoskeletal functions of an ancient monooxygenase essential for bioactive amidated peptide synthesis', *Cellular and Molecular Life Sciences*, 76(12), pp. 2329–2348. doi: 10.1007/s00018-019-03065-w.

Kumar, D., Mains, R. and Eipper, B. (2016) '60 Years of POMC: From POMC and α-MSH to PAM, molecular oxygen, copper, and vitamin C', *Journal of Molecular Endocrinology*, 56(4), pp. T63–T76. doi: 10.1530/JME-15-0266.

Kurland, C. and Gallant, J. (1996) 'Errors of heterologous protein expression', *Current Opinion in Biotechnology*, 7(5), pp. 489–493. doi: 10.1016/S0958-1669(96)80050-4.

Latham, P. (1999) 'Therapeutic peptides revisited', *Nature biotechnology*, 17(August), pp. 755–757. Available at: http://www.nature.com/nbt/journal/v17/n8/pdf/nbt0899_755.pdf

Lau, J. L. and Dunn, M. K. (2017) 'Supplementary information. Therapeutic peptides: Historical perspectives, current development trends, and future directions', *Bioorganic & Medicinal Chemistry*. Available at: http://www.nature.com/doifinder/10.1038/nn.4505.

Lau, J. L. and Dunn, M. K. (2018) 'Therapeutic peptides: Historical perspectives, current development trends, and future directions', *Bioorganic & Medicinal Chemistry*, 26(10), pp. 2700–2707. doi: 10.1016/j.bmc.2017.06.052.

Leproust, E. (2016) 'Rewriting DNA synthesis', *Chemical Engineering Progress*, September,

pp. 30–35. Available at: https://www.aiche.org/resources/publications/cep/2016/september/sbe-supplement-synthetic-biology-rewriting-dna-synthesis.

Li, C., Oldham, C. and May, S. (1994) 'NN-Dimethyl-1, 4-phenylenediamine as an alternative reductant for peptidylglycine alpha-amidating mono-oxygenase catalysis.', *Biochem. J*, 36(300), pp. 31–36. Available at: http://www.biochemj.org/bj/300/bj3000031.htm

Li, H., Aneja, R. and Chaiken, I. (2013) 'Click chemistry in peptide-based drug design', *Molecules*, 18(8), pp. 9797–9817. doi: 10.3390/molecules18089797.

Liu, H.-H., Ji, X. and Huang, H. (2015) 'Biotechnological applications of *Yarrowia lipolytica*: Past, present and future', *Biotechnology Advances*, 33(8), pp. 1522–1546. doi: 10.1016/j.biotechadv.2015.07.010.

Madzak, C., Gaillardin, C. and Beckerich, J.-M. (2004) 'Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: A review', *Journal of Biotechnology*, 109(1), pp. 63–81. Available at: http://www.sciencedirect.com/science/article/pii/S0168165604000264

Martin, V., Egelund, P., Johansson, H., Thordal Le Quement, S., Wojcik, F. and Sejer Pedersen, D. (2020) 'Greening the synthesis of peptide therapeutics: An industrial perspective', *RSC Advances*, 10(69), pp. 42457–42492. doi: 10.1039/D0RA07204D.

McCamish, M. and Woollett, G. (2011) 'Worldwide experience with biosimilar development', *mAbs*, 3(2), pp. 209–217. doi: 10.4161/mabs.3.2.15005.

Merkler, D. J. (1994) 'C-Terminal amidated peptides: Production by the *in vitro* enzymatic amidation of glycine-extended peptides and the importance of the amide to bioactivity', *Enzyme and Microbial Technology*, 16(6), pp. 450–456. doi: 10.1016/0141-0229(94)90014-0.

Merrifield, R. B. (1963) 'Solid phase peptide synthesis. I. The synthesis of a tetrapeptide', *Journal of the American Chemical Society*, 85(14), pp. 2149–2154. doi: 10.1021/ja00897a025.

El Meskini, R., Culotta, V., Mains, R. and Eipper, B. (2003) 'Supplying copper to the cuproenzyme peptidylglycine α-amidating monooxygenase', *Journal of Biological Chemistry*, 278(14), pp. 12278–12284. doi: 10.1074/jbc.M211413200.

Migneault, I., Dartiguenave, C., Vinh, J., Bertrand, M. and Waldron, K. (2004) 'Comparison of two glutaraldehyde immobilization techniques for solid-phase tryptic peptide mapping of human hemoglobin by capillary zone electrophoresis and mass spectrometry', *Electrophoresis*, 25(9), pp. 1367–1378. doi: 10.1002/elps.200305861.

Moore, A. B. and May, S. W. (1999) 'Kinetic and inhibition studies on substrate channelling in the bifunctional enzyme catalysing C-terminal amidation.', *The Biochemical journal*, 341 (Pt 1(341), pp. 33–40. Available at: http://www.biochemj.org/bj/341/bj3410033.htm

Murthy, A., Mains, R. and Eipper, B. (1986) 'Purification and characterization of peptidylglycine alpha-amidating monooxygenase from bovine neurointermediate pituitary.', *Journal of Biological Chemistry*, 261(4), pp. 1815–1822. Available at: http://www.jbc.org/content/261/4/1815.abstract

Al Musaimi, O., Al Shaer, D., de la Torre, B. and Albericio, F. (2018) '2017 FDA Peptide harvest', *Pharmaceuticals*, 11(2), p. 42. doi: 10.3390/ph11020042.

Al Musaimi, O., Al Shaer, D., Albericio, F. and de la Torre, B. (2021) '2020 FDA TIDES (peptides and oligonucleotides) harvest', *Pharmaceuticals*, 14(2), pp. 1–14. doi: 10.3390/ph14020145.

Nicaud, J., Madzak, C., van den Broek, P., Gysler, C., Duboc, P., Niederberger, P. and Gaillardin C. (2002) 'Protein expression and secretion in the yeast *Yarrowia lipolytica*.', *FEMS Yeast Research*, 2(3), pp. 371–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12702287.

Ogle, J. M. and Ramakrishnan, V. (2005) 'Structural insights into translational fidelity', *Annual Review of Biochemistry*, 74(1), pp. 129–177. doi: 10.1146/annurev.biochem.74.061903.155440.

Paradís-Bas, M., Tulla-Puche, J. and Albericio, F. (2016) 'The road to the synthesis of "difficult peptides"', *Chemical Society Reviews*, 45(3), pp. 631–654. doi: 10.1039/C5CS00680E.

Porro, D., Gasser, B., Fossati, T., Maurer, M., Branduardi, P., Sauer, M. and Mattanovich, D. (2011) 'Production of recombinant proteins and metabolites in yeasts: when are these systems better than bacterial production systems?', *Applied Microbiology and Biotechnology*, 89(4), pp. 939–48. doi: 10.1007/s00253-010-3019-z.

Prigge, S. T. (1997) 'Amidation of bioactive peptides: The structure of peptidylglycine α-hydroxylating monooxygenase', *Science*, 278(5341), pp. 1300–1305. doi: 10.1126/science.278.5341.1300.

Prigge, S. T. Mains, R., Eipper, B. and Amzel, L. (2000) 'New insights into copper monooxygenases and peptide amidation: structure, mechanism and function', *Cellular and Molecular Life Sciences*, 57(8), pp. 1236–1259. doi: 10.1007/PL00000763.

Prohaska, J. R. (2014) 'Impact of copper deficiency in humans.', *Annals of the New York Academy of Sciences*, 1314, pp. 1–5. doi: 10.1111/nyas.12354.

Puetz, J. and Wurm, F. M. (2019) 'Recombinant proteins for industrial versus pharmaceutical purposes: A review of process and pricing', *Processes*, 7(8), p. 476. doi: 10.3390/pr7080476.

Reichart, J. (2010) *Development trends for peptide therapeutics*. Available at: http://www.peptidetherapeutics.org/PTF_report_summary_2010.pdf.

Roth, R., Moodley, V. and van Zyl, P. (2009) 'Heterologous expression and optimized production of an *Aspergillus aculeatus* endo-1,4-beta-mannanase in *Yarrowia lipolytica*.', *Molecular Biotechnology*, pp. 112–20. doi: 10.1007/s12033-009-9187-3.

Rudzka, K., Moreno, D., Eipper, B., Mains, R., Estrin, D. and Amzel, L. (2013) 'Coordination of peroxide to the Cu(M) center of peptidylglycine α-hydroxylating monooxygenase (PHM): structural and computational study.', *Journal of Biological Inorganic Chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry*, 18(2), pp. 223–32. doi: 10.1007/s00775-012-0967-z.

Scannell, J. W., Blanckley, A., Boldon, H. and Warrington, B. (2012) 'Diagnosing the decline in pharmaceutical R&D efficiency.', *Nature reviews. Drug discovery*, 11(3), pp. 191–200. doi: 10.1038/nrd3681.

Schrader, M., Schulz-Knappe, P. and Fricker, L. D. (2014) 'Historical perspective of peptidomics', *EuPA Open Proteomics*, 3, pp. 171–182. doi: 10.1016/j.euprot.2014.02.014.

Sheldon, R. A. (2007) 'The E-factor: Fifteen years on', *Green Chemistry*, 9(12), pp. 1273–1283. doi: 10.1039/b713736m.

Siebert, X. Eipper, B., Mains, R., Prigge, S., Blackburn, N. and Amzel, M. (2005) 'The catalytic copper of peptidylglycine α-hydroxylating monooxygenase also plays a critical structural role', *Biophysical Journal*, 89(5), pp. 3312–3319. doi: 10.1529/biophysj.105.066100.

Simate, G., Cluetta J., Iyuke, S., Musapatika, E., Ndlovua, S., Walubita L. and Alvarez A. (2011) 'The treatment of brewery wastewater for reuse: State of the art', *Desalination*, 273(2–3), pp. 235–247. doi: 10.1016/j.desal.2011.02.035.

Stevenson, C. L. (2009) 'Advances in peptide pharmaceuticals.', *Current Pharmaceutical Biotechnology*, 10(1), pp. 122–137. doi: 10.2174/138920109787048634.

Stoffers, D., Green, B. and Eipper, B. (1989) 'Alternative mRNA splicing generates multiple forms of peptidyl-glycine α-amidating monooxygenase in rat atrium', *Proceedings of the National Academy of Sciences of the United States of America*, 86(2), pp. 735–739. doi: 10.1073/pnas.86.2.735.

Tajima, M., Iida, T., Yoshida, S., Komatsu, K., Namba, R., Yanagi, M., Noguchi, M. and Okamoto H. (1990) 'The reaction product of peptidylglycine alpha-amidating enzyme is a hydroxyl derivative at alpha-carbon of the carboxyl-terminal glycine.', *Journal of Biological Chemistry*, 265(17), pp. 9602–9605. doi: 10.1016/S0021-9258(19)38709-5.

Thayer, A. M. (2011) 'Improving peptides', *Chemical & Engineering News Archive*, 89(22), pp. 13–20. doi: 10.1021/cen-v089n022.p013.

Thundimadathil, J. (2013) 'Challenges in the large scale production of peptides', *Speciality Chemicals Magazine*, pp. 26–28. Available at: http://www.cpcscientific.com/pdfs/spe-chemicals-magazine-article.pdf.

Toney, J., Hammond, G., Fitzgerald, P., Sharma, N., Balkovec, J., Rouen, G., Olson, S., Hammond, M., Greenlee, M. and Gao, Y. (2001) 'Succinic acids as potent inhibitors of plasmid-borne IMP-1 metallo-β-lactamase', *Journal of Biological Chemistry*, 276(34), pp. 31913–31918. doi: 10.1074/jbc.M104742200.

de la Torre, B. and Albericio, F. (2019) 'The pharmaceutical industry in 2018. An analysis of FDA drug approvals from the perspective of molecules', *Molecules*, 24(4), p. 809. doi: 10.3390/molecules24040809.

de la Torre, B. and Albericio, F. (2020) 'Peptide therapeutics 2.0', *Molecules*, 25(10), pp. 2019–2021. doi: 10.3390/molecules25102293.

Uhlig, T., Kyprianou , T., Martinelli, F., Oppici, C., Heiligers, D., Hills, D., Calvo, X. and Verhaert, P. (2014) 'The emergence of peptides in the pharmaceutical business: From exploration to exploitation', *EuPA Open Proteomics*, 4, pp. 58–69. doi: 10.1016/j.euprot.2014.05.003.

Ul-Hasan, S., Burgess, D., Gajewiak, J., Li, Q., Hu, H., Yandell, M., Olivera, B. and Bandyopadhyay, P. (2013) 'Characterization of the peptidylglycine α-amidating monooxygenase (PAM) from the venom ducts of neogastropods, *Conus bullatus* and *Conus geographus*.', *Toxicon : official journal of the International Society on Toxinology*, 74, pp.

215–24. doi: 10.1016/j.toxicon.2013.08.054.

Usmani, S. S., Bedi, G., Samuel, J., Singh, S., Kalra, S., Kumar, P., Ahuja, A., Sharma, M., Gautam, A. and Raghava, G. (2017) 'THPdb: Database of FDA-approved peptide and protein therapeutics', *PLoS ONE*, 12(7), pp. 1–12. doi: 10.1371/journal.pone.0181748.

Vigneaud, V., Ressler, C., Swan, C., Roberts, C., Katsoyannis, P. and Gordon, S. (1953) 'The synthesis of an octapeptide amide with the hormonal activity of oxytocin', *Journal of the American Chemical Society*, 75(19), pp. 4879–4880. doi: 10.1021/ja01115a553.

Vishwanatha, K., Bäck, N., Mains, R. and Eipper, B. (2014) 'A histidine-rich linker region in peptidylglycine alpha-amidating monooxygenase has the properties of a pH sensor', *Journal of Biological Chemistry*, 289(18), pp. 12404–12420. doi: 10.1074/jbc.M113.545947.

Vlieghe, P., Lisowski, V., Martinez, J. and Khrestchatisky, M. (2010) 'Synthetic therapeutic peptides: Science and market.', *Drug Discovery Today*, 15(1–2), pp. 40–56. doi: 10.1016/j.drudis.2009.10.009.

Walsh, G. (2014) 'Biopharmaceutical benchmarks 2014', *Nature Biotechnology*, 32(10), pp. 992–1000. doi: 10.1038/nbt.3040.

Wand, G., Ney, R., Baylin, S., Eipper, B. and Mains, R. (1985) 'Characterization of a peptide alpha-amidation activity in human plasma and tissues', *Metabolism*, 34(11), pp. 1044–1052. Available at: http://www.sciencedirect.com/science/article/pii/0026049585900770 (Accessed: 10 December 2013).

Wang, J. and Chow, S. (2017) 'On the regulatory approval pathway of biosimilar products', pp. 353–368. doi: 10.3390/ph5040353.

Wang, Z., Wan, L., Liu, Z., Huang, X. and Xu, Z. (2009) 'Enzyme immobilization on electrospun polymer nanofibers: An overview', *Journal of Molecular Catalysis B: Enzymatic*, 56(4), pp. 189–195. doi: 10.1016/j.molcatb.2008.05.005.

Wapinski, I., Pfeffer, A., Friedman, N. and Regev, A. (2007) 'Natural history and evolutionary principles of gene duplication in fungi.', *Nature*, 449(7158), pp. 54–61. doi: 10.1038/nature06107.

Wimalasena, K. and Wimalasena, D. (1991) 'Continuous spectrophotometric assays for dopamine β-monooxygenase based on two novel electron donors: N,N-dimethyl-1,4-phenylenediamine and 2-aminoascorbic acid', *Analytical Biochemistry*, 197(2), pp. 353–361. doi: 10.1016/0003-2697(91)90404-H.

Zhang, G., Pomplun, S., Loftis, A., Loas, A. and Pentelute, B. (2020) 'The first-in-class peptide binder to the SARS-CoV-2 spike protein, *bioRxiv*, p. 2020.03.19.999318.

Zhou, X., Mikhailopulo, I., Cruz Bournazou M. and Neubauer, P., (2015) 'Immobilization of thermostable nucleoside phosphorylases on MagReSyn® epoxide microspheres and their application for the synthesis of 2,6-dihalogenated purine nucleosides', *Journal of Molecular Catalysis B: Enzymatic*, 115, pp. 119–127. doi: 10.1016/j.molcatb.2015.02.009.

Zompra, A., Galanis, A., Werbitzky, O. and Albericio, F. (2009) 'Manufacturing peptides as active pharmaceutical ingredients.', *Future medicinal chemistry*, 1(2), pp. 361–77. doi: 10.4155/fmc.09.23.

Chapter 2: Assay development for measuring peptidylglycine alpha-amidating monooxygenase products

2.1 Introduction

A critical problem inherent to the study of PAM, PHM and PAL enzymes has been the development of sensitive and reproducible enzyme assays. An incredible diversity of peptide products has been observed, including most bioactive peptides in mammalian endocrine and nervous systems, which are amidated at the C-terminus (Tatemoto and Mutt, 1978; Bradbury and Smyth, 1987). The initial studies of these peptides and their biogenesis identified glycine extended precursors to the α-amide terminating peptide sequences (Bradbury *et al.*, 1982). The precursors implied an enzyme catalysed the post-translational modification with any of the 20 standard amino acids at the penultimate position relative to the C-terminal glycine. The development of a PAM assay was further compounded by PAM's low turnover rate, which required both specialised and sensitive methods to detect PAM activity.

PAM's initial discovery and characterisations were only made possible with HPLC separation of products and increased assay sensitivity enabled due to radioactive substrates. HPLC, used in conjunction with radiolabelled substrates, has remained the standard technique for measuring PAM, PHM, and PAL activity, despite the drawbacks of multiple steps, longer run times, hazardous materials, and solvents (Mueller and Driscoll, 2002; Yin *et al.*, 2011). Although ongoing research for simpler assays without expensive equipment or the use of hazardous substances has yielded many assays, these have not been widely adopted (Carpenter and Merkler, 2003; Merkler *et al.*, 2008).

2.1.1 Current PAM activity detection assays

With the development of mass spectroscopy (MS) methods and equipment, radiolabelled substrates have become unpopular due to the disposal of radioactive waste and safety considerations (An *et al.*, 2012). The sensitivity and enzyme-substrate specificity of radio-isotope peptides make it a helpful method, but only for appropriately equipped laboratories (Mueller and Driscoll, 2002). Conversely, while HPLC techniques are more popular, they suffer from low throughput and require expensive equipment, sample preparation, and trained personnel (Kolhekar *et al.*, 1997; Mueller and Driscoll, 2002; An *et al.*, 2012). Increased throughput and accessibility, without specialised HPLC equipment, has been made possible

with fluorescence and absorbance assays for enzymes with similar activity to PAM (Carpenter and Merkler, 2003; Carpenter, 2006).

A promising enzyme-coupled absorbance assay based on measuring glyoxylate with the reduction of nicotinamide adenine dinucleotide was developed by Carpenter and Merkler (2003). Unfortunately, the essential enzymes for this assay were no longer commercially available for this study, and the patents related to the methods had seemingly curtailed its adoption (Alwan *et al.*, 2019). Enzyme-coupled assays can impose additional constraints on the original enzyme's assay composition and conditions or subsequent assays due to the coupling enzymes characteristics. These potential complications can compound any problems inherent to an enzyme and its assay.

2.1.2 Hippuric acid as a peptide substrate for PHM activity assays

PAM has documented activity with C-terminal glycine extended peptides, irrespective of the penultimate amino acid (Brock *et al.*, 1989; Shimoi *et al.*, 1992). Consequently, many peptide analogues can act as peptide substrates for PHM and PAM. Hippuric acid (HA) is a standard PAM substrate due to its simple single amino acid analogue structure, conferring ease of handling, low cost, and widespread availability (Katopodis and May, 1990; Chew, 2003; Merkler *et al.*, 2008). The reaction catalysed by PHM with HA is represented in Figure 2.1.

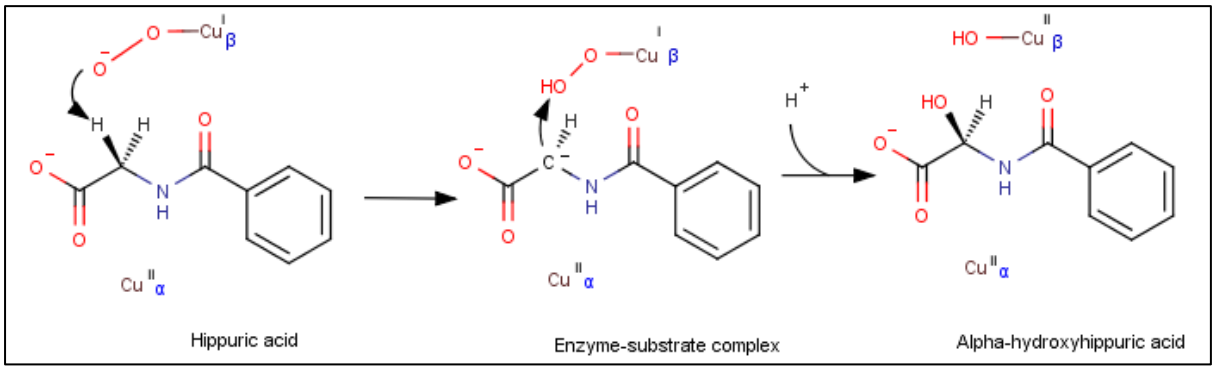


Figure 2.1: Reaction of peptidylglycine alpha-hydroxylating monooxygenase bound copper ions with a hippuric acid substrate.

An interesting study for hippuric acid measurement was a diagnostic test of urine for toluene abuse (Yoshida *et al.*, 2005; Zinalibdin and Yacob, 2013). The study further developed a previous derivatization reaction for HA by combining pyridine, hippuric acid, benzenesulphonyl chloride (PHAB) to form a red chromophore (Ogata and Sugihara, 1977; Zinalibdin and Yacob, 2013). The original assay was an absorbance measurement of HA concentrations in the millimolar range in urine (Tomokuni and Ogata, 1972). The methods

could theoretically be adapted to quantify PAM activity by monitoring a decrease in HA concentrations. Measuring a single peptide substrate's disappearance is a limiting drawback, with an assay monitoring the creation of a product, such as glyoxylate, providing a broader measure of PAM activity.

2.1.3 2-Aminobenzaldehyde, glyoxylate, glycine assay

Initial studies into the activity and enzyme active site reaction demonstrated that it was possible to dealkylate alpha-hydroxylated hippuric acid (αHA) under alkaline conditions (pH > 7.5) (Katopodis *et al.*, 1991). Notably, both PAM and PAL are located in the secretory granules, with a pH of 5.5 to 6.0 (Gainer *et al.*, 1985). Consequently, the non-enzymatic hydrolysis rate of C-terminal α-hydroxyglycine peptides is negligible *in vivo* for many organisms without PAL (Katopodis *et al.*, 1991). To catalyse the hydrolysis of glyoxylate with PAL's hydrolysis reaction, it is essential to increase this rate to biologically relevant levels. The singular expression of PHM without PAL implies that certain non-biological conditions, such as an alkaline pH, catalyse hydrolysis producing an amidated peptide. The hydrolysis reaction will occur spontaneously, at pH > 7.5 and is catalysed by the nucleophilic attack of free hydroxyl radicals to the hydroxylated amide (Figure 2.2).

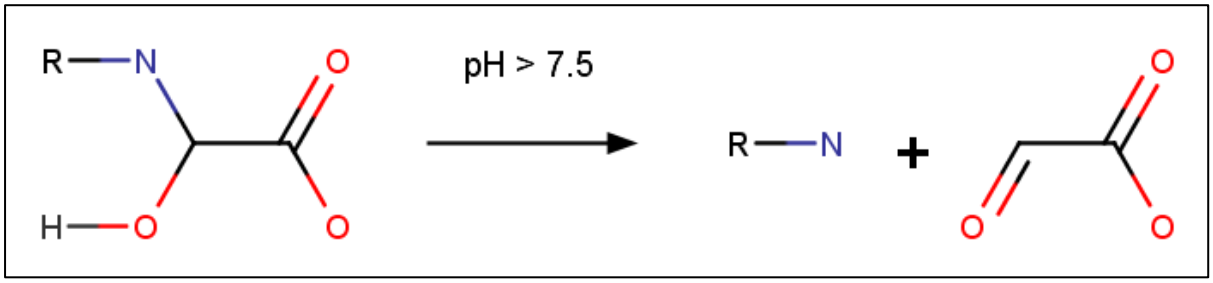


Figure 2.2: Alkaline pH-induced hydrolysis of the glyoxylate residue from the C-terminal's alpha-hydroxylated glycine residue (R= Rest of protein).

The use of alkaline conditions to catalyse the hydrolysis reactions renders PAL activity non-essential in producing the amidated peptides *in vitro* (Tajima *et al.*, 1990). An alkaline incubation step could allow for the potential of assays of PHM activity by circumventing PAL activity using pH hydrolysis and monitoring the resultant glyoxylate concentrations. Also, the sequential activity of PHM then PAL, which can instead be replicated by pH-induced hydrolysis of hydroxylated peptides, has also had the effect of obscuring the pH optima of PAM and PHM isoforms (Takahashi *et al.*, 1990). A suitable glyoxylate-based assay could be used with PAM and appropriate peptide substrates to determine its pH optima and the PHM and PAL subunits.

2.1.4 Aim

To establish a medium-throughput, non-hazardous absorbance assay to measure PAM activity.

2.1.5 Objectives

To develop an absorbance based assay for PAM, PHM and PAL activity.

To validate the absorbance assay developed and to measure PAM products with UPLC-MS.

2.2 Material and Methods

2.2.1 Pyridine, hippuric acid, benzenesulphonyl chloride concentration assay

A modified assay adapted from Zinalibdin and Yacob (2013) was investigated to measure PAM activity. Initial studies with the polystyrene 96-well plates had the plate's plastic react with the pyridine (Merck; 270970-2L) and benzenesulphonyl chloride (Merck; 108138-5G), fouling the plate and generating precipitate. Organic solvent resistant plates (Star labs, PlateOne®; S1837-9600) were used to prevent plate fouling and the deposition formed with the PHAB reaction. Standard curves of 0.1 to 2 mM HA and 2 to 7 mM HA (Sigma-Aldrich; 112003-100G) with the PHAB assay were performed. The resulting solutions had their absorbance at 430 nm measured (Zinalibdin and Yacob, 2013). Problems relating to the assay required developing a less hazardous and more sensitive assay for glyoxylate, which was investigated with a 2-aminobenzaldehyde and glycine assay.

2.2.2 2-Aminobenzaldehyde glycine glyoxylate concentration assay

The methods for determining glyoxylate concentration in solution were developed from an assay of 2-aminobenzaldehyde (2AB) (Merck; A9628-1G), glycine (Merck; G8898-1KG), glyoxylate (Merck; G4502-1G), as described in Soda *et al.*, (1973). The original paper's volumes were adapted for the 96 well plate format to a reduced final reaction volume of 350 µL. The glyoxylate containing sample solution's relative volume increased to 250 µL with the reagent volumes reduced to maintain the final reaction volume. The 2AB stock concentration was increased to 75 mM as the volume of the 2AB stock solution was reduced to 50 µL added to the 2AB, glycine, glyoxylate (AGG) reaction solution. The final concentration of 10.7 mM 2AB was twice the minimum 5 mM 2AB required for maximal absorbance development, at 440 nm from Soda *et al.*, 1973. The glycine stock solution was doubled to 2 M, while the volume added was halved to 50 µL. The reduction produced a slightly lower final concentration of 0.286 M glycine than the original paper at 0.313 M (Soda *et al.*, 1973). The decreased glycine concentration was higher than the minimum 0.156 M recommend in the original article (Soda *et al.*, 1973).

A significant alteration to the methodology was dividing the glycine stock into alkaline and acidic stocks of 2 M glycine. When the alkaline was added to a suitable hydroxylated substrate, the pH would catalyse glyoxylate hydrolysis (Katopodis *et al.*, 1991). As glycine had two dielectric points in appropriate ranges, it was possible to use a single reagent; thereby simplifying the reaction. Two 2 M glycine stock solutions, at pH 2.5 and pH 10.0, were used

instead of a single 2 M glycine pH 6.0 solution (Soda *et al.*, 1973). At the assay reaction termination, 16.7 µL of the alkaline 2 M glycine, pH 10.0, was added, and the multi-well plate was incubated at 37°C for 40 minutes, for sufficient time for pH-induced dealkylation. Subsequently, 33.3 µL of 2 M glycine, pH 2.5 was added to neutralise the solution, to ~pH 6.50. With the addition of 50 µL of 50 mM 2 AB, the plate was incubated at 37°C for 15 minutes with absorbance measured at the 450 nm peak, as opposed to the 440 nm in the original paper (Soda *et al.*, 1973). These volumes were experimentally determined.

Significant precipitation and solubility issues were encountered with the 2AB stock solution using only H_2O as a solvent. Instead, 0.3 mg 2AB was dissolved in 50 µL of 99.9 % ethanol (5 M 2AB), then diluted to 25 mM with H_2O. The 2AB will polymerise at room temperature and was stored at -20°C in powder form with 2AB stock solutions prepared immediately before assays. As 2AB can be a pulmonary irritant, the samples were kept sealed and solutions handled in a fume hood. The 2 M glycine pH 10.0 solution similarly had precipitation and degradation issues when stored for prolonged periods at RT and was made fresh weekly.

2.2.3 Liquid chromatography and mass spectroscopy methods for separation and analysis of AGG and PAM activity solutions

Ultra performance liquid chromatography (UPLC) was used on specific samples to determine PAM activity levels and validate the AGG assay. UPLC separations were performed with a Waters UPLC, utilising a Water HSS T3 C18 column, coupled in series to a Water SYNAPT G1 HDMS mass spectrometer (MS). Professor Paul Steenkamp kindly performed the chromatographic separation procedures, optimisation of techniques, and his help was invaluable for the analysis of chromatographic results with the UPLC-MS. Column dimensions were 150 mm x 2.1 mm, with 1.8 µm particles, and sample separation was performed with a constant column temperature of 60°C. Samples were prepared with centrifugation at 20 000 *g* for 15 minutes, with supernatant used for analysis. Samples were frozen at -20°C, then thawed and kept at 6°C in the Waters sample manager for analysis.

A binary solvent system of eluent A (10 mM formic acid (Merck, 533002) in H_2O, pH 2.5) and eluent B (10 mM formic acid in acetonitrile (Merck, 100029) were used for analysis. The total run time per sample was 10 minutes, with 1 µL of sample injection volumes. After the sample injection, initial running conditions were a flow rate of 98% A at 0.4 mL/min for 1 minute, with a subsequent linear-gradient to 5% B by 6 min. After 1 min of 5% B, the solvent was changed to the initial conditions.

The analytes' identity produced after chromatographic separation was obtained with an SYNAPT G1 mass spectrometer used in V-optics. The mass spectrometer (MS) was operated in electrospray mode to detect all electrospray ionisation (ESI) compatible compounds. A reference calibrant (lock mass) of leucine enkephalin (50 pg/mL) (Waters; 186006013) was used to obtain mass accuracies of 1.0-5.0 milliDalton. The MS was operated in both ESI positive and negative modes to increase coverage of the ions produced. The capillary voltage was set to 2.5 kV, with the sampling cone at 30.0 V, and the extraction cone at 4.0 V. A mass range of 50 to 1000 Daltons was covered with a scan time of 0.1 s and an interscan time of 0.02 s. The source temperature was set to 120°C and a solvation temperature of 450°C. Purified nitrogen gas was used for nebulisation with a flow rate of 550 L/h, and cone gas was set to 50 L/h. MassLynx v4.1 (SCN 872) software was used to control the hyphenated system and data analysis and manipulation.

Compound identification was enhanced for all samples by analysis with the collision cell's high and low collision energy settings. The compound fragmentation was minimised with a low energy setting of 3 eV, but to enhance the fragmentation of molecules, a collision energy ramp of 10 – 50 eV was used.

2.3 Results

Several methods were investigated to measure PAM activity, but this chapter focuses on absorbance-based assays employed due to their ease of use and equipment availability.

2.3.1 Pyridine, hippuric acid, benzenesulphonyl chloride concentration assay development

As HA was a common peptide substrate for PAM, the PHAB assay was developed to measure PAM activity. The absorbance assay method from Zinalibdin and Yacob (2013) was adapted to the multi-well plate high throughput format. The proposed reaction mechanism for the chromophore produced is shown in Figure 2.3.

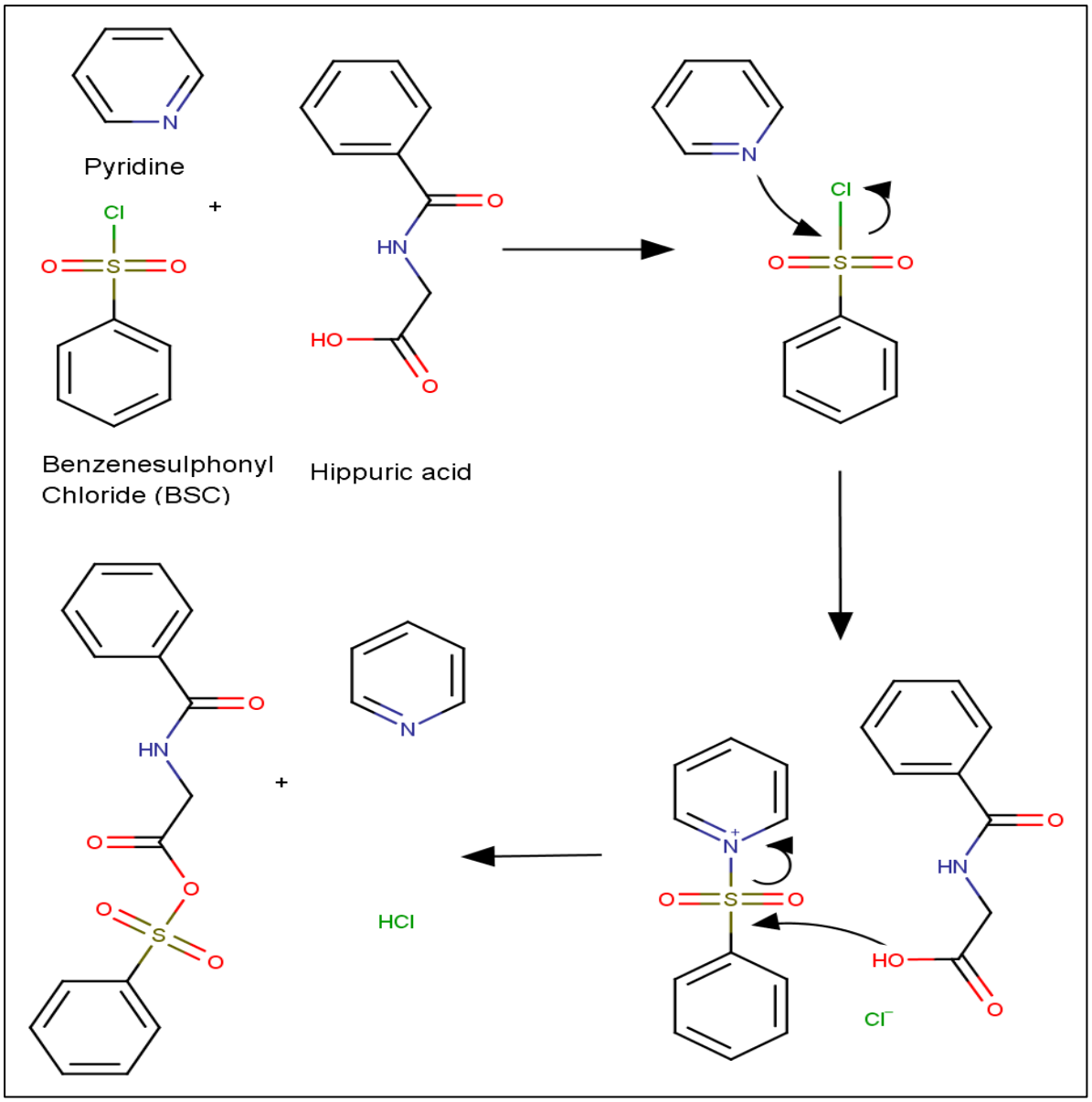


Figure 2.3: The proposed reaction mechanism between benzenesulphonyl chloride, pyridine, and hippuric acid to form a chromophoric product BSCH.

The reaction between benzenesulphonyl chloride, pyridine, and HA generates a chromophore end product 1-(benzenesulfonyl)pyridin-1-ium tris(2- (phenylformamido)acetic acid) tris(benzenesulfonyl chloride) tetrakis(pyridine) bis(benzenesulfonyl 2-(phenylformamido)acetate) chloride hydrochloride (BSCH). A spectrum scan of the PHAB solution was performed, with absorbance measured between 200 nm and 1000 nm, Figure 2.4.

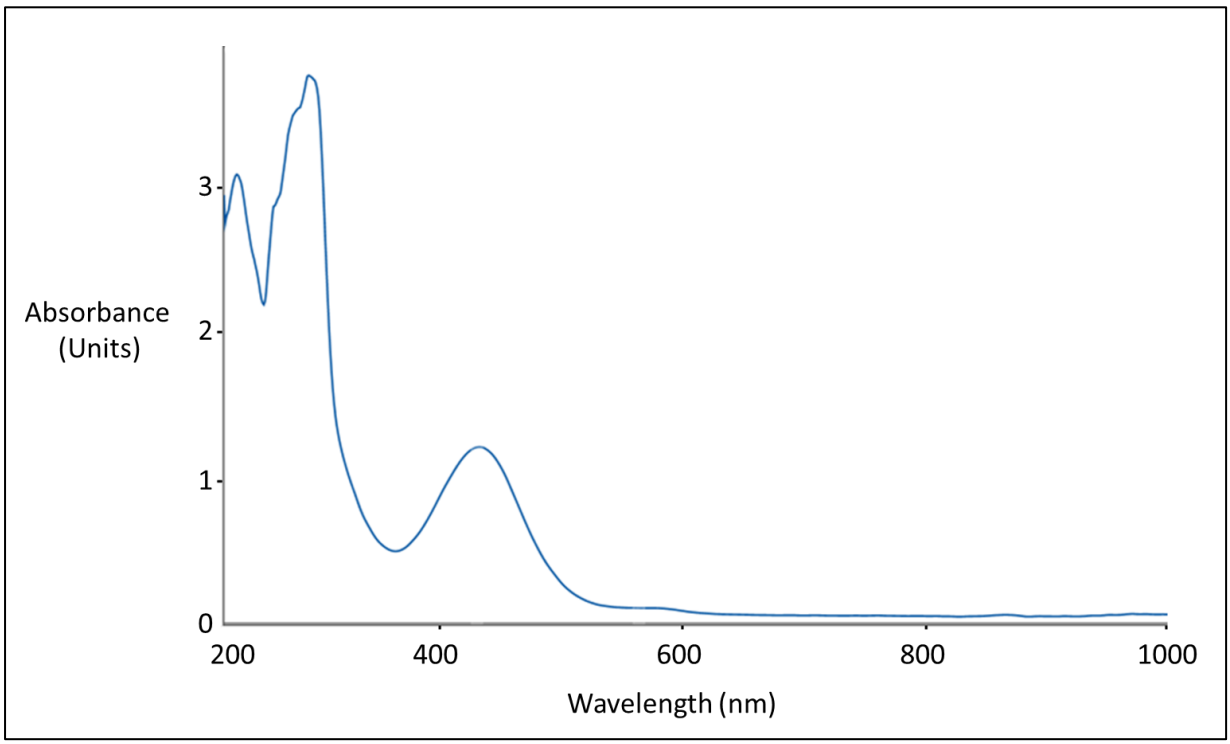


Figure 2.4: The spectrum scan between 200 nm and 1000 nm of the PHAB assay solution with 2.0 mM hippuric acid.

The product had the highest absorbance peak at 430 nm with a secondary, lower peak at 564 nm. Standard curves for the PHAB assay based on the HA substrate concentration, from 0.1 mM to 7 mM, are depicted in Figure 2.5.

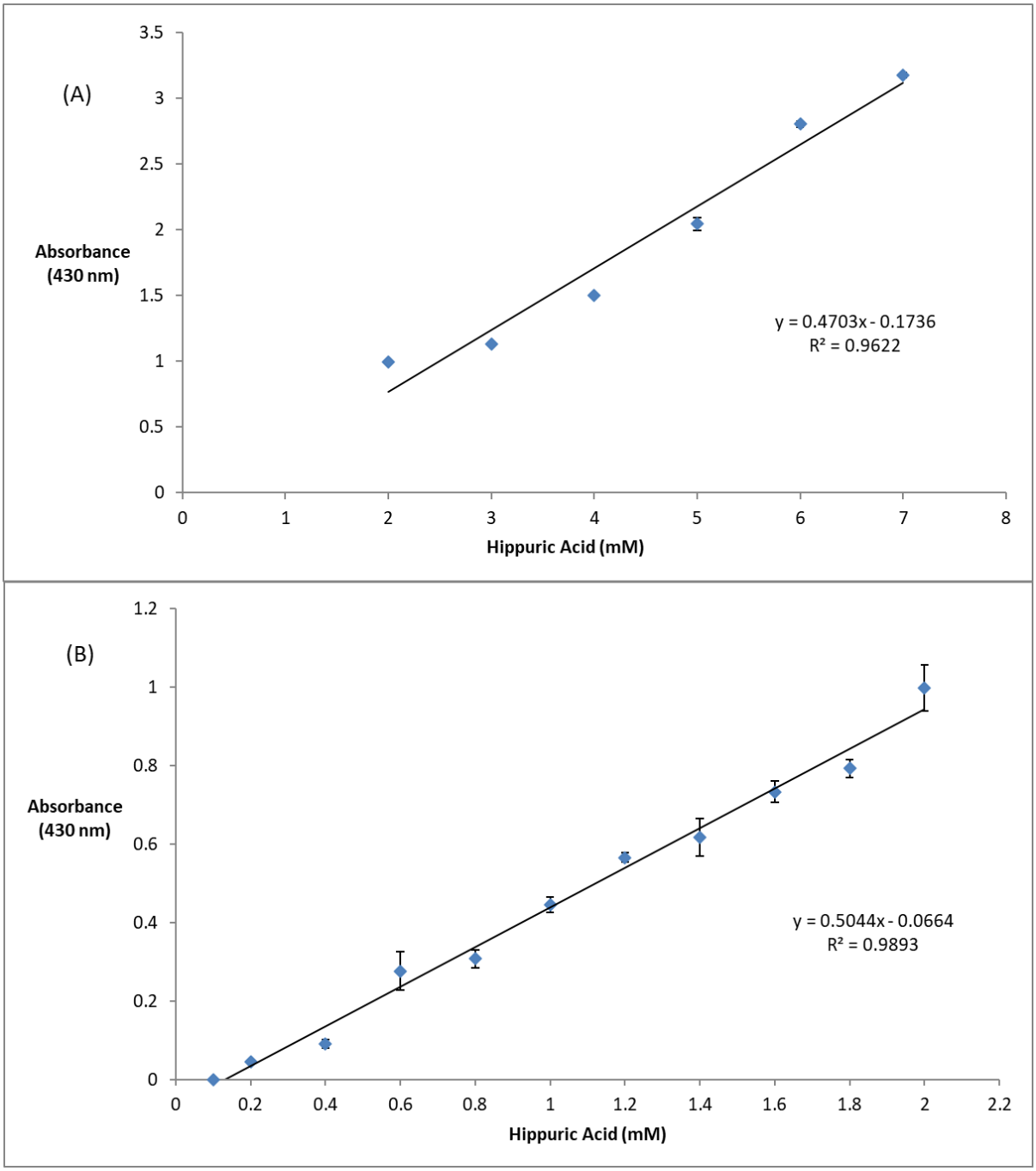


Figure 2.5: The standard curve with PHAB assay, measured at 430 nm for **(A)** 2 mM to 7 mM hippuric acid and **(B)** 0.1 mM to 2 mM hippuric acid (SD ±; n=3).

The standard curve for HA measured with the PHAB assay showed a linear relationship between HA concentrations and absorbance at 430 nm, allowing HA measurements from 0.2 mM to 7 mM, in Figure 2.5. The lowest HA concentration of 0.2 mM implied a limit of detection in the millimolar range. Concentrations of HA, under 0.2 mM, were investigated but did not have a linear relationship between absorbance and HA concentration. The low sensitivity levels and high standard deviation below 1 mM HA were not acceptable to measure

PAM activity. The relevant concentrations of enzymes produced with *Yarrowia lipolytica* and purified from flask fermentations would not be compatible with the PHAB assay mM sensitivity range.

Alternative assays were investigated, which would not be limited to measuring a single substrate and would allow the measurement of a variety of PAM peptide substrates. Notably, as glyoxylate is produced by PAM activity, it could be used to measure the sum of PHM and PAL activity.

2.3.2 2-Aminobenzaldehyde glycine glyoxylate concentration assay

As the development of a high throughput assay should ideally minimise the total number of reagents, steps, and enzymes added, the AGG assay for glyoxylate concentration from absorbance measurements was a promising method to use in a multi-well plate format. A spectrum scan from 200 nm to 1000 nm found the maximum absorbance peak at 450 nm, similar to the 440 nm used in the original paper (Soda *et al.*, 1973). A standard curve for glyoxylate was generated with the absorbance of the AGG chromophore measured at 450 nm (Figure 2.6).

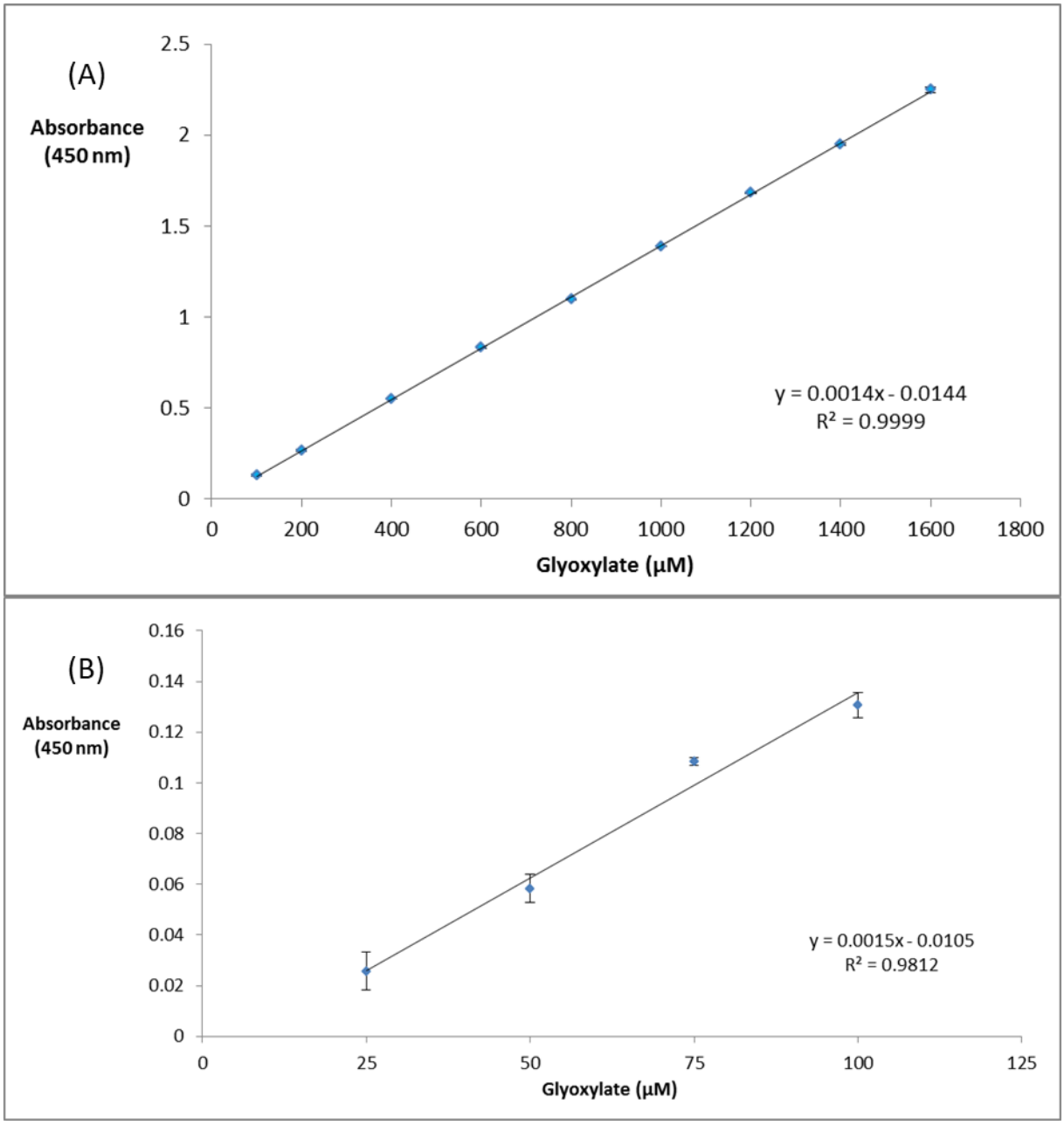


Figure 2.6: Standard curve for AGG assay, measured at 450 nm, for **(A)** 100-1600 µM glyoxylate and **(B)** 25-100 µM glyoxylate (SD ±; n=3).

The sensitivity of the glyoxylate standard curve, in the µM range, with AGG assay measured at 450 nm, was acceptable for the glyoxylate concentrations likely produced by PAM or PAL activity. A potential issue with the assay in this format was an inability to measure PHM products or activity without further modifications. The inclusion of an alkaline hydrolysis step could potentially expand the AGG assay's application to include glyoxylate produced from the hydrolysis of PHM hydroxylated peptide products.

An alkaline reagent would have to be added to a solution containing a suitable PHM hydroxylated product such as αHA to ascertain if this were possible. Care was taken to select

a suitable alkaline buffering reagent as a rapid uncontrolled change in pH, such as concentrated strong bases, could alter or degrade the analytes of interest.

The sequential addition of concentrated glycine at pH 10.0 before the addition of glycine at pH 2.5 was considered. Glycine was part of the original assay and could serve the dual purpose of changing the solution to alkaline before neutralising it. The final pH would have had to be measured regardless as it affects the AGG's chromophore synthesis (Soda *et al.*, 1973). A study using different volumes of 2 M glycine at pH 10.0 then pH 2.5, but keeping the final glycine concentration constant, was performed (Table 2.1).

Table 2.1: The pH of AGG solution generated from various volumes of 2 M Glycine pH 10.0 or pH 2.5.

2 M Glycine, pH 10 (µL)	pH I (basic)	2 M Glycine, pH 2.5 (µL)	pH II (final)
75	9.86	75	8.96
60	9.86	90	8.33
50	9.86	100	6.40
25	9.80	125	3.00

The pH study showed a 1:2 molar ratio of 2 M glycine pH 10.0 sequentially followed by 2 M glycine pH 2.5, which would result in a pH between 5.0 and 7.5 as recommended by Soda *et al.* (1973) to maximise chromophore synthesise. This conclusion was dependant on the assumption that the assay buffer used and other reagents in the activity solution had minimal effects on pH, which was likely given the high concentrations of glycine used. As a precaution, the pH was measured after any reaction with novel reagents, then adjusted if necessary to between pH 5.0-7.5 with concentrated NaOH or HCl.

To test the proposed changes and procedure, an alpha-carbon hydroxylated peptide substrate, in the form of αHA, was assayed with the AGG assay and pH hydrolysis step. The test would confirm that PHM activity would be measurable with the AGG assay as it was the final product after PHM activity with HA substrate. A potential issue with such an assay would be the time required for any hydrolysis step. To ensure all the hydroxylated substrates present, or those produced by preceding PHM activity, underwent hydrolysis, various lengths of time for alkaline hydrolysis were tested (Figure 2.7).

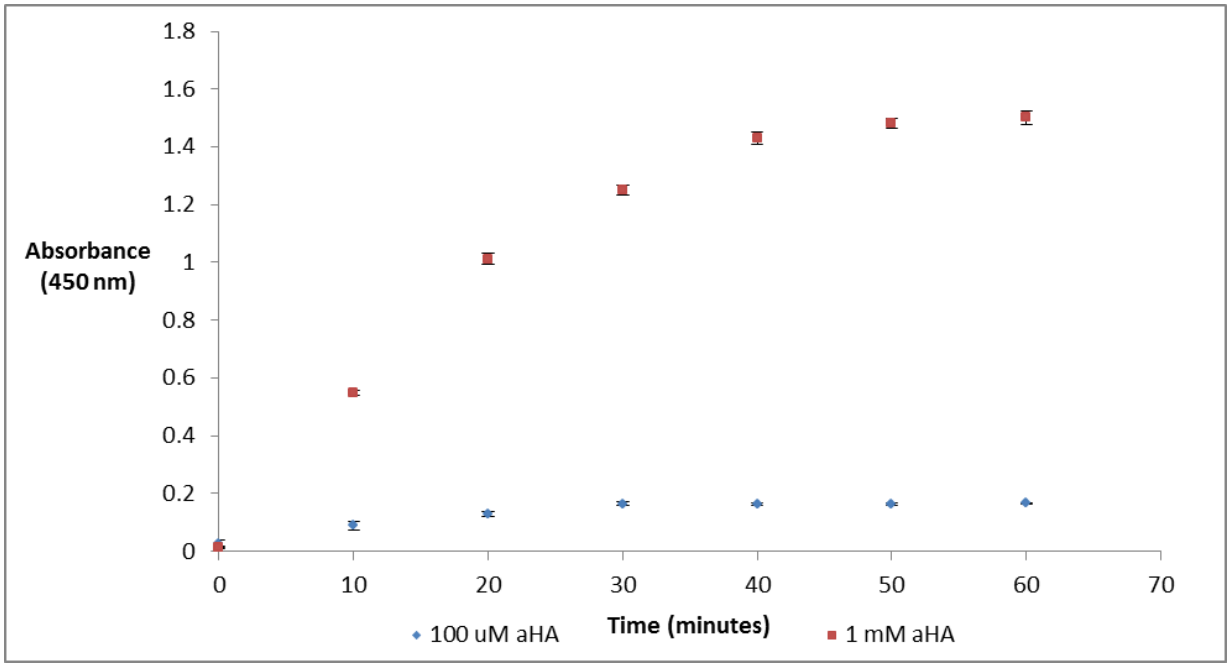


Figure 2.7: Time dependence for alkaline pH 10.0 induced hydrolysis of 100 µM and 1000 µM α-hydroxyhippuric, as measured with AGG assay at 450 nm. (SD ±; n=3)

The alkaline incubation step was observed to have achieved complete hydrolysis of both concentrations of αHA after 40 minutes at pH 10.0 at 37°C. These absorbance readings for 0.1 mM and 1.0 mM αHA were equivalent to the 100 µM and 1000 µM glyoxylate measurements, confirming their molar equivalency when measured with the AGG assay. The standardised alkaline hydrolysis methods could then be used to perform αHA standard curves (Figure 2.8).

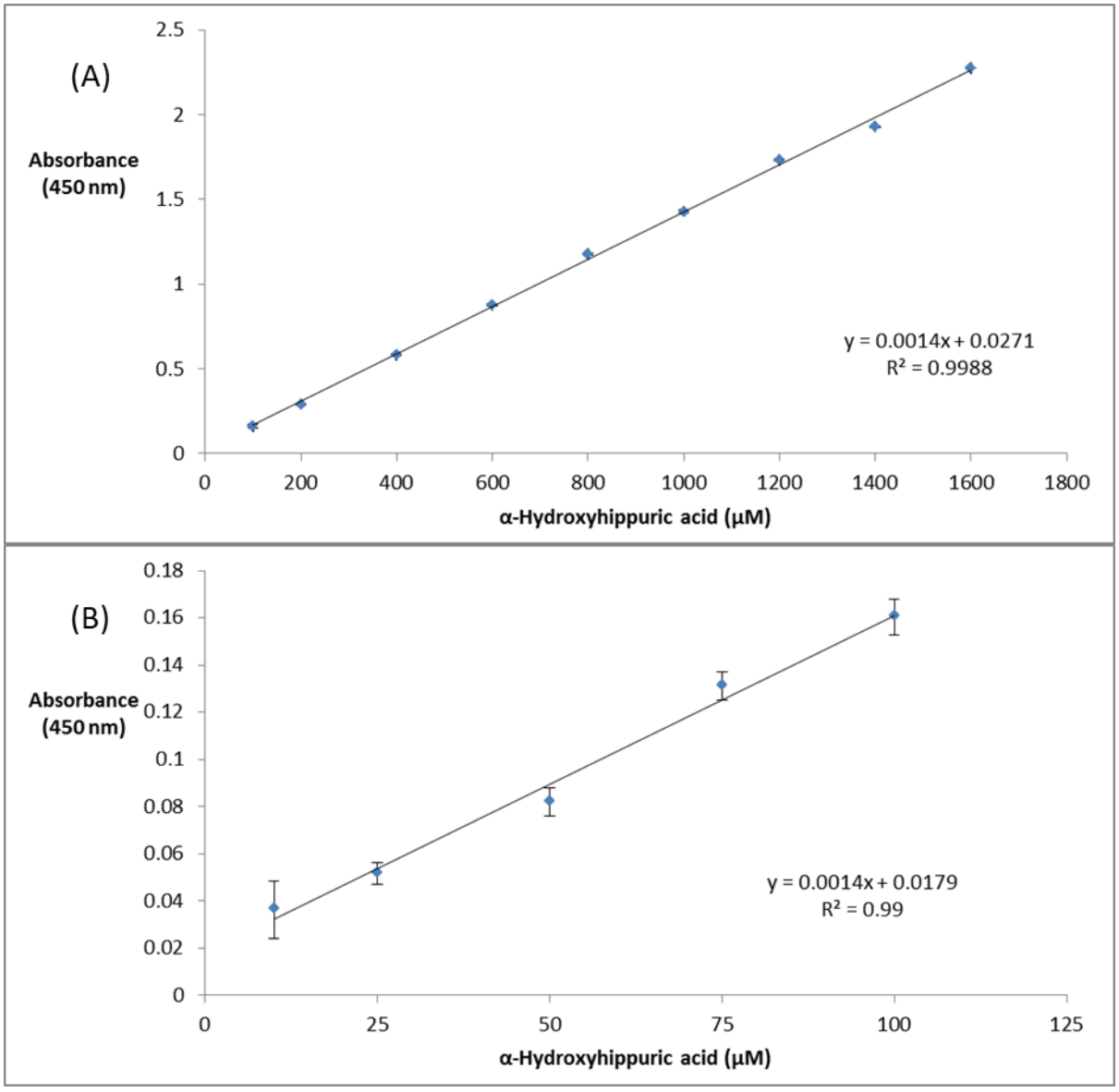


Figure 2.8: Standard curve for AGG assay, measured at 450 nm, for **(A)** 100-1600 µM α-hydroxyhippuric acid and **(B)** 10-100 µM α-hydroxyhippuric acid, after 40 minutes pH 10.0 induced hydrolysis (SD ±; n=3).

The regression equation produced for the 100-1600 µM αHA standard curve (Figure 2.8 A) was similar to the 100-1600 µM glyoxylate standard curve (Figure 2.6A). Overlaying the chromatogram produced by equimolar concentrations of αHA and glyoxylate allows a comparison of the chromatogram peaks produced with the chromophore from each reaction. The MS analysis of the chromophore and products generated by each reaction would also confirm the analytes' identity. The chromophore's identity for samples of 100 µM glyoxylate and 100 µM αHA were analysed with UPLC-MS under identical conditions (Figure 2.9).

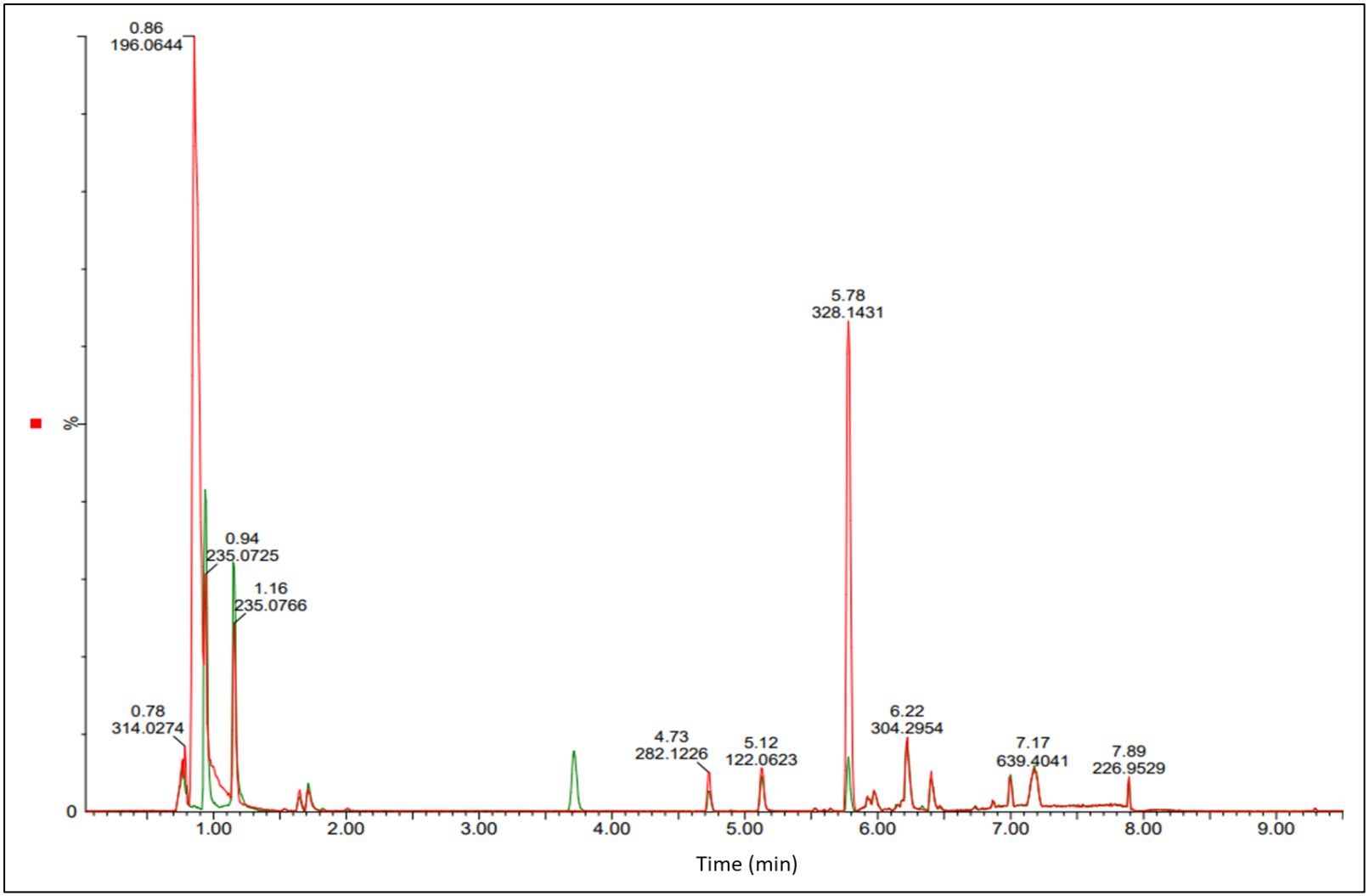


Figure 2.9: An overlaid chromatogram of 1 mM glyoxylate (red) and 1 mM αHA (green) after AGG assay, in electrospray positive mode for percentage peak height (%) versus time (minutes). The elute time, in minutes, and analyte mass, in Daltons, are presented above each relevant peak.

The masses of the final AGG chromophore (235.0725 Da and 235.0766 Da) are in good agreement with the theoretical/calculated mass (234.211) (Figure 2.9). At 0.94 and 1.16 minutes, the area under the peaks corresponding to the AGG chromophore (at 235.0725 Da and 235.0766 Da) was>95% equivalent between the respective glyoxylate and αHA samples. The maximum peak produced in the chromatogram, seen at 0.86 min of 196.0644 Da, was the 100 mM MES buffer, pH 6.0. The confirmed masses of the final chromophore and predicted chemical formula of $C_{11}H_{10}N_2O_4$ provided by the MassLynx software were compared and seen to be equivalent. A proposed synthesis pathway for the possible chromophore from the AGG assay was proposed (Figure 2.10).

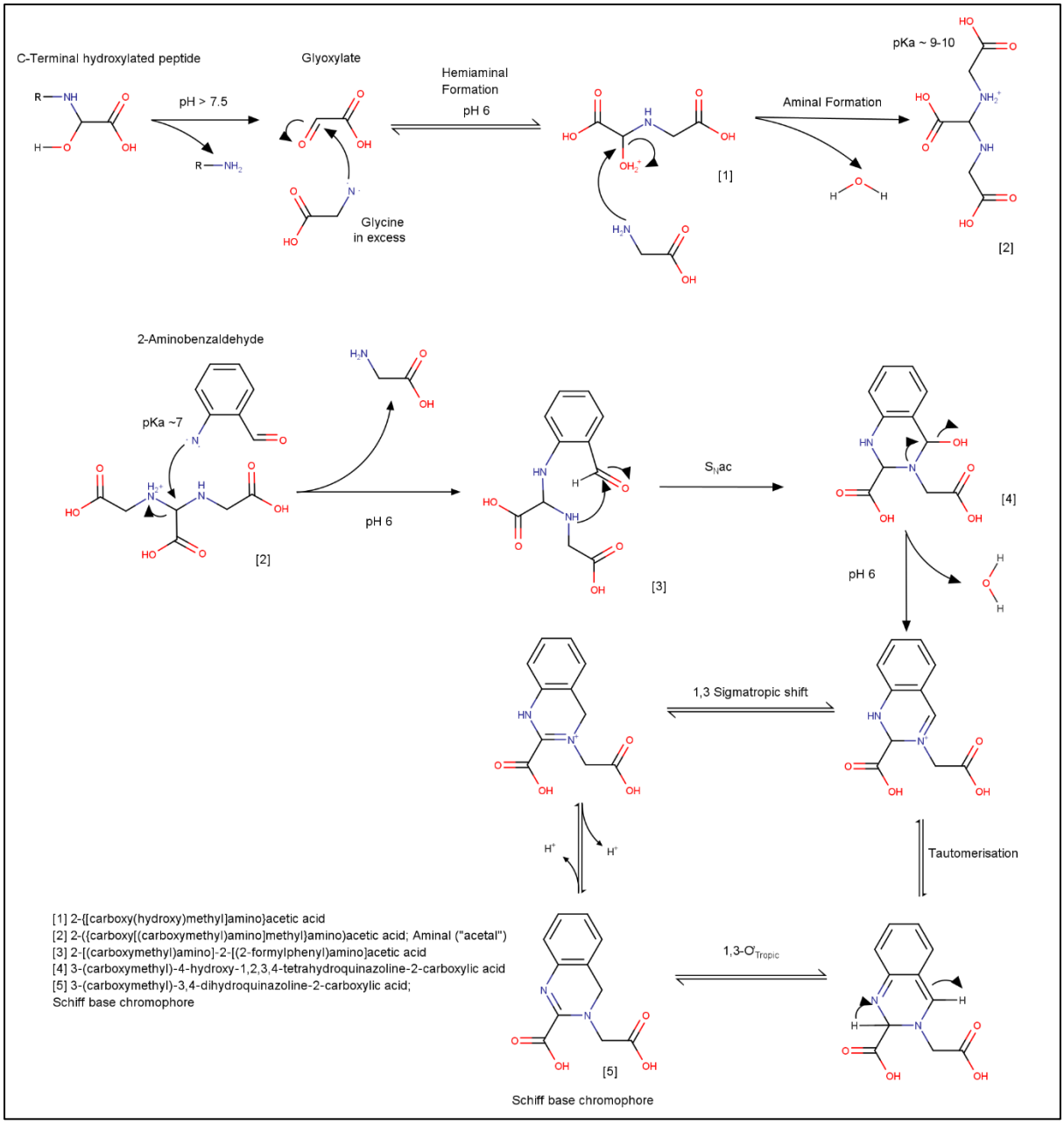


Figure 2.10: The theorised reaction mechanism for the AGG assay, including hydrolysis of glyoxylate from an amidated peptide to the final chromophore.

The chromophore could be classified as a Schiff base as per Moss, Smith and Tavernier, (1995) and was stable, being stored at 4°C for multiple weeks with no significant changes in absorbance recorded after multiple measurements.

2.4 Discussion

A fundamental problem with detecting PAM activity has been relatively low levels of specific activity of the PHM catalytic core. The 0.42 to 9.2 s^{-1} product turnover rate measured with purified PHM protein mean micrograms of active PAM present in a reaction solution can produce in nM to µM activity range with (Francisco *et al.*, 2003; Satani *et al.*, 2003; Bauman *et al.*, 2006). While higher theoretical k_{CAT} values for the *Rattus norvegicus* PHM have been calculated, these rates are not achievable as the oxygen concentrations would be above oxygen saturation limits in a reaction solution (Francisco *et al.*, 2003). The millimolar range of sensitivity recorded for the PHAB assay HA standard curve is not sensitive enough for detecting PAM or PHM activity levels with commercially available quantities of PAM (10 µg) is required, if µg quantities are required per assay (Figure 2.4). Unless a novel PAM or PHM was discovered which possessed a turnover rate of >1000 s^{-1}, this assay would not be applicable and would require prohibitively large quantities of protein.

The low turnover rate in PHM activity is a physio-chemical constraint inherent due to the bound copper ions requiring valence state reduction between each reaction to regenerate their catalytic ability. Unless a novel class of enzymes that can perform PHM's selective proteolysis with a higher turnover active site, such as serine protease, be discovered, PHM and PAM reactions will have maximum activity and turnover levels comparable to present examples. The low activity levels of PHM mean PHAB assay would be unsuitable for µg levels of PHM or PAM, as purified in a laboratory setting and commercially available.

The higher level of sensitivity with the AGG assay and the strength of the relationship between absorbance and concentration of R^2>0.99 makes this a functional assay for PAM and PAL activity by measuring the production of glyoxylate. The modification of the assay to measure PHM activity by including an alkaline hydrolysis step is a significant improvement from the previous assay. An implication from the hydrolysis time study with the αHA utilising only alkaline pH conditions is the slow rate of catalysis, requiring 40 minutes for complete hydrolysis. The delay and the endpoint assay results are circumvented by running an extensive array of numerous samples simultaneously in a high throughput fashion with a multi-well plate and suitable spectrophotometer.

The requirement of highly alkaline conditions with a pH>9.5 for *in vitro* hydrolysis is significantly higher than the neutral pH of cellular cytoplasm, explaining the co-expression of PAL and PHM in many organisms (Attenborough *et al.*, 2012). The absence of these cellular

conditions would form a significant rate-limiting step for intracellular peptide C-terminal amidation. The lack of consistent natural hydrolysis conditions implies the hydroxylation may serve an alternate purpose or function. The advantages of high turn-over rate and consistency of activity demonstrate the evolutionary advantage to retaining a PAL enzyme. Conversely, the lack of PAL in many organisms implies that there may be independent functions of the PHM enzyme and its activity outside of hydroxylating a substrate peptide for the subsequent lyase by PAL. The exploration of such a hypothesis would likely require highly sensitive assays, such as UPLC-MS, which greatly assisted in the validation of the identity of the proposed AGG chromophore and alkaline hydrolysis step.

The dual peaks observed for the AGG chromophore may be generated from the point of chirality introduced during synthesis. An alternate explanation would be alternating resonance forms of the final chromophore due to the benzene ring in the 3-(carboxymethyl)-3,4-dihydroquinazoline-2-carboxylic acid. The two stereoisomers likely explain the 235.0725 Da and 235.0766 Da masses, which differ in elution times despite similar structures. The presence of two peaks for the AGG chromophore did not alter the conclusion that the AGG assay developed would provide a viable method for measuring glyoxylate produced by PAM activity.

The development of a glyoxylate based assay in conjunction with a pH hydrolysis step allows for the determination of both PAL and PAM activity by measuring any glyoxylate produced. The sensitivity and accuracy of the AGG assay enabled the absorbance measurement of glyoxylate concentrations in the μM range of glyoxylate αHA, likely sufficient for the measurement of PAM and PAL activity. The inclusion of an alkaline hydrolysis step with glycine, as demonstrated with αHA, expands the scope of the assay to include measuring PHM activity. A variety of peptide substrates could be studied with PHM and AGG, provided the glyoxylate residue is hydrolysed.

2.5 References

Alwan, K. B., Welch, E. F. and Blackburn, N. J. (2019) 'Catalytic M center of copper monooxygenases probed by rational design. Effects of selenomethionine and histidine substitution on structure and reactivity', *Biochemistry*, 58(44), pp. 4436–4446. doi: 10.1021/acs.biochem.9b00823.

An, Z., Chen, Y., Koomen, J. and Merkler, D. (2012) 'A mass spectrometry-based method to screen for α-amidated peptides', *Proteomics*, 12(2), pp. 173–182. doi: 10.1002/pmic.201100327.

Attenborough, R. M. F., Hayward, D., Kitahara, M., Miller, D. and Ball, E. (2012) 'A "neural" enzyme in nonbilaterian animals and algae: Preneural origins for peptidylglycine α-amidating monooxygenase', *Molecular Biology and Evolution*, 29(10), pp. 3095–3109. doi: 10.1093/molbev/mss114.

Bauman, A. T., Yukl, E. T., Alkevich, K., McCormack, A. L. and Blackburn, N. J. (2006) 'The hydrogen peroxide reactivity of peptidylglycine monooxygenase supports a Cu(II)-superoxo catalytic intermediate', *Journal of Biological Chemistry*, 281(7), pp. 4190–4198. doi: 10.1074/jbc.M511199200.

Bradbury, A. F., Finnie, M. D. A. and Smyth, D. G. (1982) 'Mechanism of C-terminal amide formation by pituitary enzymes', *Nature*, 298(5875), pp. 686–688. doi: 10.1038/298686a0.

Bradbury, A. F. and Smyth, D. G. (1987) 'Biosynthesis of the C-terminal amide in peptide hormones', *Bioscience Reports*, 7(12), pp. 907–916. doi: 10.1007/BF01122123.

Brock, T., Humm, J. and Kizer, J. S. (1989) 'Assay of peptidylglycine monooxygenase: Glycine-directed amidating enzyme', in *Methods in Enzymology*, pp. 351–358. doi: 10.1016/0076-6879(89)68026-3.

Carpenter, S. E. and Merkler, D. J. (2003) 'An enzyme-coupled assay for glyoxylic acid', *Analytical Biochemistry*, 323(2), pp. 242–246. doi: 10.1016/j.ab.2003.09.012.

Carpenter, S. E. (2006) Enzyme linked spectroscopic assays for glyoxylate: The use of peptidylglycine alpha-amidating monoxygenase for the discovery of novel alpha-amidated hormones. PhD thesis. University of South Florida. Available at: http://scholarcommons.usf.edu/etd/2472 (Accessed: 1 November 2013).

Chew, G. H. (2003) Substrate-based inhibitors of peptidylglycine α-amidating monooxygenase (PAM) as anti-proliferative drugs for cancer. MSc thesis. University of South Florida. Available at: http://scholarcommons.usf.edu/etd/1341.

Francisco, W. A., Blackburn, N. J. and Klinman, J. P. (2003) 'Oxygen and hydrogen isotope effects in an active site tyrosine to phenylalanine mutant of peptidylglycine α-hydroxylating monooxygenase: Mechanistic implications', *Biochemistry*, 42(7), pp. 1813–1819. doi: 10.1021/bi020592t.

Gainer, H., Russell, J. T. and Loh, Y. P. (1985) 'The enzymology and intracellular organization of peptide precursor processing: The secretory vesicle hypothesis', *Neuroendocrinology*, 40(2), pp. 171–184. doi: 10.1159/000124070.

Katopodis, A., Ping, D., Smith, C. and May, S. (1991) 'Functional and structural characterization of peptidylamidoglycolate lyase, the enzyme catalyzing the second step in peptide amidation', *Biochemistry*, 30(25), pp. 6189–6194. doi: 10.1021/bi00239a016.

Katopodis, A. G. and May, S. W. (1990) 'Novel substrates and inhibitors of peptidylglycine α-amidating monooxygenase', *Biochemistry*, 29(19), pp. 4541–4548. doi: 10.1021/bi00471a006.

Kolhekar, A. S., Keutmann, H., Mains, R., Quon, A. and Eipper, B. (1997) 'Peptidylglycine α-hydroxylating monooxygenase: Active site residues, disulfide linkages, and a two-domain model of the catalytic core', *Biochemistry*, 36(36), pp. 10901–10909. doi: 10.1021/bi9708747.

Merkler, D. J., Asser, A., Baumgart, L., Carballo, N., Carpenter, S., Chew, G., Cosner, C., Dusi, J., Galloway, L., Lowe, A., Lowe, E., King, L., Kendig, R., Kline, P., Malka, R., Merkler, K., McIntyre, N., Romero, M., Wilcox, B. and Owen, T. (2008) 'Substituted hippurates and hippurate analogs as substrates and inhibitors of peptidylglycine α-hydroxylating monooxygenase (PHM)', *Bioorganic & Medicinal Chemistry*, 16(23), pp. 10061–10074. doi: 10.1016/j.bmc.2008.10.013.

Moss, G. P., Smith, P. A. S. and Tavernier, D. (1995) 'Glossary of class names of organic compounds and reactive intermediates based on structure (IUPAC recommendations 1995)', *Pure and Applied Chemistry*, 67(8–9), pp. 1307–1375. doi: 10.1351/pac199567081307.

Mueller, G. P. and Driscoll, W. J. (2002) 'α-Amidated peptides: Approaches for analysis', in *Posttranslational Modification of Proteins*. New Jersey: Humana Press, pp. 241–257. doi: 10.1385/1-59259-181-7:241.

Ogata, M. and Sugihara, R. (1977) 'An improved direct colorimetric method for the quantitative analysis of urinary hippuric acid as an index of toluene exposure', *Acta Medica Okayama*, 31(4), pp. 235–242.

Satani, M., Takahashi, K., Sakamoto, H., Harada, S., Kaida, Y. and Noguchi, M. (2003) 'Expression and characterization of human bifunctional peptidylglycine α-amidating monooxygenase', *Protein Expression and Purification*, 28(2), pp. 293–302. doi: 10.1016/S1046-5928(02)00684-8.

Shimoi, H., Kawahara, T., Suzuki, K., Iwasaki, Y., Jeng, A. and Nishikawa, Y. (1992) 'Characterization of a *Xenopus laevis* skin peptidylglycine alpha-hydroxylating monooxygenase expressed in insect-cell culture', *European Journal of Biochemistry*, 209(1), pp. 189–194. doi: 10.1111/j.1432-1033.1992.tb17276.x.

Soda, K., Toyama, S., Misono, R., Rirasawa, T. and Asad, K. (1973) 'Spectrophotometric determination of glyoxylic acid with o-aminobenzaldehyde and glycine, and its application to enzyme assay', *Agricultural and Biological Chemistry*, 37(6), pp. 1393–1400. doi: 10.1080/00021369.1973.10860828.

Tajima, M., Iida, T., Yoshida, S., Komatsu, K., Namba, R., Yanagi, M., Noguchi, M. and Okamoto H. (1990) 'The reaction product of peptidylglycine alpha-amidating enzyme is a hydroxyl derivative at alpha-carbon of the carboxyl-terminal glycine.', *Journal of Biological Chemistry*, 265(17), pp. 9602–9605. doi: 10.1016/S0021-9258(19)38709-5.

Takahashi, K., Okamoto, H., Seino, H. and Noguchi, M. (1990) ‘Peptidylglycine α-amidating reaction: Evidence for a two-step mechanism involving a stable intermediate at neutral pH’, *Biochemical and Biophysical Research Communications*, 169(2), pp. 524–530. doi: 10.1016/0006-291X(90)90362-Q.

Tatemoto, K. and Mutt, V. (1978) ‘Chemical determination of polypeptide hormones’, *Proceedings of the National Academy of Sciences of the United States of America*, 75(9), pp. 4115–4119. doi: 10.1073/pnas.75.9.4115.

Tomokuni, K. and Ogata, M. (1972) ‘Direct colorimetric determination of hippuric acid in urine’, *Clinical Chemistry*, 18(4), pp. 349–351. doi: 10.1093/clinchem/18.4.349.

Yin, P., Bousquet-Moore, D., Annangudi, S., Southey, B.,Mains, R., Eipper, B. and Sweedler, J. (2011) ‘Probing the production of amidated peptides following genetic and dietary copper manipulations.’, *PloS One*, 6(12), p. e28679. doi: 10.1371/journal.pone.0028679.

Yoshida, M., Akane, A., Mitani, T. and Watabiki, T. (2005) ‘Simple colorimetric semiquantitation method of hippuric acid in urine for demonstration of toluene abuse’, *Legal Medicine*, 7, pp. 198–200. doi: 10.1016/j.legalmed.2005.02.001.

Zinalibdin, M. R. and Yacob, A. R. (2013) ‘Detection of hippuric acid: A glue solvent metabolite, using a mobile test kit’, *Arabian Journal of Chemistry*, 6(1), pp. 115–120. doi: 10.1016/j.arabjc.2010.09.029.

Chapter 3: Bioinformatics, cloning, and construction of PAM, PHM, and PAL expression vectors

3.1 Introduction

Bioinformatics is essential in the process of searching, analysing, designing, and constructing genes and vectors for protein expression (Apweiler *et al.*, 2004). Computational modelling is used for finding protein sequences, designing genes, constructing vectors, and assessing protein structures, which are essential tasks in protein expression studies. Bioinformatics systems were vital for discovering PAM and handling the large amounts of data in determining the DNA and protein sequences associated with PAM (Prigge *et al.*, 2000; Vishwanatha *et al.*, 2014; Antošová and Sychrová, 2016).

3.1.1 PAM protein discovery, description, and characterisation

Since the early discovery and characterisation of C-terminal amidation of peptides, especially neuropeptides, the identity of the enzyme responsible for this post-translational modification has been pursued (Bradbury *et al.*, 1982). The mechanism of amidation by enzymes was first described by Bradbury *et al.*, (1982), and active PAM protein was initially purified from porcine pituitary fractions (Glembotski, 1985; Wand *et al.*, 1985). Further protein purification studies led to the identification of the PAM protein as the enzyme primarily responsible for C-terminal amidation activity (Glembotski *et al.*, 1984; Wand *et al.*, 1985). Subsequent studies have identified the catalytic cores (cc), PHM (EC: 1.14.14.3) and PAL (EC: 4.3.2.5), and their essential metal cofactors (Bradbury and Smyth, 1987; Eipper *et al.*, 1992; Kolhekar *et al.*, 1997). The presence of two catalytic cores revealed that the PAM reaction proceeds by a two-step mechanism. An initial hydroxylation reaction of the α-carbons at the C-terminal glycine occurs, followed by the hydrolysis of the peptide bond, yielding an amidated C-terminus (Katopodis *et al.*, 1990).

With C-terminal amidation ranking among the most common peptide modifications, it was important to identify the enzymes responsible for this modification in translational, post-translational, and communication pathways (Bauman *et al.*, 2007). In addition to the role of C-terminal amidation in cellular peptide communication, it also affects antimicrobial peptide production, protein translation via ubiquitin, and cell mobility via cellular cilia (Chew *et al.*, 2005; Huang *et al.*, 2010; Attenborough *et al.*, 2012; Kumar *et al.*, 2016). The majority of these

processes have been found across all eukaryotes, with ubiquitin found in all eukaryotes (Kumar *et al.*, 2016). PAM amino acid (AA) sequences, and the corresponding DNA sequences, have been purified from many source organisms, including the catalytic cores PHM and PAL, forming a base for searches of novel PAMs or subunits in eukaryotes. (Prigge *et al.*, 2000; Vishwanatha *et al.*, 2014; Antošová and Sychrová, 2016; Kumar *et al.*, 2016).

3.1.2 The importance of bioinformatics in studying PAM

The generation of large amounts of data by current sequencing efforts has necessitated the development of numerous tools, programs, and online platforms to manage it (Apweiler *et al.*, 2004). In particular, the Universal Protein Resource (UniProt) consortium, managed by the European Bioinformatics Institute, the Swiss Institute of Bioinformatics, and the Protein Information Resource, is an invaluable database of resources for sequence analysis, such as sequence searches and multiple sequence alignments (MSA) (Thompson *et al.*, 1994; Apweiler *et al.*, 2004; Sievers and Higgins, 2018). Using MSA of sequences of interest allows for the fast and accurate comparison of conserved sequences and structural elements of protein and nucleic acids from multiple organisms, revealing key features and sequences of a protein. The original search engines and software for MSA have improved dramatically for ease of use and faster searches, as typified by the development of the Clustal software program (Apweiler *et al.*, 2004; Sievers and Higgins, 2018).

The original Clustal software packages were designed to assist MSA by offering a simplified and faster method for clustering genetic and protein sequences (Sievers and Higgins, 2018). The Clustal Omega alignments' efficiency was increased by creating “guide trees” consisting of an initial clustering of sequences used progressively for later phases of sequences alignments (Sievers and Higgins, 2018). These guide trees consisted of closely aligned sequence pairs which in turn were aligned with other similar paired lines of sequences and previous alignments to form clusters. The clusters are then topologically positioned to create a guide tree based on established phylogenetic relationships and prior alignments. The second significant improvement to the Clustal Omega platform was the inclusion of the “HHalign” alignment engine to align the profile hidden Markov models to each other instead of relying on dynamic conventional programming and profile alignment (Sievers and Higgins, 2018). These improvements meant thousands of sequences could be aligned without losing accuracy and provide a platform for finding and comparing protein sequences from diverse origins

3.1.2.1 PAM DNA and protein sequence searches and identification

The use of protein MSA is a key tool when searching for similar proteins, because, unlike DNA alignments and searches, protein structure, as well as critical active site residues, are conserved to a higher degree at the protein level (Illergård *et al.*, 2009; Sievers and Higgins, 2018). In finding PAMs, it is crucial to limit searches to only the PHM subunit, as it performs the essential enzymatic modification, the hydroxylation of the αC of the C-terminal AA residue, and PHM can be expressed separately to the PAL subunit. It is necessary to understand how conserved active site residues and structural elements like cysteine bonds determine enzyme structure and function.

Figure 3.1: Protein sequence feature map of *Rattus norvegicus* PHM catalytic core from PAM-1 isoform protein. Map created with sequences from Stoffers, Barthel-Rosa Green and Eipper, (1989).

The most important features of the PHM subunit are the metal-binding sites for the two copper ions Cu_H and Cu_M, which form despite the distal sequence length between the six metal-binding residues (Figure 3.1), adapted from Stoffers *et al.*, (1989). The copper-binding sites consist of five histidines and a single methionine, with the native structure forming the two sites within 11 Å (Prigge, 1997). Mutations of these residues, which disrupt the active site or prevent metal-binding, have been shown to inhibit PAM activity (Kolhekar *et al.*, 2002). The multiple disulphide bonds found throughout the protein sequence likely assist in forming these metal-

binding sites, as well as in maintaining the overall enzyme tertiary structure, which is annotated with red lines in Figure 3.1 as created with sequences from Prigge (1997). Since eukaryotes require post-translational modification of proteins, eukaryotic organisms are likely to possess many PAM homologues.

The massive expansion of environmental genome sequencing has revealed that the lower eukaryote kingdoms, and specifically fungi, have higher levels of diversity in their genomes than that of other biological kingdoms (Sello *et al.*, 2015). This diversity has been observed across families of enzymes of industrial interest, for example, mono-oxygenases such as laccases and cytochrome P450, which PAM and PHM are structurally similar to (Chen *et al.*, 2014). The identification of a fungal homologue of PAM is of particular interest. To date, no fungal PHM or PAM has been found, and its expression in a yeast platform has been unsuccessful at industrially relevant levels (El Meskini *et al.*, 2003; Kumar *et al.*, 2016). In particular, a yeast homologue, or a homologue in its close relatives, is likely to increase heterologous expression and thus has the potential to be expressed in high quantities (Çelik and Çalık, 2012).

3.1.2.2 Protein structural sequences of PAM and its subunits

To date, though only PHM and PAL have been recombinantly expressed from *Homo sapiens*, *Conus bullatus*, *R. norvegicus*, PAM has been purified from many other animals, including *Bos taurus* and *Caenorhabditis elegans*. The complete domain schema of the archetypal PAM from *Rattus norvegicus* is graphically represented in Figure 3.2.

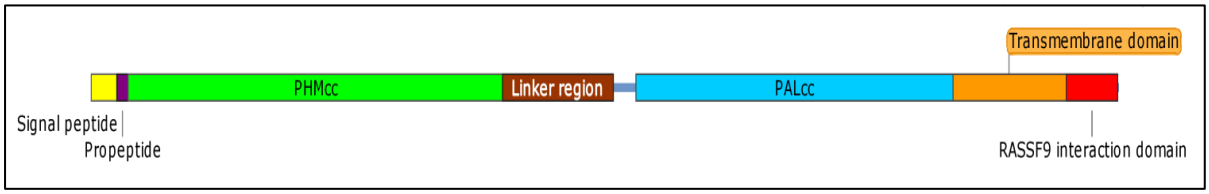


Figure 3.2: Protein domain map of *Rattus norvegicus* PAM-1 isoform protein used as the canonical reference for PAM gene. Map created with sequences from Stoffers, Barthel-Rosa Green and Eipper, (1989).

The structure of the *R. norvegicus* PAM-1 isoform is typical of a complete PAM protein, including all the possible protein regions associated with PAM for the majority of eukaryotic PAM's putatively identified and expressed to date. Various isoforms with fewer regions have been documented from this canonically complete PAM-1 protein, including isoforms with only PHM or PAL regions in some species (El Meskini *et al.*, 2003; Kumar *et al.*, 2016). These catalytic cores can also variably contain some of the proximate regions of the complete PAM

protein. Thus, PHM isoforms have been documented with complete or partial sequences of the signal peptide, propeptide, and linker regions (Prigge, 1997). Similarly, PAL isoforms can contain complete or partial sequences of the linker, transmembrane, and RASSF9 regions. The variability of PAM isoforms suggests that the structure and activity of each catalytic core is independent, which has been crucial in expression attempts of non-membrane associated PHM and PAL in heterologous expression systems (Husten and Eipper, 1991; Milgram *et al.*, 1992). This sequence paradigm for identified PAMs and their associated elements is not observed in the potential PHM identified in *Cyphellophora europaea*, which is shown as a protein domain map in Figure 3.3 created from sequence data from Cuomo *et al.*, (2014).

Figure 3.3: Protein domain map of *Cyphellophora europaea* putative PAM protein with domain identity based on similarity to *Rattus norvegicus* PAM gene.

The differences between the potential fungal PHM domains and the standard sequence include three regions not previously associated with eukaryotic PAM, and PAL is absent (Figure 3.3). The differences may be due to evolutionary changes in the cellular topology and the function of PHM or PAM activity in fungi.

A striking difference between the *C. europaea* putative PAM/PHM sequence and the standard PAM sequence is a dopamine-monooxygenase N-terminal (DOMON) domain on the N-terminus of the protein, immediately before the PHM domain (Cuomo *et al.*, 2014). DOMON domains have been associated with dopamine-β-monooxygenases (DBM), cellobiohydrolases, and xylanases, and can have heme and sugar moieties' binding abilities (Iyer *et al.*, 2007; Vendelboe *et al.*, 2016). This domain has not been found with a PAM or PHM previously. Its association with DBM, the closest protein analogue to PHM, implies it shares a common evolutionary lineage with PAM (Aravind, 2001). The presence of the DOMON domain, which is typically associated with protein trafficking to the extracellular compartment, indicates that these enzymes are likely to be extracellular and therefore more pH- and temperature-tolerant than typical PHMs (Aravind, 2001).

The cytochrome b561 domains of the putative *C. europaea* PAM have ascorbate regeneration activity by supporting oxidation from the semi-dehydroascorbate state to the active L-ascorbate form (Cuomo *et al.*, 2014). The ascorbate oxidation activity may aid the activity of the PHM

region, as it depends on a reducing agent like ascorbate, and the availability and concentration of ascorbate outside the cell are most likely to be quite variable. Thus, the cells cannot control PHM conditions and supply co-factors to the extent they can with intracellular proteins within the cytosol or the Golgi.

For the canonically confirmed PAM variants identified in hosts or heterologous expressed recombinant proteins, their expression in an active state seems to require post-translational modifications (PTM). Prominent amongst the PTM for PHM cc are disulphide bond formations. The specific folding conditions and chaperones found in the ER and Golgi may be essential for the nascent protein to fold correctly and form disulphide bridges (El Meskini *et al.*, 2003). These requirements theoretically limit PAM, plus its PHM and PAL subunits, to recombinant expression in eukaryotic organisms (Attenborough *et al.*, 2012). The expense and difficulties associated with the manipulation and expression of this protein in industrial quantities limit the likely expression hosts to fungi, particularly yeast. However, no yeast homologues of PAM have been found as of 2019 (Attenborough *et al.*, 2012; Kumar *et al.*, 2019). As the heterologous expression of genes can be variable, the design of the recombinant gene is essential to counteract problems in the expression of proteins. The use of recombinant gene design has been essential in the expression of active recombinant PAM and subunits (Prigge, 1997).

3.1.2.3 Recombinant protein selection, design, and construction

Once the identity and structure of the PAM, PHM, and PAL genes were established, gene selection became a priority in the successful expression of these genes (Prigge, 1997). Four aspects that govern the successful selection, design, and construction of a recombinant gene in yeast were proposed by Antošová and Sychrová (2016): the selection of a promoter, selection of a secretion signal sequence, optimisation of codon usage, and genetic engineering. For the majority of expression studies in yeast, codon optimisation allowed for higher protein yields (Antošová and Sychrová, 2016). The truncation and editing of genes to remove pre-proteins, transmembrane domains, and other regions like linker regions are also important to controlling yield in recombinant expression systems, as these unnecessary regions could negatively impact protein expression (Madzak *et al.*, 2004; Çelik and Çalık, 2012). During gene synthesis, common protease recognition sites in potential yeast and expression hosts, such as *Y. lipolytica*, with the amino acid sequence X-KR-X, should be removed if possible (Peberdy, 1994; Young *et al.*, 1996). The selection of promoter sequences and secretion signal sequences can influence protein production and activity levels (Bulani *et al.*, 2012).

3.1.2.4 Selection of promoter and secretion signal sequences

The Lip2 secretion signal sequence has been previously successfully used in our laboratory and in others (Bulani *et al.*, 2012; Barth *et al.*, 2013; Govender, 2015). Given the extracellular activity of PAM, protein secretion is a key objective in the expression of recombinant PAM. Depending on the isoform, its intracellular localisation within the regulatory secretory granules, endosomes, and the trans-Golgi network must be accounted for in the expression system (Eipper *et al.*, 1993; Milgram *et al.*, 1997). The localisation of PAM may result in intracellular protein accumulation and aggregation, the formation of inclusion bodies, and potential protein-protein interactions, which could result in the cleavage of host proteins (Rajagopal *et al.*, 2010; Barth *et al.*, 2013). The requirement for PAM protein secretion limits the use of any of the source protein signal sequences. A secretion signal sequence on a suitable plasmid compatible with the expression system, such as Lip2, would enable secretion and aid the expression of PAM (Nicaud *et al.*, 2002).

3.1.3 Plasmids

A key element of genetic manipulation and cloning is the safe storage and multiplication of genetic sequences. The current laboratory standard for gene manipulation is bacterial plasmid systems due to their ease of use, low cost, cryopreservation ability, and error reading to minimise mutations. For this purpose, naturally occurring plasmids have been redesigned and refined by removing unnecessary elements and adding selection markers, such as antibiotic resistance. However, there is no comparable system for yeast cloning and expression, and an alternative to free, self-replicating plasmids must be used to avoid yeast plasmid instability and replication issues (Madzak and Beckerich, 2013).

3.1.3.1 pKOV410 plasmid features

As non-integrative plasmids are not generally stable in yeast, genomic integrative plasmids were developed to ensure the permanent retention of desired genes in yeast expression systems (Madzak and Beckerich, 2013). A commonly used plasmid for the transformation of *Yarrowia lipolytica* and the production of recombinant proteins is pKOV410 (Figure 3.4).

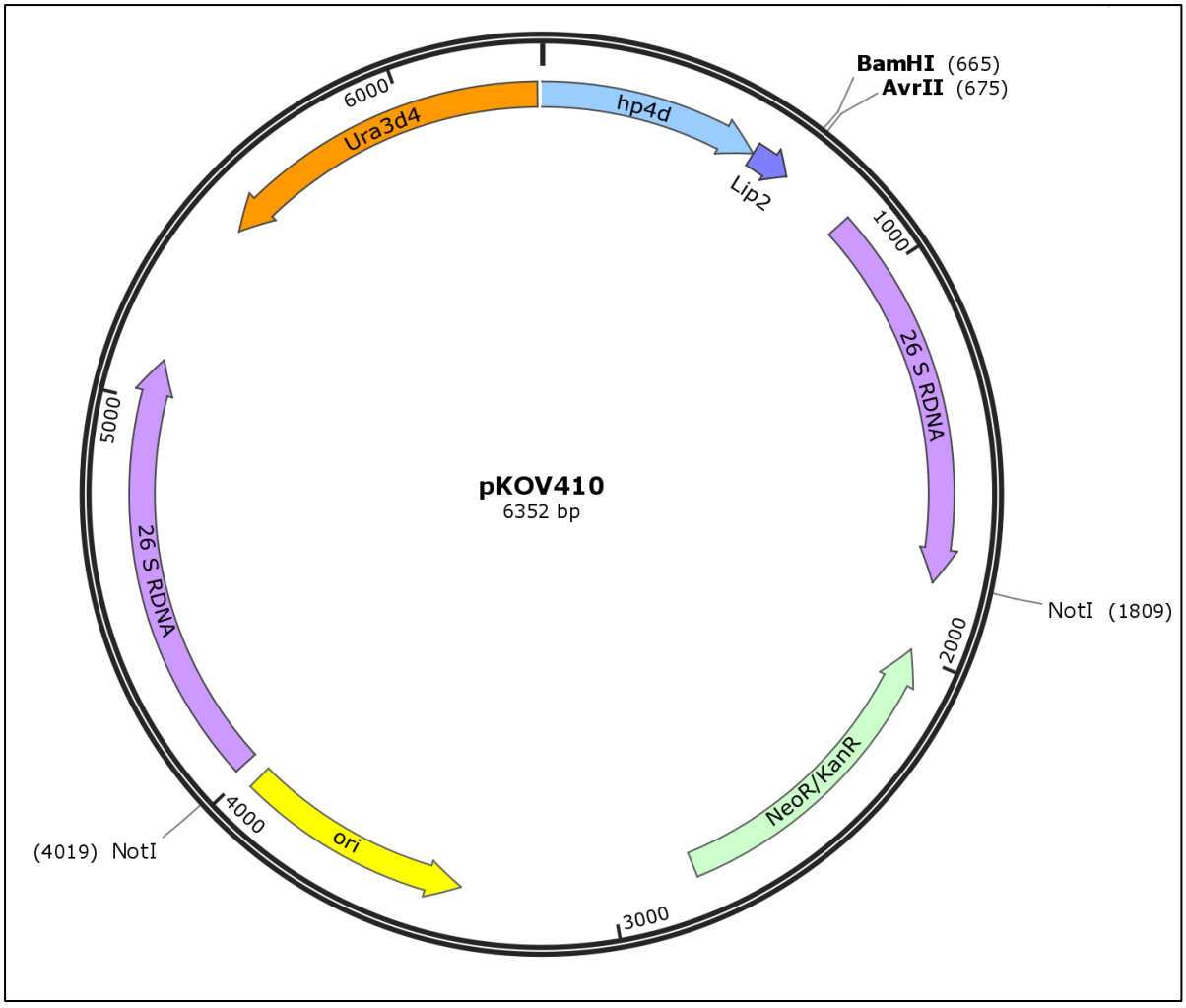


Figure 3.4: Plasmid map of LipSS pKOV410, annotated with primary features of a pKOV410 plasmid. Map created in SnapGene® Viewer with sequence data from Hofmeyer *et al.* (2014).

To achieve genomic integration, pKOV410 has two integrative regions surrounding a bacterial origin of replication (ori) and an antibiotic selection region (NeoR/KanR). The bacterial replication sequence is removed with NotI restriction endonuclease during the linearisation of the plasmid for integration, which increases its transformation efficiency and removes unwanted bacterial elements from the yeast cassette. Other key features include the hybrid promoter four direct (hp4d) region, a quadruple replication of a truncated sequence of the naturally occurring alkaline extracellular protease promoter region in *Y. lipolytica* (Blazeck *et al.*, 2011). The promoter allows for increased transcription of the downstream secretion signal peptide, Lip2, and the recombinant gene downstream, which is integrated with the BamHI and AvrII restriction sites (Madzak *et al.*, 2004; Hofmeyer *et al.*, 2014). To allow for the selection of transformed *Y. lipolytica* PolHI, which are Ura-, the pKOV410 plasmid includes a uracil gene (Ura3d4), which allows transformed cells to form colonies on uracil deficient media plates (Hofmeyer *et al.*, 2014).

3.1.4 Protein structural modelling

Prior to any protein expression experiment, it is crucial to confirm the identity and function of potential expression sequences, both of which are determined by protein structure (Yang and Zhang, 2015). Experiments based on blind structural modelling, such as Critical assessment of protein structure prediction (CASP), have found that template-based modelling methods have the highest level of accuracy in establishing the native structures of modelled but unpublished proteins. The use of previously established PAM structures is, therefore, a necessity to achieving PAM expression. However, to date, no models of the complete PAM protein that include the linker regions, signal sequences, and transmembrane domains have been reported.

Fortunately, there are several protein models of *H. sapiens* and *R. norvegicus* PHM and PAL, as well as *H. sapiens* DBM. The homology modelling enabled by these structures has been essential for identifying conserved residues, determining their function, and illuminating the catalytic activity of the active sites in both catalytic cores. Homology modelling and threading of the protein sequence can be used to assess the potential protein structures, and increase the probable identity of novel PAM variants and PHM, which catalyses the hydroxylation reaction essential for PAM activity.

3.1.4.1 Protein structural modelling with I-TASSER

The current winner of the past several CASP competitions, as of 2016, is I-TASSER from the Zhang lab at the University of Michigan (Yang and Zhang, 2015; Moult *et al.*, 2016). I-TASSER uses templates with the highest significance from its alignments of the query sequence against the known structures accessible to the database, which includes the published RSCB protein structures in the protein data bank (PDB) library (Berman *et al.*, 2002). The measure it uses for significance is the Z-score of the alignments, which is the difference between the raw and average scores in the unit of standard deviation. Therefore, the more minor deviation of the query sequence from the template, the closer the alignment of the structures. The degree of alignment is also reflected in the root mean square deviation (RMSD), which reflects the differences of a predicted model's structure from a known structure.

A key feature of the I-TASSER program, aside from the overall structural scores, is the coverage metric, which represents the coverage of the threading alignment of the query protein sequence against the template sequence. Coverage scores are calculated by dividing the number of aligned sequence residues by the length of the query protein (Yang and Zhang, 2015). This is important because it includes essential structural residues like disulphide bridges or residues

vital for secondary structures, but especially the conservation of critical residues in enzyme active sites. For PHM, these essential active site residues consist of the 5 histidines and one methionine that form the Cu_M and Cu_H copper-binding sites, vital for PHM activity (Siebert *et al.*, 2005). These key features can only form with the correct folding and topological positioning of the protein structure.

Protein ranking is especially insightful for researchers evaluating protein alignments, and these can be partly based on the TM-score of the structural alignment between the query protein structure and the known structures in the PDB library (Roy *et al.*, 2010). The RMSD between residues structurally aligned by TM-align can be used to measure structural alignment. While structural alignment is an important metric, the conservation of key sequences like active site residues is also essential. The level of identity is a vital metric for measuring conservation. Identity can be scored as the percentage sequence identity in the structurally aligned region between the query and template sequences. Finally, the relative size of the query and template sequences must be measured, as the enormous variety within the sequenced protein databases means that, given an increasingly long query sequence, the chance to generate spurious alignments and false identities increases. This mandates the use of shorter protein sequence queries to negate the generation of false-positive matches. A metric for coverage is therefore used, wherein coverage represents the coverage of the alignment by TM-align and is equal to the number of structurally aligned residues divided by the length of the query protein (Roy *et al.*, 2010).

COFACTOR and COACH modelling can generate confidence scores for predictions as part of the predicted modelling of an enzyme's potential ligand binding and function (Yang *et al.*, 2013). This is referred to as the C-score, and it represents the confidence score of the prediction of protein function and classification in I-TASSER (Zhang *et al.*, 2017). The C-scores ranges between 0 and 1, where a higher score indicates a more reliable prediction. The cluster size refers to the total number of templates in a cluster, indicating an increase in the likelihood of the enzyme being part of this enzyme class and thus sharing the same activity.

Homology modelling is particularly important for novel proteins with low sequence identity, as it can inform the researcher on potential protein structure, activity, and classification (Yang and Zhang, 2015; Zhang *et al.*, 2017). The protein expressionThe potential co-factors, metal-binding sites, and potential s requirements will also influence the gene design, as this must be compatible with both the host used for the molecular cloning and the host for the final expression.

3.1.5 Molecular cloning

The successful construction and maintenance of recombinant plasmids incorporating the gene of interest require a vector that can both store the plasmid and amplify the gene. One of the most commonly used cell lines for the transformation and propagation of plasmids is *E. coli* DH10B (Durfee *et al.*, 2008). This cell line has beneficial properties, including high transformation efficiency, capacity to accept and maintain large plasmids (>10 kbp), no methylation-dependent restriction systems, and the ability to be screened with LacZ-based α-complementation (Kovarik *et al.*, 2001). This strain started as a wild type K-12, with the MC1060 strain serving as the progenitor strain for the DH10B line (Durfee *et al.*, 2008). The DH10B line was selectively mutated several times, including *lacZΔM15*, which allowed for selective plasmid expression and blue-white colony screening (Grant *et al.*, 1990). Two other mutations were *recA1* and *endA1*, the former to improve clone stability by inhibiting the host homologous recombination system, and the latter to inactivate the host periplasmic DNA endonuclease, thereby improving the transfer of recombinant pDNA into the host cytoplasm (Grant *et al.*, 1990). Using an easy-to-use system such as DH10B enables the reliable amplification of plasmids, which aids the cloning of genes between vector systems.

3.1.6 Aim

To find, design, synthesise and transform PAM, PHM, and PAL-containing plasmid constructs for the transformation of *Y. lipolytica* cells for protein expression.

3.1.7 Objectives

- To analyse and compare several confirmed sequences of PAM, PHM, or PAL from phylogenetically distinct organisms.
- To perform protein searches and modelling of a novel PHM for fungal expression.
- To synthesise and transform PAM, PHM, or PAL constructs into a plasmid for *Y. lipolytica* transformation.

3.2 Methods and materials

3.2.1 PAM gene selection and identification

Only previously expressed and conclusively identified PAMs and their PHM and PAL were used for sequence searches, identified with basic local alignment search tool (BLAST) and DNA synthesis for initial expression studies. The closest phylogenetic relative of the fungal kingdom branch with an identified PAM and PHM, namely *C. elegans*, had its PAM AA sequence used to identify a potential fungal version of PHM or PAM. As some prokaryotes and algae express only PHM and not the whole PAM, the greatest likelihood of identifying a PAM or PHM amongst these kingdoms was achieved from using basic local alignment search tool for protein (BLASTP) with only the *C. elegans* PHM. The 300 AA-long PHM of *C. elegans* (>sp|P83388|1-300) was searched against the fungal database (uniprotkb_fungi (Protein) generated for BLAST on 8th May, 2017) using BLASTP (BLASTP 2.2.29+), matrix (blosum62). The threshold was set to 10, the results were not filtered, and the resultant sequences had gaps allowed.

To identify potential novel enzyme sequences such as fungal PAM, PHM, or PAL, the closest phylogenetic relative of fungi with a confirmed PAM gene was used as a search sequence. The closest candidate was *C. elegans*, as it had evidence of PAM at the protein level, but a recombinant version of its PAM gene had not been expressed. No other organism with verified PAM or PHM has been found that is more closely related to fungi than *C. elegans*, which forms part of the opisthokonts grouping that includes animals and fungi. Thus, the *C. elegans* PAM gene could have the closest sequence identity to fungal PAM. To circumvent the lack of suitable matches in this grouping, multiple sequence alignment (MSA) searches were performed with the PHM from *C. elegans* but limited to the fungal kingdom.

3.2.2 BLAST and MSA

The selection of known and putative PAM genes was performed using the AA sequence of *R. norvegicus* PAM, which was acquired from the European Bioinformatics Institute (http://www.ebi.ac.uk/). Both BLAST (basic local alignment search tool) and MSA searches were performed to search for similar protein and nucleotide sequences. Protein sequence features, domains, and plasmid maps were constructed in SnapGene® Viewer (version 5.0.7).

Multiple sequence alignments were performed utilising the CLUSTAL-Omega tool hosted by Uniport.org (Thompson *et al.*, 1994; Chenna *et al.*, 2003; Apweiler *et al.*, 2004; Apweiler, Bairoch and Wu, 2004). The default transition matrix was Gonnet with a gap opening penalty

of 6 bits, and a gap extension of 1 bit (Sievers and Higgins, 2018). The algorithm used in the Clustal-Omega tool was HHalign, which utilised CLUSTAL O (version 1.2.4) with default settings with its core alignment engine (Soding, 2005; Sievers and Higgins, 2018).

3.2.3 Gene synthesis and molecular cloning

Gene sequence modifications, such as the addition of purification tags, were performed *in silico* and analysed with SnapGene Viewer 4.1.6, and earlier versions. DNA was synthesised by Genscript Biotech Corporation, with codon optimisation for expression in *Y. lipolytica*. All methods were based on those from Green and Sambrook, 2012. Modifications made to these procedures and materials are detailed here. pDNA was purified from *E. coli* using several commercial plasmid purification products: Zymo Research, Zyppy Plasmid Miniprep kit, (Catalogue number: D4036); Bioline, Isolate II Plasmid mini kit (Catalog number: BIO-52055); Thermo Fisher Scientific, GeneJET Plasmid Miniprep Kit (Catalog number: K0502). The GeneJET kit was used predominantly for all pDNA purification from *E. coli*. The pDNA concentrations were quantified with a Nanodrop ND-1000 Spectrophotometer by obtaining absorbance measurements at 230, 260, and 280 nm.

Luria Broth (LB) was prepared as per Green and Sambrook, 2012, and Bertani, 1951, with 1.5% (w/v) agar added for plates. Antibiotics were added after autoclaving once the agar was 37°C or less (Bertani, 1951; Green and Sambrook, 2012). Post electroporation, a cell recovery solution of super optimal broth with catabolite repression (SOC) solution was used (Inoue, Nojima and Okayama, 1990).

Electroporation was performed according to Inoue, Nojima and Okayama (1990) and Dower, Miller and Ragsdale (1988). Typically, 80 µL of transformation-competent DH10B *E. coli* cells were combined with 1 µL of >20 ng/µL DNA for a pDNA and 2 µL of >2 ng/µL for ligated plasmids. The cell and pDNA solutions were added to a chilled (4˚C) sterile electroporation cuvette with a 0.2 mm gap (Sigma Aldrich; Z706086). The cuvette was electroporated using Bio-Rad Genepulser mark I at 1.6 kV, 200 Ω resistance, and 25 µFD capacitance. Immediately post electroporation, 1 mL of ice-cold SOC solution was added to the cuvette, and the entire solution was transferred to a sterile 20 mL MacCartney bottle for incubation at 37˚C for 1 hour at 200 rpm.

Post-electroporation, cultures were spread-plated onto separate LB plates with relevant antibiotic concentrations. After overnight growth at 37˚C, single colonies were picked for DNA polymerase chain reaction (PCR) amplification to confirm plasmid and insert sequence

identity. After successful identification, cultures were grown in the relevant antibiotic broth overnight at 37˚C, or to an OD600 of 6.0 and stored in 25 % (w/v) glycerol at -80˚C.

3.2.4 Agarose gel electrophoresis and construction of expression vectors

Agarose gel electrophoresis (AGE) was performed according to Green and Sambrook (2012), unless otherwise explicitly stated, with either 1% or a 0.8% (w/v) agarose in 40 mM Tris, 20 mM acetic acid, and 1 mM EDTA (TAE) buffer (pH 8.3). A molecular weight ladder of DNA, typically 5 µL of O'Generuler 1 kb (Thermo Scientific™; #SM1163) was included in the left lane. AGE gels used either 0.5 µg/mL of ethidium bromide or 0.5 µg/mL of Pronasafe dye (Laboratorios CONDA; CK130) for visualisation. The samples were separated electrophoretically with a constant 100 V or 120 V current using a BioRad PowerPac™ Basic power supply.

The codon-optimised genes acquired from Genscript (Hong Kong, China) were delivered in pUC57 plasmids. The genes were transformed into *E. coli* DH10B cells, as per section 3.2.3 for plasmid production and purification. The relevant genes were restricted from the host-vector with *Bam*HI and *Avr*II restriction endonucleases (Thermo Scientific™; ER0051 and ER1561), then ligated into pKOV410 with DNA ligases from either CapeBio (CapeBio: Express DNA Ligation Kit) or T4 DNA Ligase (Thermo Scientific™; EL0014).

3.2.5 PCR and DNA sequencing

The PCR amplification of PAM genes requires a DNA polymerase that can tolerate high guanine and cytosine content (GC-content) in DNA template sequences, as PAM sequences such as uPHM have a GC-content above 60%. Both Kapa 2G (Kapabiosystems; Cat no.: KK5009) and Ampliqon red (Ampliqon; Cat no.: A190303) kits were used in the present study for PCR. The PCR primers used are shown in Appendix II. The primary primer sets used for the pKOV410 plasmid were pKOV forward and reverse (pKOV f+r) and pKOV HF+HR. The pKOV HF+HR primers were designed with increased lengths of 58 and 64 bp, respectively, by adding 27 and 29 bp, respectively, to pKOV f+r. The additional sequence length raised the annealing temperature of the primer pairs to overcome the high GC-content in PAM genes and *Y. lipolytica* genomes. The primer sequences, their properties, and other associated PCR information is shown in Appendix II. Inqaba Biotech synthesised all primers. Primer annealing temperatures were calculated with Oligoanalyzer at https://eu.idtdna.com/calc/analyzer.

These cloning procedures were utilised with all transformations and ligations for all PAM, PHM, and PAL sequences in pKOV410 vectors. PCR utilising these synthesised PAM

sequences created additional PHM and PAL gene truncations from amplification of the relevant sequences, as listed for each template and primer as per Appendix II.

To establish if any mutations had resulted from the amplification and cloning steps, the purified pDNA of pKOV410 with ligated PAM gene or subunits, or the purified genomic DNA (gDNA) of the transformed *Y. lipolytica*, were sequenced by Inqaba Biotec™ using the pKOVf + pKOVr or the pKOVHF + pKOVHR primers. Sanger sequencing was performed using BrilliantDye™ Terminator v3.1 Cycle Sequencing on an ABI3500XL genetic analyser with a high-fidelity polymerase.

3.3 Results

3.3.1 Genomic and proteomic searches

Searches for a complete PAM DNA sequence and its corresponding protein sequences were initially performed. Since *Y. lipolytica* was the fungal (yeast) host for expression, BLAST searches were performed of the fungi, and specifically in the Ascomycota phylum. However, no PAM genetic homologues were found, nor has any PAM isoform been documented in fungi to date. Therefore, a known and previously expressed PAM was chosen for the subsequent experiments as a proof-of-concept for the expression of phylogenetically diverse PAMs, and finally for unknown PAM sequences.

To date, several isoforms of PAM, PHM, and PAL have been recombinantly expressed and purified from tissues of native PAM-expressing organisms. These isoforms have confirmed activity, and have genetic and proteomic identities. These sequences formed the basis for the search for novel PAM sequences in the present study. The protein structures of PHM and PAL have been crystallographically modelled, allowing for the resolution of conserved residues and their functions in the active sites of catalytic cores. Unfortunately, the complete bi-functional PAM crystal structure has yet to be modelled. However, using the confirmed identities of these PHM and PAL catalytic cores and their genetic and proteomic information, it is possible to establish key requirements for comparing known sequences and query the identity of novel sequences.

3.3.1.1 Phylogenetic comparison

With an established collection of sequences from the published genomes of model organisms containing PAM, the conserved AA residues vital to PAM functionality and the structure and activity of the subunits PHM and PAL were assessed. The AA sequence of *R. norvegicus* PAM was acquired from the European Bioinformatics Institute (EBI) (http://www.ebi.ac.uk/) and a BLAST search was performed to identify similar protein sequences. A list of matching protein sequences was constructed (Table 3.1).

Table 3.1: A list of the PAM or PHM species sequences for multiple sequence alignments using the canonical *Rattus norvegicus* PAM-1 isoform as reference.

Source Organism (Common name)	**UniProt accession number (Entry name)**	**Enzyme**	**Proof of identity**	**Expression hosts or source**
***Homo sapiens* (Human)**	P19021 (AMD_HUMAN)	PAM	Activity from recombinant protein	CHO cells
***Rattus norvegicus* (Rat)**	P14925 (AMD_RAT)	PAM	Activity from recombinant protein	CHO cells
***Xenopus laevis* (African clawed frog)**	P08478 (AMDA_XENLA)	PAM	Activity from recombinant protein	CHO cells
***Caenorhabditis elegans* (Nematode)**	P83388 (AMDL_CAEEL)	PAM	Activity from the host protein	Purified from host cells
***Conus bullatus* (Bubble cone snail)**	G8EWC9 (G8EWC9_CONBU)	PAM	Activity from recombinant protein	Insect cells
***Cyphellophora europaea* (Black rock yeast)**	W2RTG2 (W2RTG2_9EURO)	PHM	Predicted putative protein	Not tested
***Sorangium cellulosum* (Gram^{+} bacterium)**	A9GYG9 (A9GYG9_SORC5)	PHM	Predicted putative protein Novel and unconfirmed	Not tested
***Drosophila melanogaster* (Fruit fly)**	O01404 (PHM_DROME)	PHM	Activity from the host protein	Purified from host cells
***Chlamydomonas reinhardtii* (Green algae)**	A8J7Q6 (A8J7Q6_CHLRE)	PAM	Activity from recombinant protein	CHO cells

The sequences contained in Table 3.1 were used in a multiple sequence alignment (Figure 3.5). As the sequences used in the MSA do not contain PAL sequences, only the PHM isoforms of the aligned sequences are displayed. The overall level of identity for the PHM isoforms was relatively low, with 19 identical positions, 1.355% identity, and only 34 similar positions. The level of identity was improved when the alignment was restricted to sequences homologous to the PHM. Of primary interest was the identification of key conserved residues, which may be important for PHM activity. These are the metal-binding residues present in all sequences and are essential for binding two copper residues. As protein structures are conserved to a higher degree than the underlying sequence, the presence of structurally associated amino acids can indicate an underlying conserved structure in a protein sequence. Cysteine residues, which can form disulphide bonds essential for protein structure, are conserved across the PHM sequences, in the conserved canonical sequences and the candidate PHM sequences. The exception to this was a cysteine residue at the N-terminal of the PHM, another between the CuH and CuM binding sites, and cysteines downstream of, but not directly adjacent to, the final metal-binding site.

The sequences contained in Table 3.1 were used in the multiple sequence alignment, performed as per methods in section 3.2.2, as depicted in Figure 3.5.

Figure 3.5: Protein sequence alignment of confirmed and potential PHM and PAM sequences according to ClustalO analysis. The sequence is annotated with the following colours to describe structural elements from canonical PAM enzymes: Disulphide bonds in green; metal binding in violet; and beta-sheets in yellow. Additionally, the following symbols are for conserved or similar amino acid residues: conserved sequence residue (*); conservative mutations (:); semi-conservative mutations (.); and non-conservative mutations are not marked.

As the sequences used in the MSA in Figure 3.5 do not contain PAL sequences, only the PHM isoforms of the aligned sequences are displayed. For the PHM isoforms, the overall level of identity for all the query sequences was resultantly low with 19 identical positions, 1.355% identity and only 34 similar positions. The levels of identity were improved with the restriction of all the sequences homologous to the PHM. Of primary interest with this MSA was identifying the conservation of key conserved residues of interest for PHM activity. These are the metal-binding residues, which are present in all sequences in Figure 3.5. These are essential for binding the two copper residues, and mutagenesis studies that have substituted these residues produced inactive proteins or enzymes with reduced activity levels (Siebert *et al.*, 2005). As protein structures are conserved to a higher than sequence, the presence of structurally associated amino acids can indicate an underlying conserved structure of a protein sequence. The cysteine residues, which can form essential disulphide bonds for structure, are conserved across the PHM sequences, in the conserved canonical sequences and the candidate PHM sequences. The exceptions to these results were cysteine residues at the start of the PHM and another between the CuH and CuM binding sites and those downstream, but not directly adjacent, to the final metal-binding site.

3.3.1.2 Catalytic core modification and MSA

The signal peptide, propeptide, variable interdomain region, and sequences post-PHM were removed prior to MSA to maximise the sequence identity. These sequences were re-aligned with an unmodified *R. norvegicus* PAM included as a reference sequence to identify key residues and features (Figure 3.6).

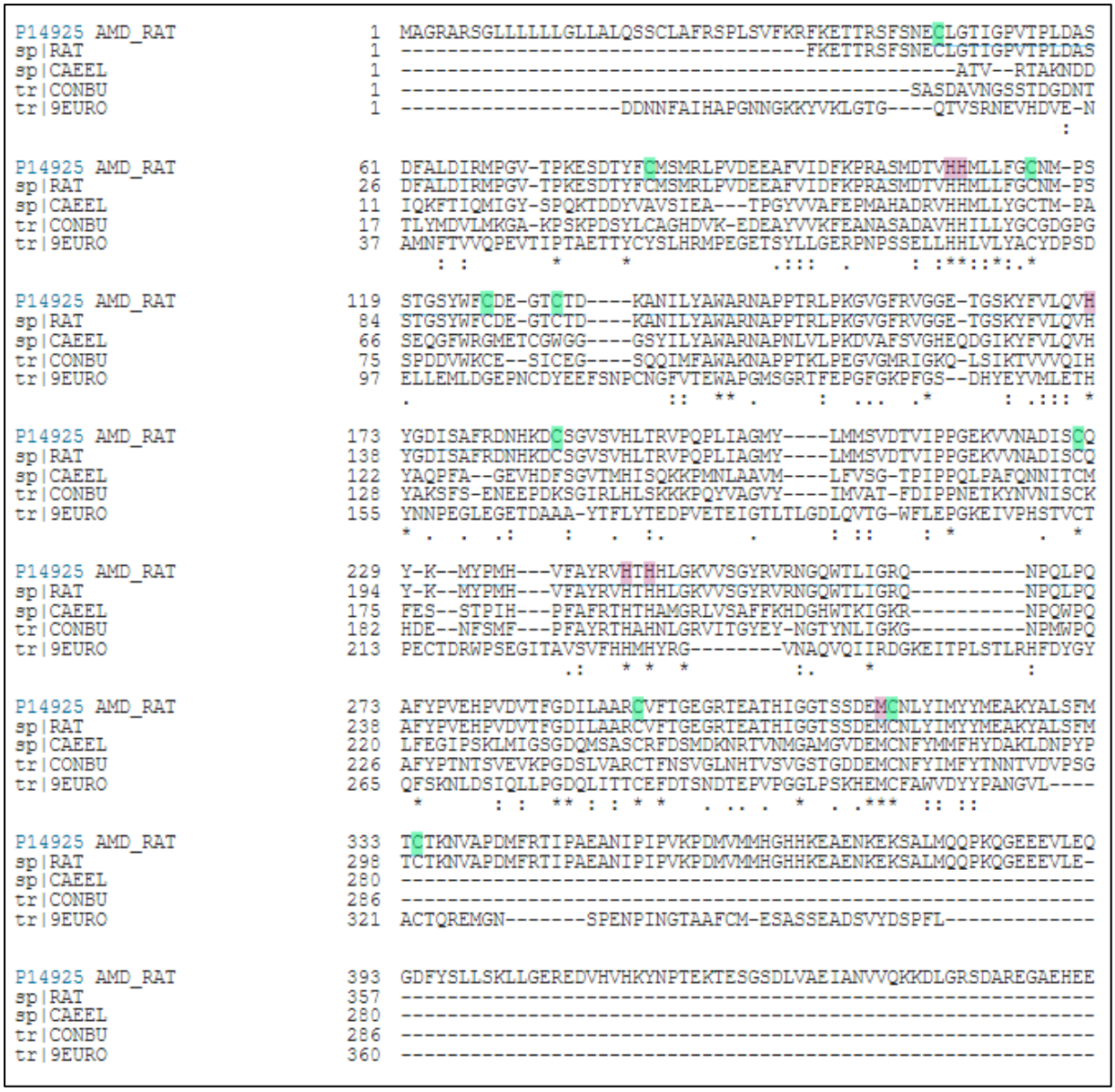


Figure 3.6: A multiple sequence alignment of the canonical *Rattus norvegicus* PAM with the truncated PHM regions of *R. norvegicus* (RAT), *Caenorhabditis elegans* (CAEEL), *Conus bullatus* (CONBU), and *Cyphellophora europaea* (9EURO) PHM sequences. Colour legend as per Figure 3.5.

All the metal-binding residues, essential for PHM activity, are conserved in all the aligned sequences (Figure 3.6). Based on the available data, the query PHM sequences chosen for synthesis and expression were *R. norvegicus*, *Conus bullatus*, *C. elegans*, and *C. europaea*, corresponding to the Uniprot accessions P14925 (AMD_RAT), P83388 (AMDL_CAEEL), G8EWC9 (G8EWC9_CONBU), and W2RTG2 (W2RTG2_9EURO).

W2RTG2_9EURO was included, although its identity as a PAM was not experimentally proven because it potentially contained a PHM subunit and was fungal in origin. The planned expression system for subsequent experiments was a yeast host, and W2RTG2_9EURO was

the nearest potential phylogenetic PHM match. An MSA of the genes synthesised for PHM expression was performed (Figure 3.7).

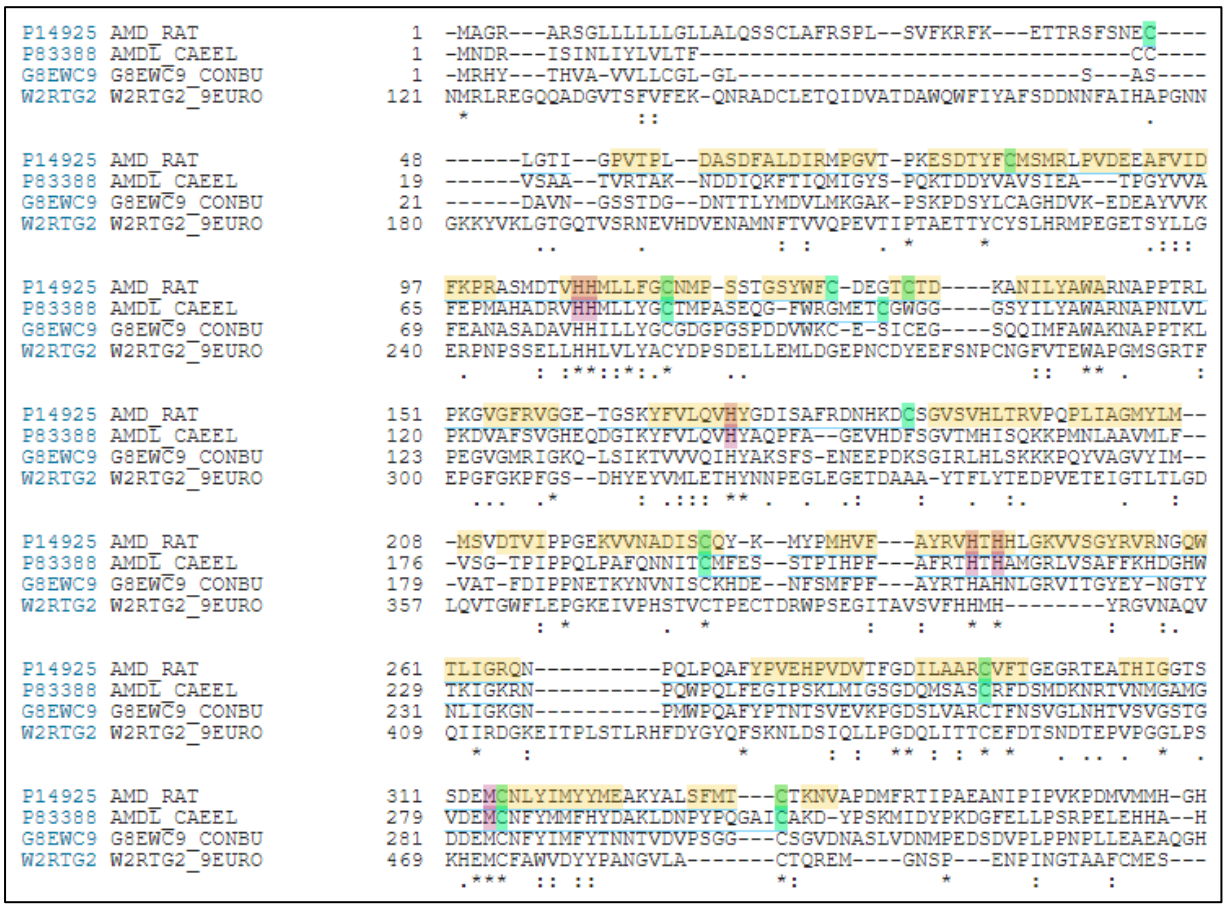


Figure 3.7: Multiple sequence alignment of *Rattus norvegicus*, *Caenorhabditis elegans*, *Conus bullatus*, and *Cyphellophora europaea* PHM sequences.

The aligned sequences had an identity score of 6.516%, with 64 similar positions and 26 identical positions. While the identity score was low, the identical positions crucially included the key metal-binding residues and the cysteine bonds, indicating the potential for a conserved structure. The low identity score offered no confidence of the *C. europaea* gene identity as a PHM. To assist with identification, an additional MSA was performed between the *C. europaea* PHM and its probable nearest phylogenetic match, *C. elegans* PHM (Figure 3.8).

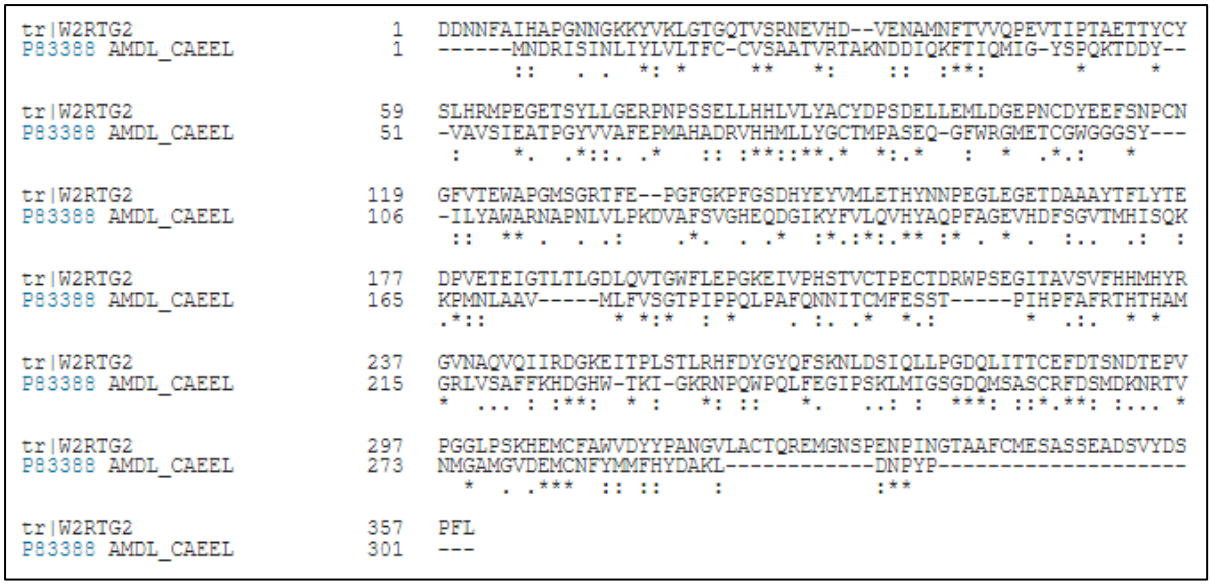


Figure 3.8: Multiple sequence alignment of *Caenorhabditis elegans* (AMDL_CAEEL) and *Cyphellophora europaea* (W2RTG2) uPHM sequences.

The protein sequence alignment between *C. elegans* and *C. europaea* uPHM sequences had 61 identical positions, an identity of 16.804%, and 99 similar positions. A higher degree of similarity was observed than in the MSA in Figure 3.8. In comparison, an MSA between the PHM sequences of *R. norvegicus* PHM and P83388 *C. elegans* PHM had 110 identical positions, an identity of 22%, and 92 similar positions. The *C. elegans* PHM is 300 AA, and the W2RTG2 *C. europaea* PHM catalytic core (uPHM) is 357 AA in length. The sequences aligned with an E-value of 3.2e0 and an overall score of 84.

3.3.2 Protein structure homology modelling

While the 16.8% sequence identity between ePHM and uPHM was below the 20% to 35% that is typically used as a cut off for pairwise sequence identity with protein MSAs, it is still possible that both were PHMs (Rost, 1999). It can be possible to obtain additional information on a novel protein sequence such as uPHM, especially in terms of structure and function, with protein structural modelling of the PHM sequence of the *C. europaea* PAM against known protein structures. Structural modelling provided insight into the properties and possible functions of the protein. Initial modelling of the protein was performed to determine the B-factor throughout the sequence (Figure 3.9).

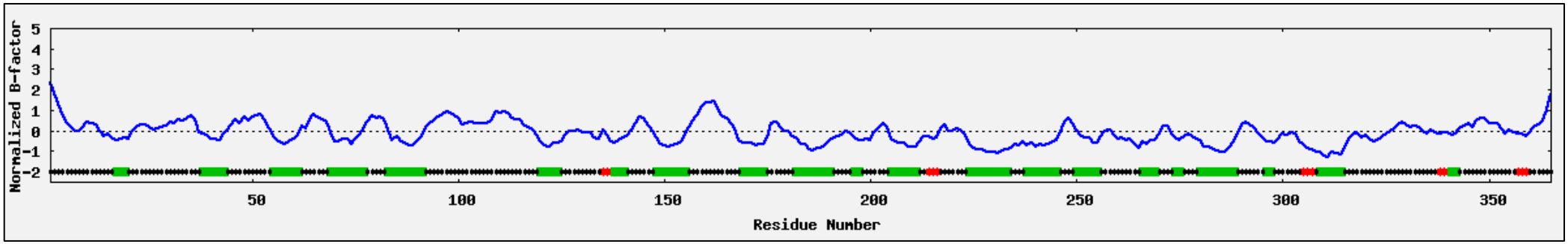


Figure 3.9: The predicted normalised B-factor for the uPHM query sequence as calculated by ResQ.

Figure Legend:

Normalized B-factor	Solid blue line
Helix	Red star
Strand	Green box
Coil	Black star

The normalised B-factor of the UPHM sequence correlates with lower scores for the residues that form predicted secondary structures, as is expected given the stabilising bonds and interactions that these structures can provide. The higher normalised B-factor scores for sequences between these secondary structure areas is also expected, as these areas require higher degrees of freedom of movement to align and fold the secondary structures into their required final structure.

3.3.2.1 Protein structural modelling

With the *de novo* protein models generated, a comparison of the theoretical models for uPHM to similar previously modelled protein structures was possible. Comparing the protein sequence coverage between a query sequence and templates within the PDB allows for the selection of templates sharing high degrees of sequence similarity. Normalized Z-scores were calculated by applying threading between the query and templates to establish the coverage between the sequences (Table 3.2).

Table 3.2: Ranked threading templates with I-TASSER from the PDB library for the uPHM query sequence.

Rank	PDB Hit	Source Organism, Enzyme name	Identity 1	Identity 2	Coverage	Normalized Z-Score
1	4zelA	*H.sapiens*, Dopamine beta-hydroxylase	0.21	0.27	0.99	2.31
2	1sdwA	*R.norvegicus*, PHM	0.19	0.20	0.85	4.00
3	1sdwA	*R.norvegicus*, PHM F	0.18	0.20	0.85	2.53
4	4zel	*Elizabethkingia meningoseptica*, PNGF-II	0.23	0.27	0.98	6.25
5	4zel	*Bacteroides fragilis*, PNGase F	0.23	0.27	0.98	4.57
6	1sdwA	*R.norvegicus*, PHM F	0.18	0.20	0.85	4.09

Using I-TASSER from the Zhang group at the University of Michigan, the uPHM protein sequence underwent a *de novo* folding routine. The resultant theoretical structure for uPHM was compared to a previously solved protein structure PDB file for a TM-score. The closest matches are shown in Table 3.3.

Table 3.3: Top 5 Identified structural analogues of uPHM in PDB identified by TM-align.

Rank	PDB Hit	Source Organism, Enzyme name	TM-score	RMSD	Identity	Coverage
1	4zelA	*H.sapiens*, Dopamine beta-hydroxylase	0.942	1.94	0.217	0.986
2	1sdwA	*R.norvegicus*, PHM	0.623	3.71	0.174	0.751
3	1pnf_	*Elizabethkingia miricola*, PNGase F	0.508	4.40	0.049	0.663
4	4r4xA	*Elizabethkingia meningoseptica*, PNGF-II	0.496	4.96	0.054	0.682
5	3ks7D	*Bacteroides fragilis*, PNGase F	0.493	4.89	0.048	0.674

Based on the uPHM structure formed from the I-TASSER folding algorithm utilising structures in the PDB server, it was possible to assign the protein structure and rank the key features of the matches. Interestingly, the top match for uPHM was for a human dopamine beta-hydroxylase, the copper-containing catalytic core of DBM, followed by the rat PHM. The order of ranking was because the DBM model had more coverage than the PHM model. The highest initial match for the uPHM sequence was for *R. norvegicus* DBM (Figure 3.10), but the second and third place ranks were for *R. norvegicus* PHM.

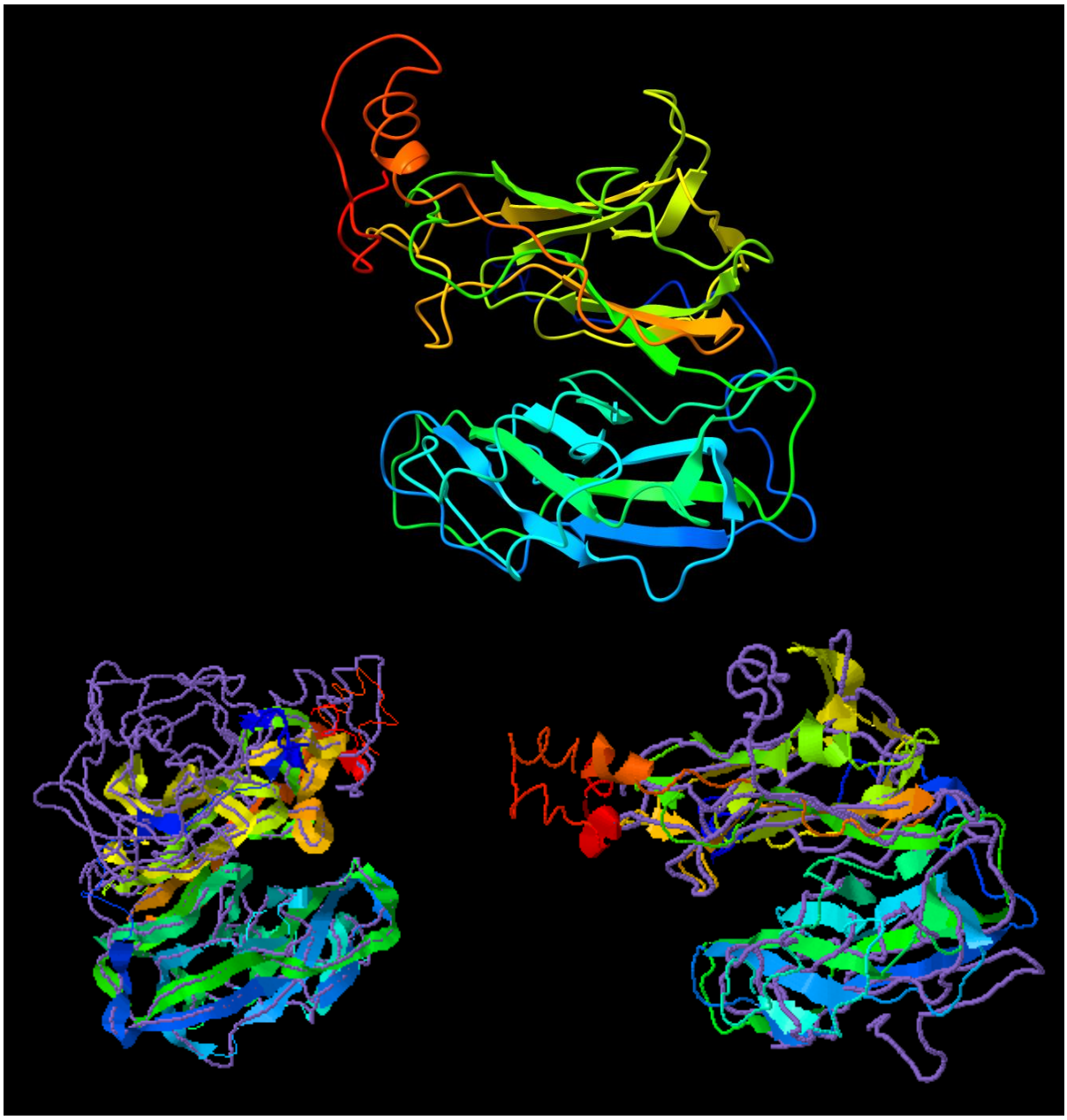


Figure 3.10: The peptide backbone for the predicted protein model for uPHM diagram (top), made with ChimeraX, and two overlaid ribbons (template structure) and lines (query sequence) diagram of the predicted uPHM model in cartoon form overlaid the structural analogue of *Rattus norvegicus* DBM (bottom left) and PHM, made with I-TASSER (bottom right).

The length of the uncovered structural analogue for the DBM modelling was greater than that of the PHM overlay (Figure 3.10). Thus, for the DBM, the sequence had greater coverage of the relatively small uPHM query structure, as confirmed by the higher coverage in Table 3.3

3.3.2.2 Predicted protein function and enzyme classification using COFACTOR and COACH

The potential co-factors, metal-binding sites, and potential substrates of the potential uPHM are of particular interest, as various metals have been recorded as increasing PHM activity. These attributes were assessed using the I-TASSER web server, which uses the COFACTOR webserver and consensus approach (COACH) to modelling.

Table 3.4: Top 5 Identified structural ligand binding site analogues of uPHM in the PDB identified using COFACTOR.

Rank	C-score	Cluster size	PDB Hit	Source Organism, Enzyme name	Rank	C-score
1	0.23	9	3micA	*R.norvegicus*, PHM	AZI	232,234,306
2	0.07	3	3mihA	*R.norvegicus*, PHM	AZI	245,246,247,248,278,280
3	0.04	2	1opmA	*R.norvegicus*, PHM	IYG	88,118,187,189,230,232,306,308,310
4	0.04	2	1OPMA	*R.norvegicus*, PHM	1OPMA01	32,33,34,35,36,37,38,39,40,172,173
5	0.04	2	1SDWA	*R.norvegicus*, PHM	1SDWA04	134,227,269

All the matches from the ligand-binding PBD files in Table 3.4 were for PHM models, which offer a high likelihood that the uPHM query sequence has PHM ligand binding affinity. Importantly, the ligand-binding sites H232, H234, and M306 were conserved spatially. These are likely to form the essential metal-binding residues identified by MSA and would theoretically form the Cu_M binding site. The substrates bound to rank 1 and 2 were azide molecules used in the crystallographic process. The 1OPMA and 1SDWA PDB files, in ranks

4 and 5, respectively, were both PHM models with peptide substrates bound in the active site (Prigge, 1997).

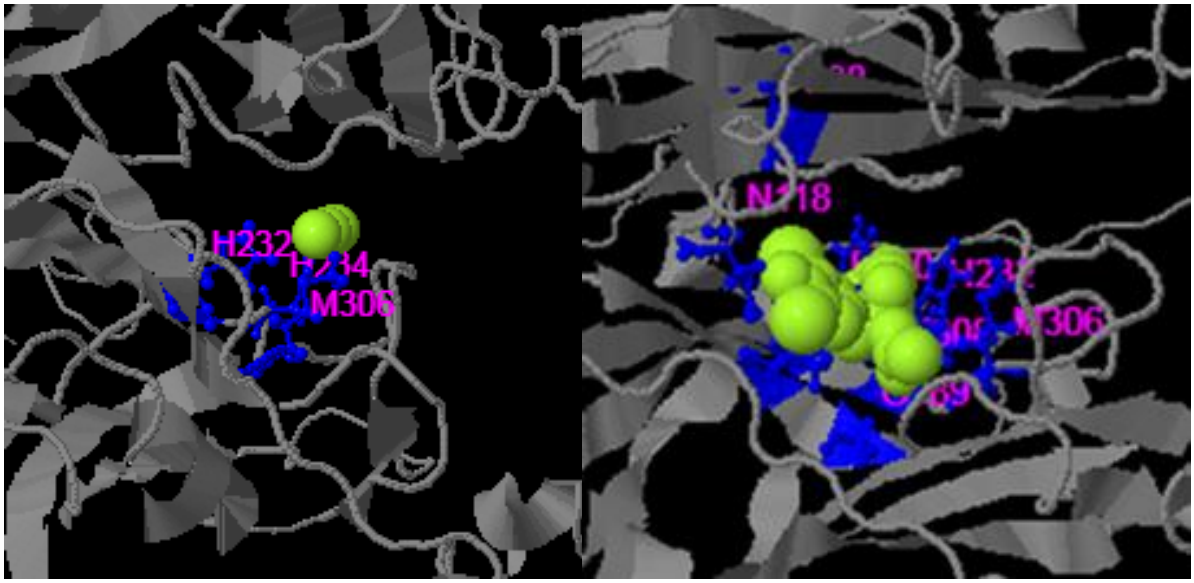


Figure 3.11: A ribbons diagram of the predicted protein structural model for uPHM overlaid with the *Rattus norvegicus* PHM Cu_H metal-binding site (left) from 3mic model, and the N-alpha-acetyl-3,5-diiodotyrosylglycine substrate from 1OPMA model (right), with substrates displayed in space fill style, made with I-TASSER.

The modelling predicted that the key conserved histidine residues of the uPHM query sequence were a match for the Cu_H metal-binding site, using the ligand-binding structural information from the PHM model 3mic (Figure 3.11). This conformation was favoured over that of the DBM model. However, using the ligand-binding information from the 1OPMA model, the PHM substrate isoleucine-valine-glycine (IVG), which is structurally similar to the common PAM substrate acetyl-L-tyrosyl-L-valyl-glycine (acYVG), indicating that uPHM has similar ligand binding potential to confirmed PHM models. In particular, the methionine (M306) and histidine (H232) binding copper-binding residues forming the potential Cu_M site is sterically proximate to the C_α on the glycine residue of IYG, which would be necessary for an active enzyme. Finally, after combining the structural, ligand, and additional modelling data, the algorithm assessed potential enzyme activity, which was crucial for identifying this uPHM query sequence.

3.3.2.3 Enzyme commission numbers and active sites

The primary purpose of this structural modelling was to attempt to determine the activity of candidate uPHM proteins. The uPHM query sequence was modelled against similar protein models in the PDB to determine potential EC number, and therefore determine its probable

activity (Table 3.5). The EC numbers and the top two PDB models are PHM, indicating a high likelihood that the uPHM query sequence has PHM activity.

Table 3.5: Top five potential enzyme classification activity numbers from uPHM analogues in PDB identified using COACH.

Rank	C-score	PDB Hit	Source Organism, Enzyme name	TM-score	RMSD	Identity	Coverage	EC Number
1	0.299	1sdwA	*R.norvegicus*, PHM	0.623	3.71	0.174	0.751	4.3.2.5 1.14.17.3
2	0.231	1yjlA	*R.norvegicus*, PHM	0.551	4.20	0.157	0.699	1.14.17.3 4.3.2.5
3	0.215	1pnfA	*E.miricola*, PNGase F	0.508	4.40	0.049	0.663	3.5.1.52
4	0.181	1qxpB	*R.norvegicus*, Cysteine Protease: m-calpain	0.388	6.05	0.033	0.592	3.4.22.53 3.4.22.52
5	0.175	1kfxL	*H.sapiens*, Cysteine protease: m-calpain	0.399	5.88	0.049	0.592	3.4.22.53

3.3.3 Gene design, synthesis and cloning

Using the codon optimised sequences for expression in *Y. lipolytica*, the DNA sequences were synthesised by Genscript, as shown in Appendix I. The sequences were shipped in pUC57 plasmids, excised using *Bam*HI and *Avr*II, and subsequently ligated into a pKOV410 plasmid (similarly digested), and the ligations transformed into *E. coli* DH10B. An example of the pKOV410 plasmid transformed with the *R. norvegicus PAM* is shown in Figure 3.12.

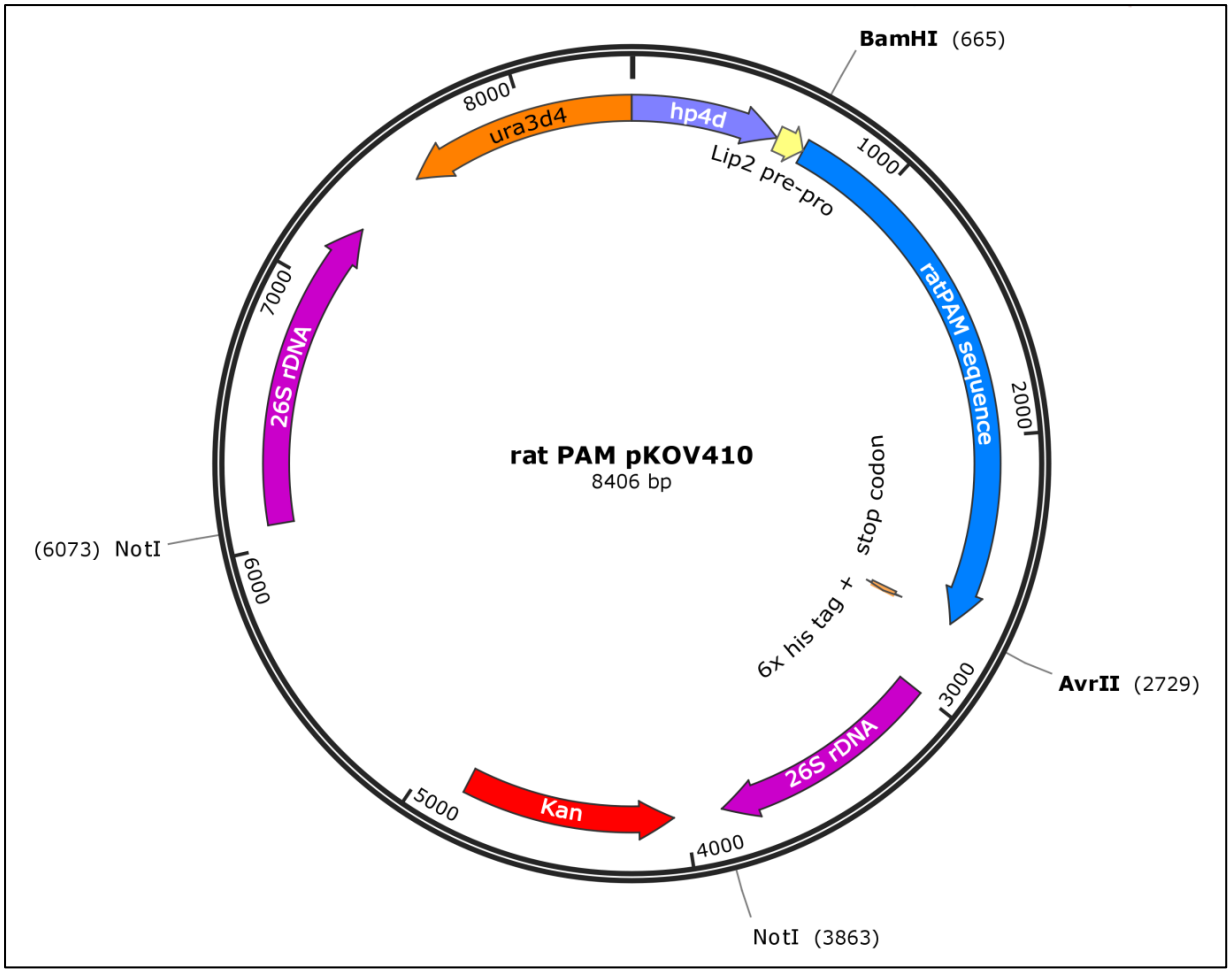


Figure 3.12: Plasmid map of pKOV410 with the synthetic recombinant *Rattus norvegicus* PAM gene inserted, annotated with primary features of pKOV plasmid.

From the original four PAM genes from *R. norvegicus* (AMD_RAT), *C. bullatus* (G8EWC9_CONBU), *C. elegans* (AMDL_CAEEL), and *C. europaea* (W2RTG2_9EURO), an additional ten truncations were performed on *C. bullatus*, *C. elegans*, and *C. europaea* PAM to generate truncated PAM, PHM, and PAL genes. For *C. bullatus* and *C. elegans*, PAM truncation with the removal of the native signal sequence was performed, resulting in cPAMss- and EPAMss-, respectively. A summary of all the truncated protein sequences created, along with their total base pairs and predicted protein masses, is shown in Table 3.6.

Table 3.6: Topological gene diagrams for PAM, PHM, and PAL constructed in the present study.

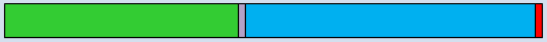
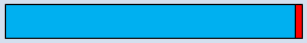
Protein abbreviation	Base pairs (bp)	Protein sequence pictogram	Predicted protein mass (kDa)
tPAM	2037		76
cPAM	2181		80
cPHM-	855		31
cPAMss-	2130		78
cPHM+	876		33
cPAL	1257		46
EPAM	2007		74
EPHM-	837		31
EPAMss-	1944		72
EPHM+	876		32
EPAL	1086		40
UPAM	2844		103
UPAM2	2373		85
UPHM	1095		41

Figure legend for protein topology in Table 3.6.

Colour	Sequence feature description
Yellow	Signal peptide
Green	PHM region
Violet	Linker region
Blue	PAL region
Brown	Predicted DOMON domain
Dark blue	Predicted cytochrome b561 domain
Orange	Predicted transmembrane and cytoplasmic domains
Red	6xHis tag and Stop codon

To establish whether a minimal truncated PHM would be expressed, a version based on the minimal *R. norvegicus* PHM expressed by Siebert *et al.* (2005) was generated. This PHM version had a peripheral sequence region including a conserved cysteine residue that forms a disulphide bond in several expressed crystallographic models removed and was constructed from *C. bullatus*, *C. elegans*, and *C. europaea*. These minimal truncated constructs were annotated cPHM- and EPHM-, respectively. In the present study, only a 6xHis tag was used, with no purification tag included, and with two *PHM* sequences to determine any deleterious effects from the expression of purification tags. To determine if the excluded sequences and purification tags removed from these PHMs would have any impact on the expression, complete versions with 6xHis tags were included, and these genes were annotated *cPHM+* and *EPHM+*. Finally, to express a functional PAL, the *PAL* sections were amplified and annotated *cPAL* and *EPAL*. The relevant sequences are shown in Appendix I. These truncations were performed with PCR and the relevant primers, using primers and parameters shown in Appendix II.

For the *C. europaea PAM* gene, no homologous *PAL* region was present in the native gene; therefore, the complete gene was synthesised and was annotated as *uPAM*. Two truncations were generated from *uPAM*, one of the probable PHMs with the DOMON domain (*uPAM2*), and one of the minimal PHM (*uPHM*). All three versions included a 6x His tag for purification. The sequences for all constructs and plasmids are shown in Appendix I.

3.3.4 Construction of PAM, PHM, and PAL pKOV410 plasmids

All the PAM, PHM, and PAL genes and their derivates were placed into pKOV410 vectors, which had been linearised with *Bam*HI and *Xma*JI (*Avr*II). PCR was performed to confirm the successful transformation of *E. coli* with pKOVHR and pKOVHR primers (Figure 3.13).

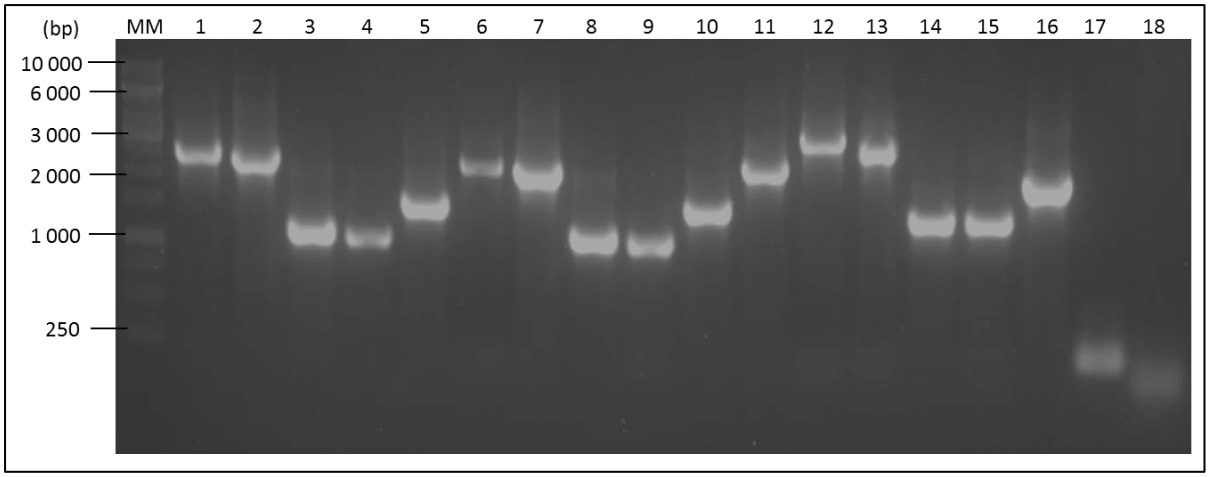


Figure 3.13: An agarose gel of PCR with all PAM, PHM, and PAL pKOV410 transformants and controls using the pKOVHF and pKOVHR primers. The numbered lanes contain the following PCR products:

MM. 2.5 µL of O'Generuler 1Kb DNA marker	10. EPHM- pKOV410 *E. coli* DH10B
1. tPAM pKOV410 *E. coli* DH10B	11. EPAL pKOV410 *E. coli* DH10B
2. cPAM pKOV410 *E. coli* DH10B	12. tPAM pKOV410 *E. coli* DH10B
3. cPAMss- pKOV410 *E. coli* DH10B	13. UPAM pKOV410 *E. coli* DH10B
4. cPHM+ pKOV410 *E. coli* DH10B	14. UPAM2 pKOV410 *E. coli* DH10B
5. cPHM- pKOV410 *E. coli* DH10B	15. UPHM pKOV410 *E. coli* DH10B
6. cPAL pKOV410 *E. coli* DH10B	16. mCherry pKOV410 *E. coli* DH10B
7. EPAM pKOV410 *E. coli* DH10B	17. Lip2 pKOV410 *E. coli* DH10B
8. EPAMss- pKOV410 *E. coli* DH10B	18. Negative control
9. EPHM+ pKOV410 *E. coli* DH10B	

The fidelity of the genes was ascertained by sequencing plasmid and genomic DNA during the cloning process. The yeast cassette had no mutations, and care was taken to avoid mutations caused by exposure to UV light during cloning or handling errors. The sequences of EPAM, cPAM and their related subunits were sequenced post-transformation into *E. coli* and ligated into pKOV410. No mutations were found in the sequenced DNA. The UPHM pKOV410 yeast cassette in the transformed *Y. lipolytica* gDNA was sequenced, and no mutations in the key sequence residues were found.

3.4 Discussion

A key issue during the identification of potential PAM genes, or their subunits, has been the lack of procedures and methodology to generate specific minimum criteria for sequence identification. Accurate identification is an inherent problem in heterologous gene expression, particularly those with high levels of diversity within a gene family, such as PAM and copper mono-oxygenases. Therefore, experimentally confirmed PAM identities are relied on, limiting expression attempts in non-conventional yeast expression hosts, like *Y. lipolytica*. The use of a suite of protein modelling and prediction programs, I-TASSER, is a pro-active and practical measure to prevent spurious expression attempts. Given the size and complexity of the PAM enzyme, it could be both costly and laborious.

PAM homologues or subunits have not been previously reported in fungi, which further complicates the expression of this critical enzyme at the industrial scale. A potential reason for the lack of homologues to date may be the inclusion of the *PAL* sequences along with the *PHM* for search queries as the entire *PAM* sequence. As the PAL is not essential to the eventual amidation of a peptide, due to the spontaneous cleavage of the C_{α}-hydroxylated peptides, this may have resulted in the apparent absence of PAM activity in the fungal kingdom (Attenborough *et al.*, 2012; Antošová and Sychrová, 2016). The identification of the potential *C. europaea* PHM reveals that previous protein searches may have been artificially limited by these overly broad sequence coverage and identity requirements.

The identification and further investigation into the other domains present on the *C. europaea* PHM gene would contribute to the elucidation of the amidation mechanism present in fungi but is not within the scope of the present study. A peptide amidation mechanism in fungi suggests a new class of amidated peptides produced from fungi, suggesting that new classes of therapeutic peptides could be possible. Further research into fungal PHM protein isoforms and *PHM* genes and fungal amidated peptides is highly recommended for the development of therapeutic peptides. Basic research into understanding this novel gene and multi-domain protein structure with PHM in fungi would also provide further insights into PAM functions.

According to the uPHM protein model generated using the I-TASSER system, the peptide sequences before the start of the β-sheets for the copper mono-oxygenase structure have only a single three peptide stretch and low normalised B-factor, implying an unorganised structure. The lack of secondary structure may be due to the sequence between the DOMON and PHM domains forming a linker region of sorts (Vendelboe *et al.*, 2016). This lack of secondary

structure could then provide a level of freedom for the folding of the nascent protein as it is translated and undergoes post-translational modification and folding.

Although protein homology modelling did not yield definitive protein identity or enzyme activity in the case of the *C. europaea* protein, based on these results, there is a high probability that the protein is a PHM. The protein constructs designed were successfully integrated into the host plasmid, pKOV410.

3.5 References

Antošová, Z. and Sychrová, H. (2016) 'Yeast hosts for the production of recombinant laccases: A review', *Molecular Biotechnology*, 58(2), pp. 93–116. doi: 10.1007/s12033-015-9910-1.

Apweiler, R., Bairoch, A., Wu, C., Barker, W., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M., Natale, D., O'Donovan, C., Redaschi, N. and Yeh, L. (2004) 'UniProt: the Universal Protein knowledgebase', *Nucleic Acids Research*, 32(90001), pp. 115D – 119. doi: 10.1093/nar/gkh131.

Apweiler, R., Bairoch, A. and Wu, C. H. (2004) 'Protein sequence databases', *Current Opinion in Chemical Biology*, 8(1), pp. 76–80. doi: 10.1016/j.cbpa.2003.12.004.

Aravind, L. (2001) 'DOMON: An ancient extracellular domain in dopamine beta-monooxygenase and other proteins.', *Trends in Biochemical Sciences*, 26(9), pp. 524–6. doi: 10.1016/s0968-0004(01)01924-7.

Attenborough, R. M. F., Hayward, D., Kitahara, M., Miller, D. and Ball, E. (2012) 'A "neural" enzyme in nonbilaterian animals and algae: preneural origins for peptidylglycine α-amidating monooxygenase.', *Molecular Biology and Evolution*, 29(10), pp. 3095–109. doi: 10.1093/molbev/mss114.

Barth, G. (2013) *Yarrowia lipolytica*, *Nonconventional Yeasts in Biotechnology*. Edited by G. Barth. Berlin, Heidelberg: Springer Berlin Heidelberg (Microbiology Monographs). doi: 10.1007/978-3-642-38583-4.

Bauman, A. T., Ralle, M. and Blackburn, N. J. (2007) 'Large scale production of the copper enzyme peptidylglycine monooxygenase using an automated bioreactor.', *Protein Expression and Purification*, 51(1), pp. 34–8. doi: 10.1016/j.pep.2006.06.016.

Berman, H. M., Battistuz, T., Bhat, T., Bluhm, W., Bourne, P., Burkhardt, K., Feng, Z., Gilliland, G., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J. and Zardecki, C. (2002) 'The protein data bank', *Acta Crystallographica Section D: Biological Crystallography*, 58(6 I), pp. 899–907. doi: 10.1107/S0907444902003451.

Bertani, G. (1951) 'Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*.', *Journal of Bacteriology*, 62(3), pp. 293–300. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=386127&tool=pmcentrez&rendertype=abstract.

Blazeck, J., Liu, L., Redden, H. and Alper, H. (2011) 'Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach', *Applied and Environmental Microbiology*, 77(22), pp. 7905–7914. doi: 10.1128/AEM.05763-11.

Bradbury, A. F., Finnie, M. D. A. and Smyth, D. G. (1982) 'Mechanism of C-terminal amide formation by pituitary enzymes', *Nature*, 298(5875), pp. 686–688. doi: 10.1038/298686a0.

Bradbury, A. F. and Smyth, D. G. (1987) ‘Enzyme-catalysed peptide amidation. Isolation of a stable intermediate formed by reaction of the amidating enzyme with an imino acid’, *European Journal of Biochemistry*, 169(3), pp. 579–584. doi: 10.1111/j.1432-1033.1987.tb13648.x.

Bulani, S. I., Moleleki, L., Albertyn, J. and Moleleki, N. (2012) ‘Development of a novel rDNA based plasmid for enhanced cell surface display on *Yarrowia lipolytica*.’, *AMB Express*, 2(1), p. 27. doi: 10.1186/2191-0855-2-27.

Çelik, E. and Çalık, P. (2012) ‘Production of recombinant proteins by yeast cells’, *Biotechnology Advances*, 30(5), pp. 1108–1118. doi: 10.1016/j.biotechadv.2011.09.011.

Chen, W., Lee, M. Jefcoate, C., Kim, S., Chen, F. and Yu, J. (2014) ‘Fungal cytochrome P450 monooxygenases: Their distribution, structure, functions, family expansion, and evolutionary origin’, *Genome Biology and Evolution*, 6(7), pp. 1620–1634. doi: 10.1093/gbe/evu132.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T., Higgins, D. and Thompson, J. (2003) ‘Multiple sequence alignment with the Clustal series of programs’, *Nucleic Acids Research*, 31(13), pp. 3497–3500. doi: 10.1093/nar/gkg500.

Chew, G. H., Galloway, L., McIntyre, N., Schroder, L., Richards, K., Miller, S., Wright, D. and Merkler, D. (2005) ‘Ubiquitin and ubiquitin-derived peptides as substrates for peptidylglycine α-amidating monooxygenase’, *FEBS Letters*, 579, pp. 4678–4684. doi: 10.1016/j.febslet.2005.06.089.

Cuomo, C., de Hoog, S., Gorbushina, A., Walker, B., Young, S. K., Zeng, Q., Gargeya, S., Fitzgerald, M., Haas, B., Abouelleil, A., Allen, A. W., Alvarado, L., Arachchi, H. M., Berlin, A. M., Chapman, S. B., Gainer-Dewar, J., Goldberg, J., Griggs, A., Gujja, S., Hansen, M., Howarth, C., Imamovic, A., Ireland, A., Larimer, J., McCowan, C., Murphy, C., Pearson, M., Poon, T. W., Priest, M., Roberts, A., Saif, S., Shea, T., Sisk, P., Sykes, S., Wortman, J., Nusbaum, C. and Birren, B. (2014) *The Genome Sequence of Phialophora europaea CBS 101466*. Available at: https://www.ncbi.nlm.nih.gov/protein/XP_008719583

Dower, W. J., Miller, J. F. and Ragsdale, C. W. (1988) ‘High efficiency transformation of *Escherichia coli* by high voltage electroporation’, *Nucleic Acids Research*, 16(13), pp. 6127–6145. doi: 10.1093/nar/16.13.6127.

Durfee, T., Nelson, R., Baldwin, S., Plunkett III G., Burland, V., Mau, B., Petrosino, J. F., Qin, X., Muzny, D. M., Ayele, M., Gibbs, R. A., Csörgő, B., Pósfai, G., Weinstock, G. M. and Blattner, F. R. (2008) ‘The complete genome sequence of *Escherichia coli* DH10B: Insights into the biology of a laboratory workhorse’, *Journal of Bacteriology*, 190(7), pp. 2597–2606. doi: 10.1128/JB.01695-07.

Eipper, B. A., Milgram, S., Husten, E., Yun, H. and Mains, R. (1993) ‘Peptidylglycine alpha-amidating monooxygenase: A multifunctional protein with catalytic, processing, and routing domains.’, *Protein Science : a Publication of the Protein Society*, 2(4), pp. 489–97. doi: 10.1002/pro.5560020401.

Eipper, B. A., Stoffers, D. A. and Mains, R. E. (1992) 'The biosynthesis of neuropeptides: peptide alpha-amidation.', *Annual Review of Neuroscience*, 15, pp. 57–85. doi: 10.1146/annurev.ne.15.030192.000421.

Glembotski, C. C. (1985) 'Further characterization of the peptidyl alpha-amidating enzyme in rat anterior pituitary secretory granules.', *Archives of Biochemistry and Biophysics*, 241(2), pp. 673–683.

Glembotski, C. C., Eipper, B. A. and Mains, R. E. (1984) 'Characterization of a peptide alpha-amidation activity from rat anterior pituitary.', *The Journal of Biological Chemistry*, 259(10), pp. 6385–6392. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6725255.

Govender, S. (2015) Elucidating the role of growth rate on the production of a fusion protein under regulation of the hp4d promoter by *Yarrowia lipolytica*. MSC thesis. University of Stellenbosch.

Grant, S., Jessee, J., Bloom, F. and Hanahan, D. (1990) 'Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants.', *Proceedings of the National Academy of Sciences*, 87(12), pp. 4645–4649. doi: 10.1073/pnas.87.12.4645.

Green, M. R. and Sambrook, J. (2012) *Molecular Cloning: A Laboratory Manual*. 4th edn, *Cold Spring Harbor Laboratory Press*. 4th edn. New York: John Inglis. Available at: www.cshlpress.org.

Hofmeyer, T., Bulani, S., Grzeschik, J., Krah, S., Glotzbach, B., Uth, C., Avrutina, O., Brecht, M., Göringer, H., Van Zyl, P. and Kolmar, H. (2014) 'Protein production in *Yarrowia lipolytica* via fusion to the secreted lipase Lip2p.', *Molecular Biotechnology*, 56(1), pp. 79–90. doi: 10.1007/s12033-013-9684-2.

Huang, Y., Huang, J. and Chen, Y. (2010) 'Alpha-helical cationic antimicrobial peptides: Relationships of structure and function.', *Protein & Cell*, 1(2), pp. 143–52. doi: 10.1007/s13238-010-0004-3.

Husten, E. J. and Eipper, B. A. (1991) 'The membrane-bound bifunctional peptidylglycine alpha-amidating monooxygenase protein. Exploration of its domain structure through limited proteolysis.', *The Journal of Biological Chemistry*, 266(26), pp. 17004–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1894599 (Accessed: 9 December 2013).

Illergård, K., Ardell, D. H. and Elofsson, A. (2009) 'Structure is three to ten times more conserved than sequence - A study of structural response in protein cores', *Proteins: Structure, Function, and Bioinformatics*, 77(3), pp. 499–508. doi: 10.1002/prot.22458.

Inoue, H., Nojima, H. and Okayama, H. (1990) 'High efficiency transformation of *Escherichia coli* with plasmids', *Gene*, 96(1), pp. 23–28. doi: 10.1016/0378-1119(90)90336-P.

Iyer, L. M., Anantharaman, V. and Aravind, L. (2007) 'The DOMON domains are involved in heme and sugar recognition', *Bioinformatics*, 23(20), pp. 2660–2664. doi: 10.1093/bioinformatics/btm411.

Katopodis, A. G., Ping, D. and May, S. W. (1990) ‘A novel enzyme from bovine neurointermediate pituitary catalyzes dealkylation of alpha-hydroxyglycine derivatives, thereby functioning sequentially with peptidylglycine alpha-amidating monooxygenase in peptide amidation’, *Biochemistry*, 29(26), pp. 6115–6120. doi: 10.1021/bi00478a001.

Kolhekar, A. S., Keutmann, H., Mains, R., Quon, A. and Eipper, B. (1997) ‘Peptidylglycine α-hydroxylating monooxygenase: Active site residues, disulfide linkages, and a two-domain model of the catalytic core’, *Biochemistry*, 36(36), pp. 10901–10909. doi: 10.1021/bi9708747.

Kolhekar, A. S. Bell, J., Shiozaki, E., Jin, L., Keutmann, H., Hand, T., Mains, R. and Eipper, B. (2002) ‘Essential features of the catalytic core of peptidyl-α-hydroxyglycine α-amidating lyase’, *Biochemistry*, 41(41), pp. 12384–12394. doi: 10.1021/bi0260280.

Kovarik, A., Matzke, M., Matske, A. and Koukalova, B. (2001) ‘Transposition of IS 10 from the host *Escherichia coli* genome to a plasmid may lead to cloning artefacts’, *Molecular Genetics and Genomics*, 266(2), pp. 216–222. doi: 10.1007/s004380100542.

Kumar, D., Blaby-Haas, C., Merchant, S., Mains, R., King, S. and Eipper, B. (2016) ‘Early eukaryotic origins for cilia-associated bioactive peptide-amidating activity.’, *Journal of Cell Science*, 129(5), pp. 943–56. doi: 10.1242/jcs.177410.

Kumar, D., Mains, R., Eipper, B. and King, S. (2019) ‘Ciliary and cytoskeletal functions of an ancient monooxygenase essential for bioactive amidated peptide synthesis’, *Cellular and Molecular Life Sciences*, 76(12), pp. 2329–2348. doi: 10.1007/s00018-019-03065-w.

Madzak, C. and Beckerich, J. (2013) ‘Heterologous protein expression and secretion in *Yarrowia lipolytica*’, in Barth, G. (ed.) *Microbiology Monographs*. Berlin, Heidelberg: Springer Berlin Heidelberg (Microbiology Monographs), pp. 1–76. doi: 10.1007/978-3-642-38583-4_1.

Madzak, C., Gaillardin, C. and Beckerich, J. (2004) ‘Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: A review’, *Journal of Biotechnology*, 109(1), pp. 63–81. Available at: http://www.sciencedirect.com/science/article/pii/S0168165604000264

El Meskini, R., Culotta, V., Mains, R. and Eipper, B. (2003) ‘Supplying copper to the cuproenzyme peptidylglycine alpha-amidating monooxygenase.’, *The Journal of biological chemistry*, 278(14), pp. 12278–84. doi: 10.1074/jbc.M211413200.

Milgram, S., Johnson, R. and Mains, R. (1992) ‘Expression of individual forms of peptidylglycine alpha-amidating monooxygenase in AtT-20 cells: Endoproteolytic processing and routing to secretory granules’, *The Journal of Cell Biology*, 117(4), pp. 717–728. doi: 10.1083/jcb.117.4.717.

Milgram, S., Kho, S., Martin, G., Mains, R. and Eipper, B. (1997) ‘Localization of integral membrane peptidylglycine α-amidating monooxygenase in neuroendocrine cells’, *Journal of Cell Science*, 706, pp. 695–706.

Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T. and Tramontano, A. (2016) 'Critical assessment of methods of protein structure prediction: Progress and new directions in round XI', *Proteins: Structure, Function, and Bioinformatics*, 84, pp. 4–14. doi: 10.1002/prot.25064.

Nicaud, J., Madzak, C., van den Broek, P., Gysler, C., Duboc, P., Niederberger, P. and Gaillardin C. (2002) 'Protein expression and secretion in the yeast *Yarrowia lipolytica*.', *FEMS Yeast Research*, 2(3), pp. 371–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12702287.

Peberdy, J. F. (1994) 'Protein secretion in filamentous fungi — Trying to understand a highly productive black box', *Trends in Biotechnology*, 12(2), pp. 50–57. doi: 10.1016/0167-7799(94)90100-7.

Prigge, S. T. (1997) 'Amidation of bioactive peptides: The structure of peptidylglycine α-hydroxylating monooxygenase', *Science*, 278(5341), pp. 1300–1305. doi: 10.1126/science.278.5341.1300.

Prigge, S. T. Mains, R., Eipper, B. and Amzel, L. (2000) 'New insights into copper monooxygenases and peptide amidation: Structure, mechanism and function', *Cellular and Molecular Life Sciences*, 57(8), pp. 1236–1259. doi: 10.1007/PL00000763.

Rajagopal, C., Stone, K., Mains, R. and Eipper, B. (2010) 'Secretion stimulates intramembrane proteolysis of a secretory granule membrane enzyme', *Journal of Biological Chemistry*, 285(45), pp. 34632–34642. doi: 10.1074/jbc.M110.145334.

Rost, B. (1999) 'Twilight zone of protein sequence alignments', *Protein Engineering, Design and Selection*, 12(2), pp. 85–94. doi: 10.1093/protein/12.2.85.

Roy, A., Kucukural, A. and Zhang, Y. (2010) 'I-TASSER: A unified platform for automated protein structure and function prediction', *Nature Protocols*, 5(4), pp. 725–738. doi: 10.1038/nprot.2010.5.

Sello, M. M., Jafta, N., Nelson, D., Chen, W., Yu, J., Parvez, M., Kgosiemang, I., Monyaki, R., Raselemane, S., Qhanya, L., Mthakathi, N., Mashele, S. and Syed, K. (2015) 'Diversity and evolution of cytochrome P450 monooxygenases in Oomycetes', *Scientific Reports*, 5, pp. 1–13. doi: 10.1038/srep11572.

Siebert, X., Eipper, B., Mains, R., Prigge, S., Blackburn, N. and Amzel, M. (2005) 'The catalytic copper of peptidylglycine α-hydroxylating monooxygenase also plays a critical structural role', *Biophysical Journal*, 89(5), pp. 3312–3319. doi: 10.1529/biophysj.105.066100.

Sievers, F. and Higgins, D. G. (2018) 'Clustal Omega for making accurate alignments of many protein sequences', *Protein Science*, 27(1), pp. 135–145. doi: 10.1002/pro.3290.

Soding, J. (2005) 'Protein homology detection by HMM-HMM comparison', *Bioinformatics*, 21(7), pp. 951–960. doi: 10.1093/bioinformatics/bti125.

Stoffers, D. A., Barthel-Rosa Green, C. and Eipper, B. A. (1989) 'Alternative mRNA splicing generates multiple forms of peptidyl-glycine α-amidating monooxygenase in rat atrium', *Proceedings of the National Academy of Sciences of the United States of America*, 86(2), pp. 735–739. doi: 10.1073/pnas.86.2.735.

Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) 'CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice', *Nucleic Acids Research*, 22(22), pp. 4673–4680. doi: 10.1093/nar/22.22.4673.

Vendelboe, T. V., Harris, P., Zhao, Y., Walter, T., Harlos, K., El Omari, K. and Christensen, H., (2016) 'The crystal structure of human dopamine β-hydroxylase at 2.9 Å resolution', *Science Advances*, 2(4), pp. 1–10. doi: 10.1126/sciadv.1500980.

Vishwanatha, K., Bäck, N., Mains, R. and Eipper, B. (2014) 'A histidine-rich linker region in peptidylglycine α-amidating monooxygenase has the properties of a pH sensor', *Journal of Biological Chemistry*, 289(18), pp. 12404–12420. doi: 10.1074/jbc.M113.545947.

Wand, G., Ney, R., Baylin, S., Eipper, B. and Mains, R. (1985) 'Characterization of a peptide alpha-amidation activity in human plasma and tissues', *Metabolism*, 34(11), pp. 1044–1052. Available at: http://www.sciencedirect.com/science/article/pii/0026049585900770

Yang, J., Roy, A. and Zhang, Y. (2013) 'Protein–ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment', *Bioinformatics*, 29(20), pp. 2588–2595. doi: 10.1093/bioinformatics/btt447.

Yang, J. and Zhang, Y. (2015) 'I-TASSER server: new development for protein structure and function predictions', *Nucleic Acids Research*, 43(W1), pp. W174–W181. doi: 10.1093/nar/gkv342.

Young, T. W., Wadeson, A., Glover, D., Quincey, R., Butlin, M. and Kamei, E. (1996) 'The extracellular acid protease gene of *Yarrowia lipolytica*: Sequence and pH-regulated transcription', *Microbiology*, 142(10), pp. 2913–2921. doi: 10.1099/13500872-142-10-2913.

Zhang, C., Freddolino, P. L. and Zhang, Y. (2017) 'COFACTOR: Improved protein function prediction by combining structure, sequence and protein–protein interaction information', *Nucleic Acids Research*, 45(W1), pp. W291–W299. doi: 10.1093/nar/gkx366.

Chapter 4: Protein expression, purification, and identification

4.1 Introduction

The present study aimed to facilitate the gram to kilogram-scale production of biologically synthesised alpha-amidated peptides, which requires post-translational modification enzymes at proportional scales. Fourteen *PAM, PHM,* or *PAL* constructs or truncations were made, as described in the preceding chapter, to allow for *in vitro* alpha amidation of glycine-extended C-terminal peptides.

Previous expression and secretion attempts of PAM or PHM in yeast have not produced industrially significant quantities (< 2.0 µg/L) (El Meskini *et al.*, 2003). Despite the limited quantities of PAM produced by CHO cells, they remain the primary commercial source of recombinant PAM to produce recombinantly expressed pharmaceutical peptides (Kim and Seong, 2001). Unfortunately, CHO cells incur significant production costs due to expensive reagents and specialised equipment while producing low levels of PAM (Jolivalt *et al.*, 2005). Compounding these drawbacks are the multiple-step purification processes involved with CHO cell culture, limiting its industrial applicability and increasing final product costs (El Meskini *et al.*, 2003; Jolivalt *et al.*, 2005).

Fungal expression systems do not have these inherent production disadvantages and produce a similar class of enzymes to PHM in the form of laccases, in inexpensive and industrial quantities (Alessandra *et al.*, 2010; Theerachat *et al.*, 2012). Laccases, which include copper monooxygenases, have a similar catalytic activity to PAM. Laccases use oxygen and produce water, with no harmful by-products or toxic solvents, forming part of an industry-wide movement towards green chemistry (Riva, 2006). The literature on recombinant protein expression of cuproenzymes, like laccases, by yeast hosts, provides a guide to avoiding many of the issues encountered in copper-containing enzyme expression.

4.1.1 Recombinant protein expression in yeast

Fungal systems are the predominant single-cell protein production system for generating industrial quantities of heterologous proteins that require post-translational modifications (Porro *et al.*, 2011; Liu, Redden and Alper, 2013). In particular, the yeast systems *Saccharomyces cerevisiae*, *Yarrowia lipolytica*, *Pichia pastoris*, and *Kluyveromyces lactis*

have been used to express laccases and other copper-containing enzymes primarily sourced from other fungi, as well as plants and bacteria (Antošová and Sychrová, 2016). PAM recombinant expression in *S. cerevisiae* is notoriously problematic, and a different choice of host was recommended to attempt recombinant PAM expression (El Meskini *et al.*, 2003). A promising candidate was the non-conventional yeast *Y. lipolytica*, which, since its initial identification in the early 1960s, has been developed as a platform for the recombinant expression of cuproenzymes (Jolivalt *et al.*, 2005; Hofmeyer *et al.*, 2014; Liu *et al.*, 2015).

4.1.2 *Yarrowia lipolytica* as a recombinant protein expression host

Y. lipolytica is a dimorphic yeast with generally-regarded-as-safe status for industrial processing of value-added chemicals, such as citric acid (Liu *et al.*,2015). Its potential for biotechnological applications was recognised early after its discovery due to its ability to grow on lipid-rich substrates, high production levels of inorganic acids, and prolific secretion of host proteins (Barth and Gaillardin, 1996). *Y. lipolytica's* metabolic capacity for simple media and industrial waste, such as glycerol and molasses, is an economic advantage, but its ability to secrete heterologous proteins from diverse genetic sources in concentrations as high as g/L is paramount (Dulermo and Nicaud, 2011; Hofmeyer *et al.*, 2014).

An advantage for protein secretion is conferred by *Y. lipolytica's* co-transcriptional pathway, which produces high levels of secreted proteins and its glycosylation moieties, which are similar to those of mammals (Nicaud *et al.*, 2002; Bankar *et al.*, 2009). Unlike mammalian protein expression systems, *Y. lipolytica*-expressed proteins do not contain problematic endotoxins, are not oncogenic, and do not contain viral DNA, which are inherent disadvantages of mammalian expression systems (Kim *et al.*, 2012; Groenewald *et al.*, 2014; Pardee *et al.*, 2016). *Y. lipolytica* is a non-pathogenic organism with high cell density fermentation, making it an ideal host for protein expression (Madzak *et al.*, 2004; Kahari *et al.*, 2007).

Despite these benefits, *Y. lipolytica* also has some limitations as a protein expression system. The high cell density achieved during the late stages of fermentation can be disadvantageous, as *Y. lipolytica* is an obligate aerobe and is sensitive to dissolved oxygen concentrations (Jolivalt *et al.*, 2005). Oxygen concentrations below 1.5 mg/L during growth conditions cause a transition in yeast cells to the mycelium cellular form, inhibiting protein expression (Bellou *et al.*, 2014). Furthermore, growth was inhibited at temperatures above 32°C, with 28°C being the optimal culture temperature for *Y. lipolytica* (Beopoulos *et al.*, 2009; Bellou *et al.*, 2014; Darvishi *et al.*, 2017). Despite these physical limitations, the previously documented successful

use of *Y. lipolytica* for the secretion of industrial-scale quantities of multi-copper laccases supports further investigation of this expression platform (Jolivalt *et al.*, 2005).

Multiple gene deletions have improved the ability to express proteins at industrial levels in Y. lipolytica strains, such as removing the secreted alkaline and acidic extracellular proteases (Barth and Gaillardin, 1996; Gonzalez-Lopez *et al.*, 2002). Other modifications to the *Y. lipolytica* genome include the disruption of the native uracil synthesis gene *URA3*, allowing for selection on uracil deficient media, with autotrophy restored from a vector containing the required genes, such as *ura3d4* (Kerscher *et al.*, 2001). The use of the *ura3d4* defective selection marker in the pKOV410 vector allowed for multiple integration events. A single copy of the *URA3* gene was insufficient for growth on uracil-deficient media (Bulani, 2013). Further strain improvements led to the Po series of strains becoming predominant for heterologous protein expression (Madzak and Beckerich, 2013). A genetic disadvantage of *Y. lipolytica* is their lack of endogenous *PAM* or *PHM* genes or homologous proteins (Liu and Alper, 2014). The absence of *PAM* or *PHM* in the yeast subphylum partly explains the failure of previous expression attempts with PAM in *S. cerevisiae* (El Meskini *et al.*, 2003; Attenborough *et al.*, 2012).

Finally, in comparative tests for their ability to secrete active forms of six fungal enzymes, *Y. lipolytica* had the best performance reproducibility for recombinant protein expression (Madzak and Beckerich, 2013). These advantages in protein expression in conjunction with the standardised pKOV410 plasmid system (detailed in section 3.2.1) have enabled a high-throughput method to screen for gene expression. Any gene constructs producing suitable levels of secreted proteins could be further improved by optimising yeast cultivation conditions.

4.1.3 Optimisation of yeast cultivation conditions

The fermentation of copper-containing enzymes, such as PHM, can be optimised around five primary cultivation conditions for yeast (Antošová and Sychrová, 2016). These include copper concentration, media composition, pH, temperature, and optimal cultivation times for enzyme activity and total protein yield (Antošová and Sychrová, 2016). Some important additional constraints are the enzyme activity conditions and protein stability, which must be factored into cultivation optimisation.

4.1.3.1 Composition of media for protein production

Copper concentrations from 10.0 µM Cu to as high as 1.2 mM of $CuSO_4$ have been used in media for the production of cuproenzymes by yeast (Theerachat *et al.*, 2012). Enzyme activity assays for PHM typically use a 1.0-5.0 µM Cu concentration, supporting the addition of 10 µM Cu into any cultivation media (Eipper *et al.*, 1995). High copper concentrations would be beneficial for *Y. lipolytica* cuproenzyme production and enzyme activity (El Meskini *et al.*, 2003; Jolivalt *et al.*, 2005). The ability of yeast to tolerate high copper concentrations during fermentation is an advantage compared to CHO cells for PAM production, which require the PAM to be reconstituted with copper in an incubation step after purification (Jolivalt *et al.*, 2005).

In yeasts cultivation, a standard medium is yeast peptone dextrose (YPD) (Green and Sambrook, 2012). Studies seeking to reduce media costs with *Y. lipolytica* have attempted to use minimal media, such as protein production broth (PPB), supplemented with industrial by-products, such as glycerol and molasses (Madzak and Beckerich, 2013; Hofmeyer *et al.*, 2014; Darvishi *et al.*, 2017). Sugar beet molasses-based media increased total enzyme activity compared to homogenous glucose-based media for laccase production (Darvishi *et al.*, 2017). Media composition also determines the maximum cell density and cellular morphology, both of which can result in a fivefold difference in recombinant protein activity and production in *Y. lipolytica* (Madzak *et al.*, 2005).

A closely related parameter to media composition is the media pH. It determines many of the cell growth characteristics and the late-stage cell morphology, which significantly alters protein secretion levels (Madzak *et al.*, 2005).

4.1.3.2 Media pH for protein production and stability

Changes in the media pH can cause cellular stress, and a low pH in the fermentation solution can inhibit protein production in *Y. lipolytica* (Madzak and Beckerich, 2013). During the early stages of yeast fermentation, the pH can reach below 3.0, potentially denaturing secreted proteins, damaging yeast cells, and limiting cell growth (Jolivalt *et al.*, 2005). The pH is typically controlled in automated bioreactors, but a buffered medium is not standard practice for protein expression studies conducted in shaker flasks (Jolivalt *et al.*, 2005). The inclusion of a buffer reagent can control the pH, but the buffer type and concentration require careful selection and optimisation, as these can affect cell morphology. For *Y. lipolytica*, a 20 mM citrate buffer generated higher cell density, and a yeast cell morphology, than the same media

with 200 mM citrate, which produced only mycelium and minimal protein secretion (Madzak *et al.*, 2005). Further complicating the selection of buffer reagent are any potential interactions between the recombinant enzyme and the buffer. Consequently, only buffers with minimal chelation properties have been used for PAM expression systems (Gaier *et al.*, 2013). The divalent metal chelation properties of common buffers such as Tris, phosphate, or acetate limits their use for PAM expression. As citrate production is possible at industrial scales with *Y. lipolytica* fermentation, citrate is tolerated during fermentation and is compatible with PAM activity assays (Bellou *et al.*, 2014).

4.1.3.3 Optimal cultivation time for recombinant protein production

The Lip2p secretion signal with the growth phase-dependant *h4pd* promoter has consistently been used to measure the secretion of recombinant proteins after the culture enters late cell growth stages (Jolivalt *et al.*, 2005; Bulani *et al.*, 2012; Hofmeyer *et al.*, 2014). The maximum cell density of *Y. lipolytica* in small scale shaker flasks (<100 mL) is reached at 48 to 72 hours (Hofmeyer *et al.*, 2014). With enzyme production and secretion induced by *h4pd,* the optimal time for enzyme activity cultivation and protein concentration could be studied when the cells are at maximum density (Hofmeyer *et al.*, 2014). These cultivation times would typically be between 48 h to 168 h, but prolonged incubation of the secreted protein in the culture medium could be detrimental to enzyme activity (Madzak, 2015). These considerations necessitate finding the optimal fermentation time for the best enzyme-specific activity and total activity in solution, both of which are a function of an enzyme's stability in solution, which can be influenced by the cultivation temperature used.

4.1.3.4 Optimal cultivation temperature of *Yarrowia lipolytica*

The translation and folding of *de novo* enzymes can be affected by the host's optimal cultivation temperature if it is not the same as that of the source organism of the recombinant protein. Expression temperatures below the source organism optimum are typically used, especially in yeast where the optimal fermentation temperatures are usually 20°C to 30°C (Antošová and Sychrová, 2016). For the four chosen PAM source genes, all the host organisms were mesophilic, presumably allowing their derived recombinant proteins to tolerate the lower optimal growth temperature of *Y. lipolytica* (Gale *et al.*, 1988; Madzak *et al.*, 2004; Ul-Hasan *et al.*, 2013). Fermentation temperatures also determine cell morphology and the resulting maximum cell density in *Y. lipolytica* (Beopoulos *et al.*, 2009; Bulani, 2013).

The interactions between these conditions and recombinant protein expression in yeast means there is no generalised optimal path for expression, and each protein studied requires individualised optimisation with *Y. lipolytica* (Madzak and Beckerich, 2013; Antošová and Sychrová, 2016).

4.1.4 Protein purification procedure for recombinant proteins

The design of a protein purification process is influenced by the protein's enzyme classification, structure, activity, and the need to rapidly transfer the enzyme into a stabilising solution at low temperature (Green and Sambrook, 2012; Hanke and Ottens, 2014). All buffers and purification systems that chelate divalent copper ions or irreversibly inhibit metalloprotease activity should be excluded from the purification process. A commonly used technique is immobilised metal chromatography (IMAC) with divalent-cation metal compatible buffers. The majority of purifications for recombinantly-expressed proteins and enzymes purified to homogeneity utilise IMAC, and add multiple sequential histidine tags (4 to 10 AA long) to the protein (Derewenda, 2004). Typically located at the C- or N-terminus of the protein, these tags have low immunogenicity and limited interference with the enzyme's tertiary structure, protein expression levels, or activity levels (Young *et al.*, 2012). Its ease of use and low cost make IMAC with a 6x histidine purification tag optimal for many purification procedures.

4.1.5 Aim

To optimise the protein production conditions and purification procedure to express recombinant PAM, PHM, and PAL in a *Y. lipolytica* system.

4.1.6 Objectives

- To screen transformed *Y. lipolytica* colonies for recombinant PAM, PHM, and PAL expression.
- To optimise fermentation conditions for the expression of maximum quantities of recombinant protein.
- To establish a set of purification procedures to purify uPHM to homogeneity.

4.2 Material and Methods

Due to the large culture volumes and resultant handling difficulties, a single shaker flask was used per variable tested for some studies instead of triplicate replicates.

4.2.1 *Yarrowia lipolytica* transformation and colony PCR screening

Y. lipolytica transformation was performed as per Bulani *et al.*, 2012 and Hofmeyer *et al.*, 2014. To identify which colonies were transformed with the pKOV410 cassette, a high throughput PCR method that partially removes contaminants was required (Liu *et al.*, 2000). The PCR screening with pDNA and *Escherichia coli* required modifications due to *Y. lipolytica's* high GC-content (49.0%) and yeast cell wall interference (Dujon *et al.*, 2004). Single colony scrapings were suspended in 20 µL of sterile 0.1% NaCl (Merck, SAAR5822320EM) to act as a temporary saline stock solution. A 2 µL sample of this saline solution was suspended in 20 µL of 10 mM NaOH and 10 mM Li-acetate (Sigma-Aldrich; 517992-100G), heated for 10 minutes at 60°C, then centrifuged for 5 min at 16 000 *g* at 4°C. A 2 µL sample of this partially purified gDNA supernatant provided a suitable template for colony screening PCR.

The PCR method used was described in section 3.2.5, with primers pKOV-HF+HR. To ensure strain monoculture, as multiple PAM, PHM, and PAL pKOV410 strains were generated simultaneously, primers unique to each gene were used. The sequences for these, and their relevant annealing and extension temperatures, are recorded in Appendix II. This procedure was amenable to high-throughput screening of multiple colonies for integrated pKOV410 gene cassettes. To establish the fidelity and sequence identity of the integrated cassette, gDNA was extracted using the glass bead method described by Chen, Beckerich, and Gaillardin, 1997.

4.2.2 Deep well fermentation screening for protein secretion

Multiple colonies were transformed with the gene of interest. However, the expression of the pKOV410 system can be variable depending on the integration site, which might not be an actively transcribed region (Bulani *et al.*, 2012). It was, therefore, necessary to select colonies with the highest secreted protein concentrations. Colonies with PCR-confirmed gene integration were cultured overnight on an orbital shaker at 28°C. A high throughput method with 2 mL YPD 2% (10 g/L yeast extract (Biolab, HG000BX6), 20 g/L peptone (Merck, SAAR4943300DN), 20 g/L D-glucose (Merck, SAAR2676020EM)), in 10 mL 24-well clear polypropylene deep well plates (Corning, Axygen®; P-DW-10ML-24-C) was used for

screening. Briefly, a 2 µL inoculum was suspended in 1.5 mL of YPD and incubated at 28°C with shaking at 200 rpm (Darvishi *et al.*, 2017). Brand names and catalogue numbers for media reagents are recorded in Appendix IV. Plates were sealed with an oxygen-permeable membrane to facilitate oxygen exchange (Corning, Axygen®; BF-400-S). When cultures had reached an OD600 nm of 3-6, sterile 80% (v/v) glycerol (Minema, G6020) was added at a final concentration of 25% (v/v), briefly vortexed, and frozen at -80°C.

Fresh cultures were started with inoculum solution of 0.1 at OD600 nm, from glycerol stocks, added to 1.5 mL YPD2 in 10 mL deep well plates and grown at 28°C for 72 h on an orbital shaker at 200 rpm. These cultures were transferred to 1.5 mL Eppendorf tubes, centrifuged for 20 min at 18 000 *g*, and the supernatant was separated from the cell pellet. The supernatant protein concentration was determined by Bradford's assay using known concentrations of bovine serum albumin (BSA) fraction V (Sigma-Aldrich; 10735086001) and Bradford's reagent (Sigma-Aldrich; B6916-500ML). A standard curve of BSA was generated from 0.1 to 1.2 mg/mL alongside culture samples, with a reference standard curve used for individual samples as per Appendix V. The two transformant colonies that produced the highest protein concentrations in solution after three days were used to establish a glycerol stock bank.

4.2.3 Shaker flask fermentation with *Yarrowia lipolytica*

The high dissolved oxygen concentrations requirements for production and high cell density in late growth stages with *Y. lipolytica* necessitated the use of a culture:shaker flask volume ratio below 20%, with constant orbital shaking at 200 rpm to ensure sufficient oxygen exchange. Monoseptic cultures were maintained with 50 µg/mL of Kanamycin, to which *Y. lipolytica* possesses natural resistance (Madzak, 2015). Cell morphology was observed with an Olympus microscope under 100x magnification. The methods used were previously described by Hofmeyer *et al.*, 2014.

4.2.3.1 Fermentation media and pH optimisation for recombinant protein production

The effects of media composition on maximum protein secretion and activity levels were studied. Media compositions were based on previous studies of YPD and protein production broth (PPB), as follows:

- YPD 1 % (Y1): 10 g/L yeast extract, 10 g/L tryptone, 10 g/L D-glucose (Madzak *et al.*, 2005).

- YPD 2 % (Y2): 10 g/L yeast extract, 20 g/L tryptone, 20 g/L D-glucose (Bohlin *et al.*, 2006).
- PPB (P): 20 g/L D-glucose, 1.32 g/L Yeast extract, 1.32 g/L NH_4Cl, 0.32 g/L KH_2PO_4, 0.132 g/L $MgSO_4 \cdot 7\ H_2O$; 0.015 g/L (0.1 mM) $CuSO_4$, 33 µg/L thiamine (Sigma-Aldrich; T4625-100G), and 20 mM citrate (Sigma-Aldrich; C8532-5 KG) buffer pH 6.0 (Madzak *et al.*, 2005).
- MPPB (mP): PPB with D-glucose replaced with 30 g/L sugar beet molasses (Grafschafter®; Goldsaft 450 g) to measure the effects of a change in carbon source optimised to bioreactor cultivation (Darvishi *et al.*, 2017).
- mPPBX (mPx): 110 g/L sugar beet molasses, 2.4 g/L ammonium chloride, 3.0 g/L, yeast extract g/L, 0.32 g/L KH_2PO_4, 0.132 g/L $MgSO_4 \cdot 7\ H_2O$, and 0.015 g/L $CuSO_4$ and 0.3 mg/L thiamine. This media was based on the optimal reactor media composition from Darvishi *et al.*, 2017, but was used for shaker flask culturing in the present study.

All media components were combined, dissolved, and sterilised by autoclaving at 121°C for 15 min. The trace metals and thiamine were made as concentrates, 0.2 µm filter-sterilised, and either frozen or added after inoculum. The pH of the media was measured and adjusted to pH 7.0 post-autoclave if a buffering reagent was present.

Y2 was made with 20 mM citrate for pH cultivation studies, then titrated to the required pH after autoclaving. The pH of individual shaker flasks was measured daily with 1 mL samples corrected to the desired pH with dilute acid (0.5 M H_2SO_4) or dilute base (1 M NH_4OH). The volumes used for pH adjustment were used to calculate the proportional volume of concentrated acid and base required for each shaker flask's culture volume. As the effects of pH and media reagents on secreted protein stability and activity can be inhibitory and denaturing, these contaminating reagents were rapidly removed by purification.

4.2.4 Protein purification procedure

The purification procedure was designed to have minimal effects on enzyme activity, increase enzyme stability, and was based on previous *Y. lipolytica* laccase expression and purification studies (Bulani *et al.*, 2012; Hofmeyer *et al.*, 2014).

4.2.4.1 Centrifugation

After fermentation, yeast cells were pelleted by centrifugation for 20 min at 16 000 *g* at 4°C, with the resulting supernatant designated as the crude protein solution. The centrifugation step was standard procedure for both the 1.5 mL culture samples and the pooled cultures after the

fermentation of the shaker flask. The crude and purification fractions were kept at 4°C for the duration of the purification procedure, and either stored overnight at 4°C or frozen at -20°C. Pooled supernatant had multiple additional purification steps applied to reduce working volume.

4.2.4.2 Tangential flow filtration for protein concentration and contaminant removal

Larger volumes (>20 mL) of samples were filter-sterilised using a 0.45 µm filter syringe. The crude solution was stabilised by the addition of 25 mM 3-(N-Morpholino) propanesulphonic acid (MOPS) (Sigma-Aldrich; M1254-250G) pH 7.5 and 10 mM NaCl. To reduce volume and remove contaminants below 10 000 Da MWCO, a tangential flow filtration (TFF) system (Pall corporation, OA010C12) in a Pall Minimate system was used, resulting in an average 10-fold reduction in crude volume.

4.2.4.3 Immobilised metal affinity chromatography of recombinant protein

For immobilised metal affinity chromatography (IMAC), the primary resin system used was nitrilotriacetic acid (NTA) residue-coated magnetic microspheres (MagReSyn®; MR-NTA005) (Gerber *et al.*, 2015). The buffers used were the wash, equilibration, binding buffer (WEB): 50 mM MOPS pH 7.5, 40 mM imidazole (Sigma-Aldrich; I239-100G), 1 M NaCl, and the elution buffer (EB): 50 mM MOPS pH 7.5, 500 mM imidazole, 500 mM NaCl. The primary buffer reagent was altered from the manufacturer recommendation of phosphate to MOPS, as the former can chelate metal. Before sample processing, triple strength WEB was added to the concentrated TFF retentate to allow for binding. Samples were processed in 15 mL falcon tubes at 4°C, according to the manufacturer instructions (Gerber *et al.*, 2015). The fractions containing uPHM were selectively eluted with 500 mM imidazole in the EB, with elution fractions of 5 mL.

4.2.4.4 Dialysis and protein storage solution

For protein stability, desalting and buffer exchange is an essential final step, and was performed with 10 000 Da molecular weight cut-off (MWCO) centrifugal membrane concentrator tubes (15 mL Vivaspin® 6 centrifugal concentrators; Sartorius; VS0601). Samples were kept chilled at 4°C and were centrifuged at 16 000 *g* until the desired volume was achieved, and a quadruple-strength protein storage solution (PSS) was added. The final purified protein fractions contained 20% glycerol, 20 mM NaCl, 1.0 µM CuCl, 25 mM MOPS (pH 7.5), for storage at -20°C. The protein purification fractions were analysed using electrophoresis to assess purity.

4.2.5 Electrophoresis methods

Agarose gel electrophoresis was performed according to section 2.2.3.6. Protein samples were separated and visualised with sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) to determine protein fraction mass, quantity, and purity. These methods were modified from the Mini-Protein® 3 cell protocol and polyacrylamide gel electrophoresis and detection guide (Laemmli, 1970; Bio-Rad Laboratories, 2012).

4.2.5.1 Discontinuous denaturing SDS-PAGE

Protein samples were diluted with triple-strength sodium dodecyl sulphate (SDS) reducing buffer consisting of 50 mM Tris HCl (pH 6.8) (Sigma Aldrich; 10812846001), 40% (w/v) glycerol, 3% (w/v) SDS (Sigma-Aldrich; L3771-1KG), 0.14% (w/v) bromophenol blue (Sigma-Aldrich; B0126-25G), 5% (w/v) β-mercaptoethanol (Sigma Aldrich; M6250-100ML), then heated to 95°C for 5 min (Labnet AccuBlock™ digital dry bath). Discontinuous gels composed of 12% resolving (0.375 M Tris-HCl pH 8.8 and 0.1% (w/v) SDS) and 4% stacking gels (0.125 M Tris-HCl pH 6.8 and 0.1% (w/v) SDS) diluted from degassed 30% acrylamide/bis (Bio-Rad; 1610158) were hand-poured. Gel polymerisation was initiated with the addition of 0.1% (v/v) ammonium persulphate (APS) and N,N,N',N'-tetramethylethylenediamine (TEMED) according to the Mini-Protein® 3 cell protocol (Bio-Rad Laboratories, 2012).

Electrophoresis was performed using a Mini-Protein® 3 cell tank with SDS running buffer (25 mM Tris base, 192 mM glycine, 1% (w/v) SDS). Typically 15 µL of protein sample was added and separated by discontinuous gel electrophoresis (Laemmli, 1970). A constant voltage of 150 V (Bio-Rad Power Pac™ Basic) was applied until the dye front migrated within 1 cm of the gel bottom. The gels were stained with standard Coomassie staining or washed twice with de-ionised water before staining with colloidal Coomassie stain (Dyballa and Metzger, 2009). The gels were then destained with either a destaining solution or with water for the standard Coomassie stain and the colloidal Coomassie stain, respectively. Gels were photographed using a gel documentation system (Biorad laboratories; ChemiDoc™ MP Imaging system 170-8280).

4.3 Results

Yeast expression plasmids containing *PAM*, *PHM*, and *PAL* sequences were transformed into *Y. lipolytica* Po1h hosts. The transformed cells were selected with selective media deficient in uracil. DNA sequence integration was confirmed by performing yeast colony PCR, according to the method described in section 3.2.1. This method required preparative treatment of the gDNA before PCR colony screening.

4.3.1 *Yarrowia lipolytica* colony PCR screening

PCR was performed with colonies 0.5 mm or larger grown on selective media to confirm the integration of PAM, PHM, or PAL sequences. The PCR was performed with pKOV: HF+HR primers as per section 3.2.1 (Figure 4.1).

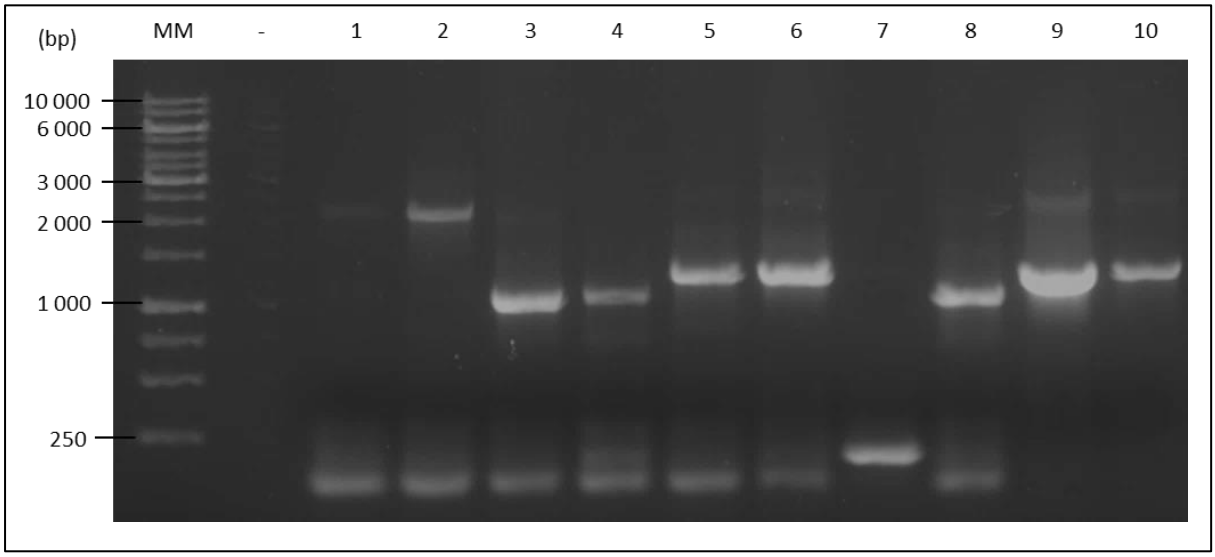


Figure 4.1: Agarose gel electrophoresis of colony screening PCR for *Caenorhabditis elegans* and *Cyphellophora europaea* PAM, PHM, and PAL pKOV410 *Yarrowia lipolytica* transformants, with *Escherichia coli* positive controls.

Figure legend:

MM. 4 µL of O'Generuler 1Kb DNA marker	6. UPHM pKOV410 *Y. lipolytica* Po1h
1. EPAM pKOV410 *Y. lipolytica* Po1h	7. Crem pKOV410 *Y. lipolytica* Po1h
2. EPAMss- pKOV410 *Y. lipolytica* Po1h	8. mCherry pKOV410 *Y. lipolytica* Po1h
3. EPHM+ pKOV410 *Y. lipolytica* Po1h	9. UPHM pKOV410 *E. coli* DH-10B
4. EPHM+ pKOV410 *Y. lipolytica* Po1h	10. mCherry pKOV410 *E. coli* DH-10B
5. EPAL pKOV410 *Y. lipolytica* Po1h	

The colony screening PCR confirmed the presence of inserted recombinant DNA in the *Y. lipolytica* Po1h colonies, as seen in lane 6 of Figure 4.1. The number of PCR cycles provided

DNA amplification similar to the transformed pDNA in *E. coli* DH-10B colonies in lane 9. This methodology allowed for up to 10's of colonies from multiple strains to be screened simultaneously. Confirmed colonies were then cultured in deep well plates, as per section 3.2.2, to identify colonies with detectable recombinant protein secretion using Bradford's assay. Only the uPHM culture secreted recombinant proteins in detectable quantities, and this isolate was therefore selected for further studies.

4.3.2 uPHM protein expression analysis

Fermentations of the uPHM isolate were scaled up to shaker flask volumes as per section 4.2.3. After fermentation, 500 mL shaker flask cultures were centrifuged for 20 min at 16 000 *g* at 4°C in JM-10 Beckman Coulter centrifuges, with the supernatant of the culture designated as the crude protein solution. The selective production and secretion of recombinant proteins were analysed by comparison to a Lip2-pKOV410 culture only secreting the Lip2 signal sequence. SDS-PAGE analysis of the cell-free supernatant for six days out of eight days of fermentation was performed (Figure 4.2).

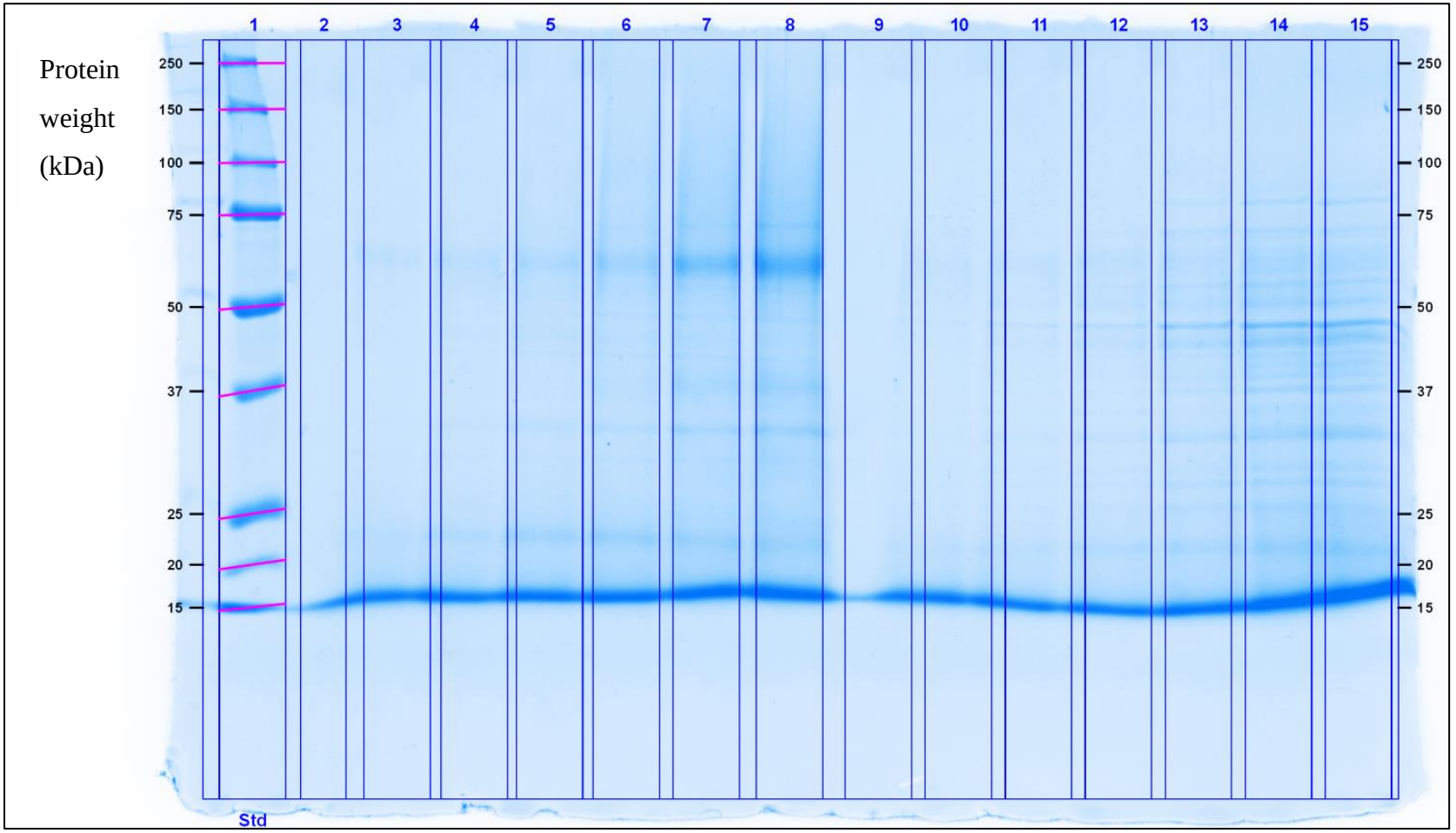


Figure 4.2: SDS-PAGE of the supernatant protein isolated from uPHM- and Lip2- pKOV *Yarrowia lipolytica* cultures grown for 1-4, 6, or 8 days in YPD 2%.
Figure legend:

Lane	Contents	Lane	Contents	Lane	Contents	Lane	Contents
1	Precision Plus Protein marker (2.5 µL)	5	uPHM supernatant day 3	9	Blank lane	13	Lip2 supernatant day 4
2	Blank lane	6	uPHM supernatant day 4	10	Lip2 supernatant day 1	14	Lip2 supernatant day 6
3	uPHM supernatant day 1	7	uPHM supernatant day 6	11	Lip2 supernatant day 2	15	Lip2 supernatant day 8
4	uPHM supernatant day 2	8	uPHM supernatant day 8	12	Lip2 supernatant day 3		

The SDS PAGE analysis confirmed the expression of a novel protein fraction at 55 kDa in lanes 3-8 (Figure 4.2). This potential recombinant protein had a higher mass than predicted at 40.7 kDa. The protein fractions of the negative control strain produced only Lip2 secretion signal and had no significant protein fraction of similar mass to the uPHM strain. The expression and accumulation of this novel protein were visually detectable from day two onwards, as expected with the *h4pd* promoter.

4.3.3 Purification of uPHM and analysis

The crude protein isolated from uPHM samples underwent IMAC protein purification as described in section 4.2.4.3. Samples of each fraction were subjected to SDS-PAGE analysis (Figure 4.3).

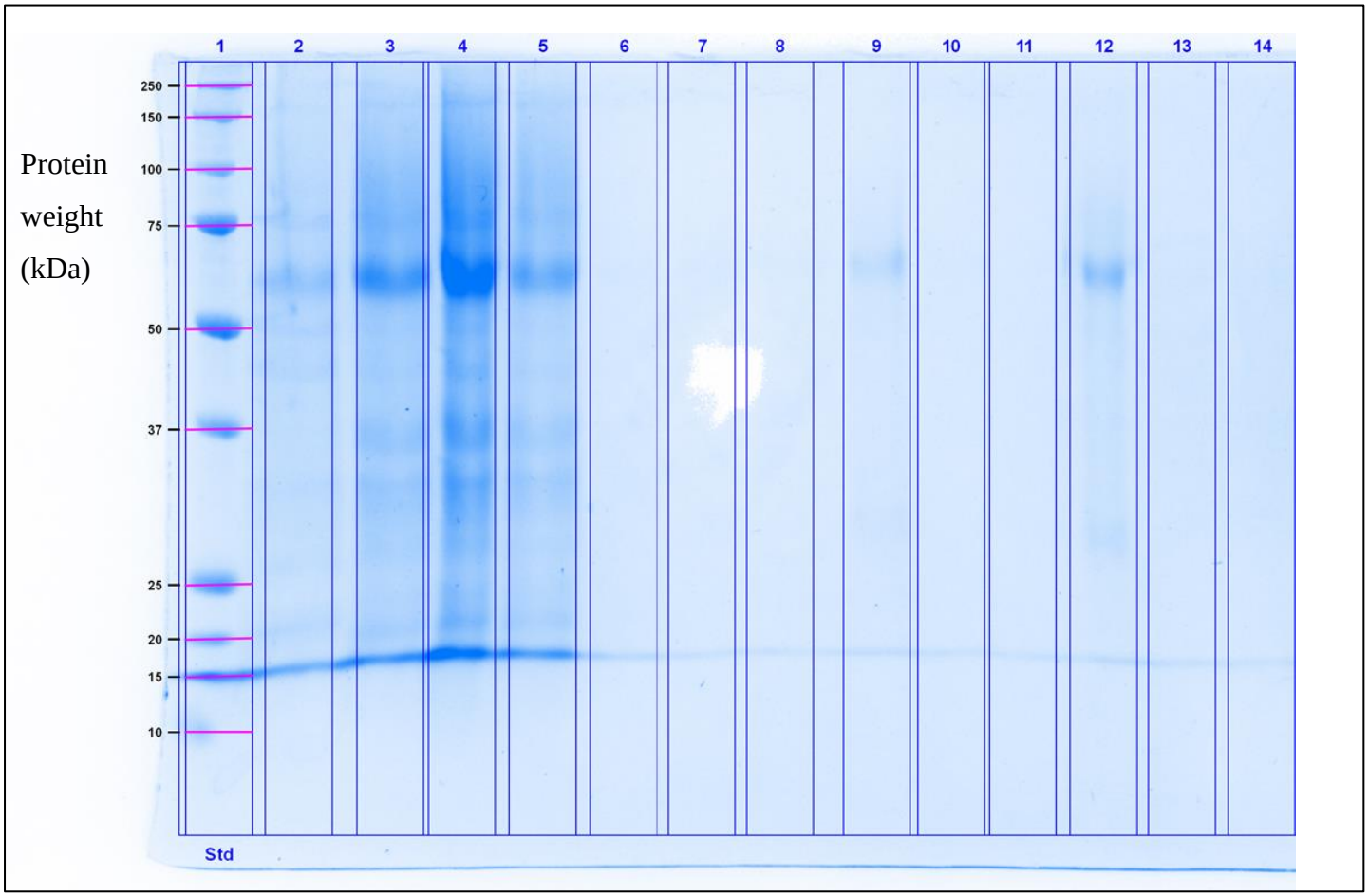


Figure 4.3: SDS-PAGE analysis of purified fractions of protein isolated from uPHM-pKOV410 *Yarrowia lipolytica* grown in YPD 2% in shaker flasks.
Figure legend:

Lane	Contents	Lane	Contents	Lane	Contents	Lane	Contents
1	Protein marker (2.5 µL)	5	Flowthrough	9	Elution 1	13	TFF 10 000 MWCO Filtrate
2	Crude culture supernatant	6	Wash 1	10	Elution 2		
3	TFF 10 000 MWCO Retentate	7	Wash 2	11	Elution 3		
4	Retentate and binding buffer	8	Wash 3-5	12	Purified final uPHM		

The novel 55 kDa protein, seen in lane 9 of Figure 4.3, was selectively retained after IMAC purification due to the 6x His-tag. The same purification procedures were performed with the Lip2 crude extract, with no detectable protein concentration after purification. Post-elution, the elution buffer and imidazole were reduced to >1.0 µM concentrations by filtration through 10 000 Da MWCO Vivaspin® 6 centrifugal concentrators. The final purified uPHM protein was diluted with a 3x protein storage solution for activity assays and storage, as seen in lane 12. The total quantity and concentrations of each step in the purification methods are summarised in Table 4.1.

Table 4.1: Protein purification table for uPHM with tangential flow filtration, immobilised metal chromatography, and dialysis buffer exchange and concentration.

Purification step	Total volume (mL)	Protein concentration (mg/mL)	Total protein (mg)
Crude	124	0.37	45.9
10 000 MWCO TFF	20	1.48	29.6
Ni IMAC elute	8.0	0.51	4.08
Membrane filtration	4.0	0.98	3.92

For purification, the pooled culture supernatant after centrifugation was designated as crude. The entire volume of 124 mL was reduced, with smaller contaminants such as yeast particulates <10 000 Da removed using TFF. An increase in concentration from 0.37 mg/mL to 1.48 mg/mL also increased the solution's clarity relative to the crude. The maximum binding capacity of the 5 mL MagResyn® Ni-NTA magnetic microbeads was 5 mg total, and the contaminating particulates could limit the column binding capacity. Therefore multiple passes with the total protein under 15 mg per sample were required to purify all uPHM in a fermentation batch. To prevent overloading the microspheres, only 10.0 mL of TFF retentate underwent IMAC purification. The first and second elution fractions were pooled before the final buffer exchange and concentration using centrifugal concentrators. The final quantity of purified uPHM (Figure 4.3) was 3.92 mg. The total purified protein equates to 9.8 mg out of the original 45.9 mg of crude protein, corresponding to 21.3% of the total protein of the crude.

The 55 kDa band, from Figure 4.3 lane 12, was excised from the gel, extracted from the polyacrylamide and SDS, then sent for peptide mass spectrometry (MS) to confirm the presence of the uPHM protein sequence. The band coverage was 34% for the uPHM amino acid sequence submitted, compared with the template amino acid sequence for uPHM. Several

other contaminating proteins were identified, including *Sus scrofa domesticus* trypsin from the protein digest and several *Homo sapiens* and *Y. lipolytica* proteins.

4.3.4 Fermentation for the production of uPHM

With the recombinant uPHM successfully identified and purified, optimisation of the protein production levels were performed. The initial fermentation parameters investigated were media composition, which influenced total cellular biomass and cell morphology and protein secretion levels. Standard YPD media had been used for all screening and initial expression attempts. Media optimisation was conducted at a small scale in shaker flask fermentation, but future fermentation media optimisation will be performed in multi-litre bioreactors, which entail unique fermentation characteristics and requirements.

4.3.4.1 Media composition optimisation for uPHM pKOV410 *Yarrowia lipolytica*

The optimisation of the fermentation media was performed as per section 4.2.3.1 used standard YPD media, Y2 and Y1, at a low-volume shaker flask scale (Antošová and Sychrová, 2016). Media that had been optimised for low cost and bioreactor production, such as protein production broth (P), and specialised media mP and mPx, were also tested. The effects of the medium on of total cell density, according to OD600 nm measurements, and the amount of secreted protein over seven days are shown in Figure 4.4.

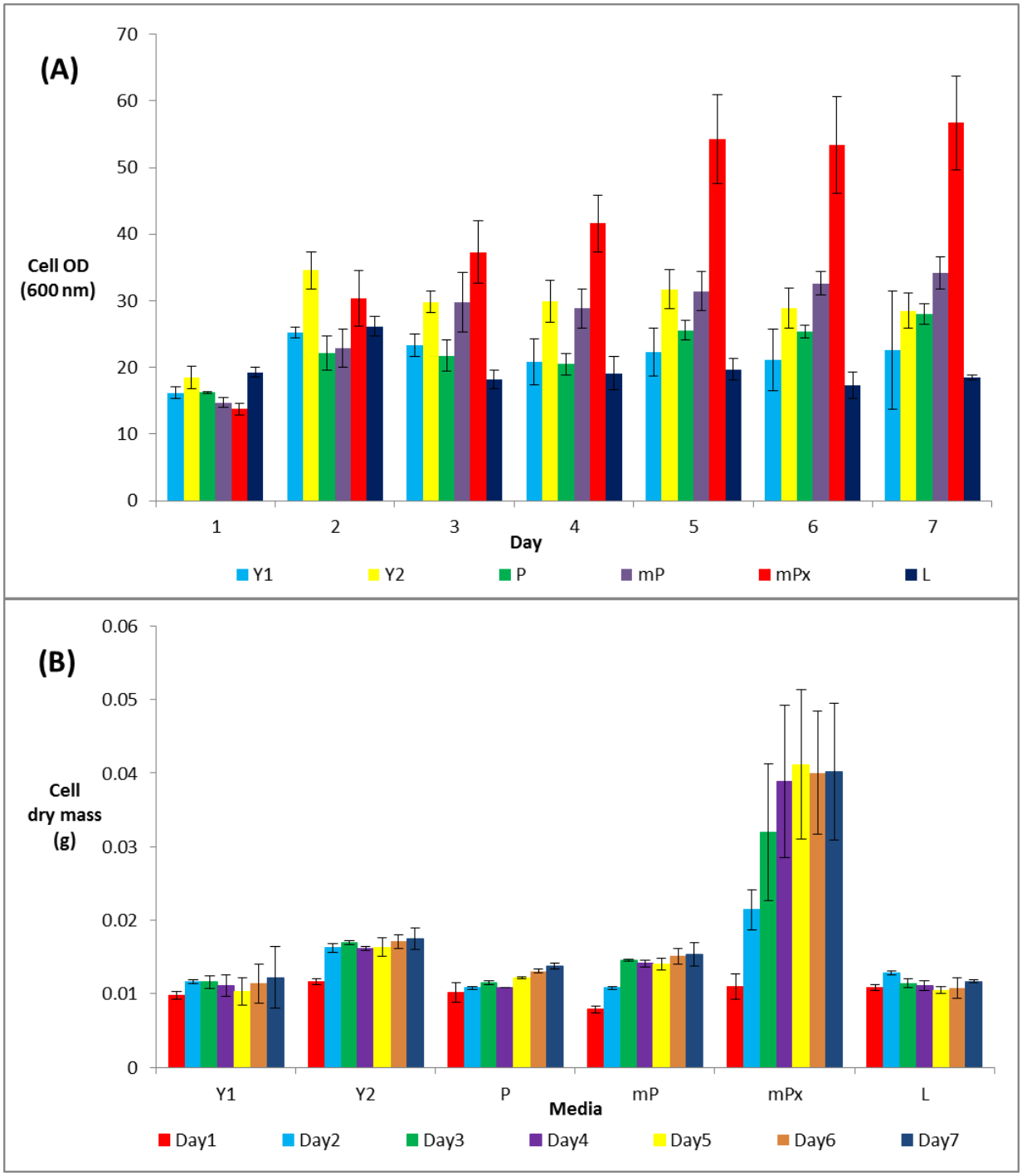


Figure 4.4: The effect of media composition on cell growth. **(A)** OD600 measurements to determine the cell densities of uPHM pKOV410 *Yarrowia lipolytica* grown in Y1, Y2, P, mP, mPx, and Lip2 with Y2 over seven days of fermentation. (n=3; SD ±). **(B)** The weight of dry cell biomass for uPHM pKOV410 *Yarrowia lipolytica* grown in Y1, Y2, P, mP, mPx, and Lip2 with Y2 over seven days of fermentation. (n=3; SD ±).

The OD600 nm and dry cell biomass measurements produced comparable data (Figure 4.4). Based on their similarity, only the OD600 nm measurement was used for subsequent investigations due to its high throughput and same-day results. The final cell mass and OD600 nm were comparable between Y1, Y2, P, and mP media, as each had an equivalent amount of

carbon- and nitrogen- sources (Figure 4.4 A). The highest cell density and dry mass were obtained using mPx media, which contained higher carbon-source concentrations and was adapted from Darvishi *et al.*, 2017. The OD600 nm of cells grown in mPx medium peaked on day 5, while the other media obtained maximum cell density on day 2 or 3. The protein concentration of all secreted proteins in the cell-free supernatant for each medium was monitored daily with Bradford's assays (Figure 4.5).

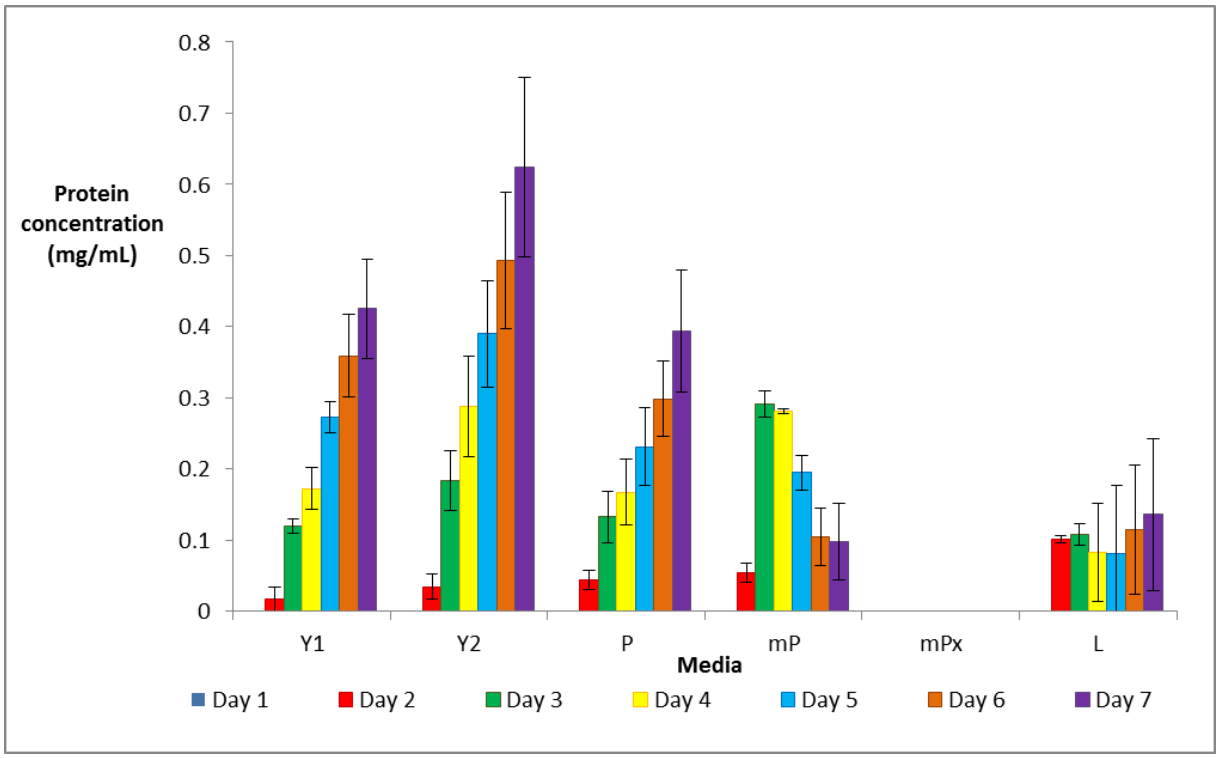


Figure 4.5: The protein concentration in the cell-free supernatant of uPHM pKOV410 *Yarrowia lipolytica* grown in Y1, Y2, P, mP, mPx, and Lip2 with Y2 media for seven days of fermentation. (n=3; SD ±)

The lack of protein production by cells grown in the bioreactor-optimised mPx medium was notable, as all other cultures produced detectable quantities of protein (Figure 4.5). The protein concentration produced by cells grown in mP, which contains the same reagents as mPx but with lower molasses and yeast extract concentrations, reached 0.31 mg/mL by day 3. Cells grown in the negative control medium containing Lip2, produced limited protein at 0.1 mg/mL. The low protein concentrations remained relatively constant throughout the fermentation compared to cells grown in Y2 medium, which had the highest protein secretion at 0.63 mg/mL by day 7 (Figure 4.5).

The lack of protein secretion by cells grown in mPx required further analysis and could be explained by cell morphology changes induced by the media composition (Madzak *et al.*,

2005). The effect of media on cell morphology was observed by comparing cells grown in Y2 and mPx by microscopy (Figure 4.6).

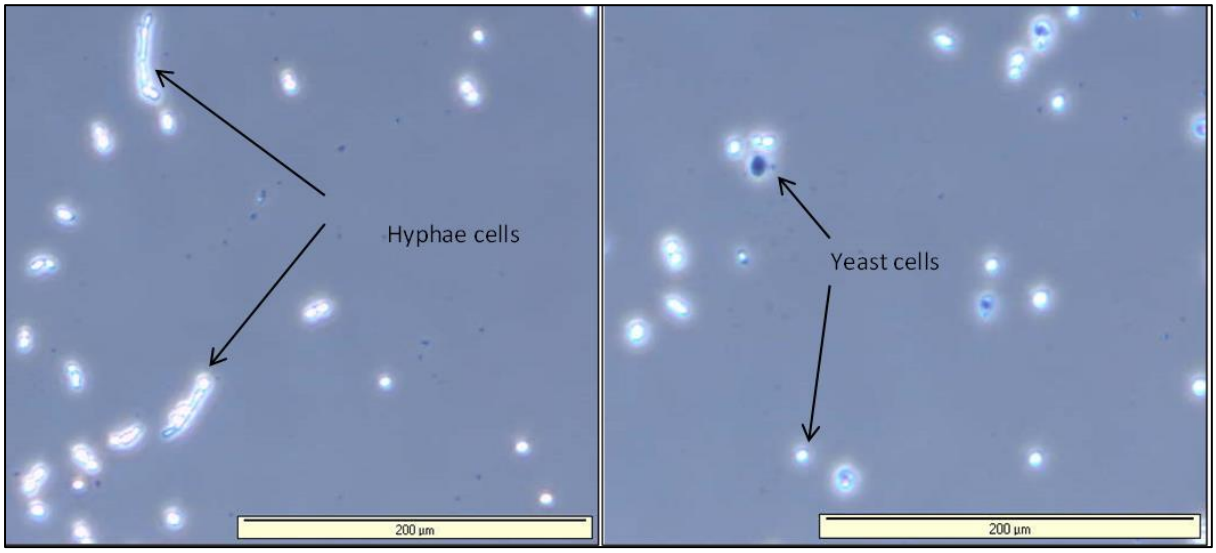


Figure 4.6: Images obtained by light microscopy of the cell morphology of the uPHM *Yarrowia lipolytica* cells with fermentations of mPx (left) and Y2 (right) media after seven days of fermentation. 100x magnification. Scale bar is indicated in the bottom right of each image.

Cells grown in mPx had both hyphae and yeast bud morphology, while cells grown in Y2 had only yeast cell morphology. Cells grown in Y1 and Y2 had only yeast bud cell morphology, but cells grown in P, mP, and mPx media had a combination of both cell morphologies.

4.3.5 Optimised fermentation pH for uPHM pKOV410 *Yarrowia lipolytica*

The effects of other culture conditions were subsequently analysed as described in section 4.2.3.1. An explanation for the observed changes in cell morphology and protein secretion could be culture pH changes during fermentation, suggesting the need for pH buffering in the culture media. Therefore, the effects of pH were analysed using Y2 media at 28°C with one unit increments of pH between 3.0 and 7.0 by measuring the cell density and supernatant protein concentration (Figure 4.7).

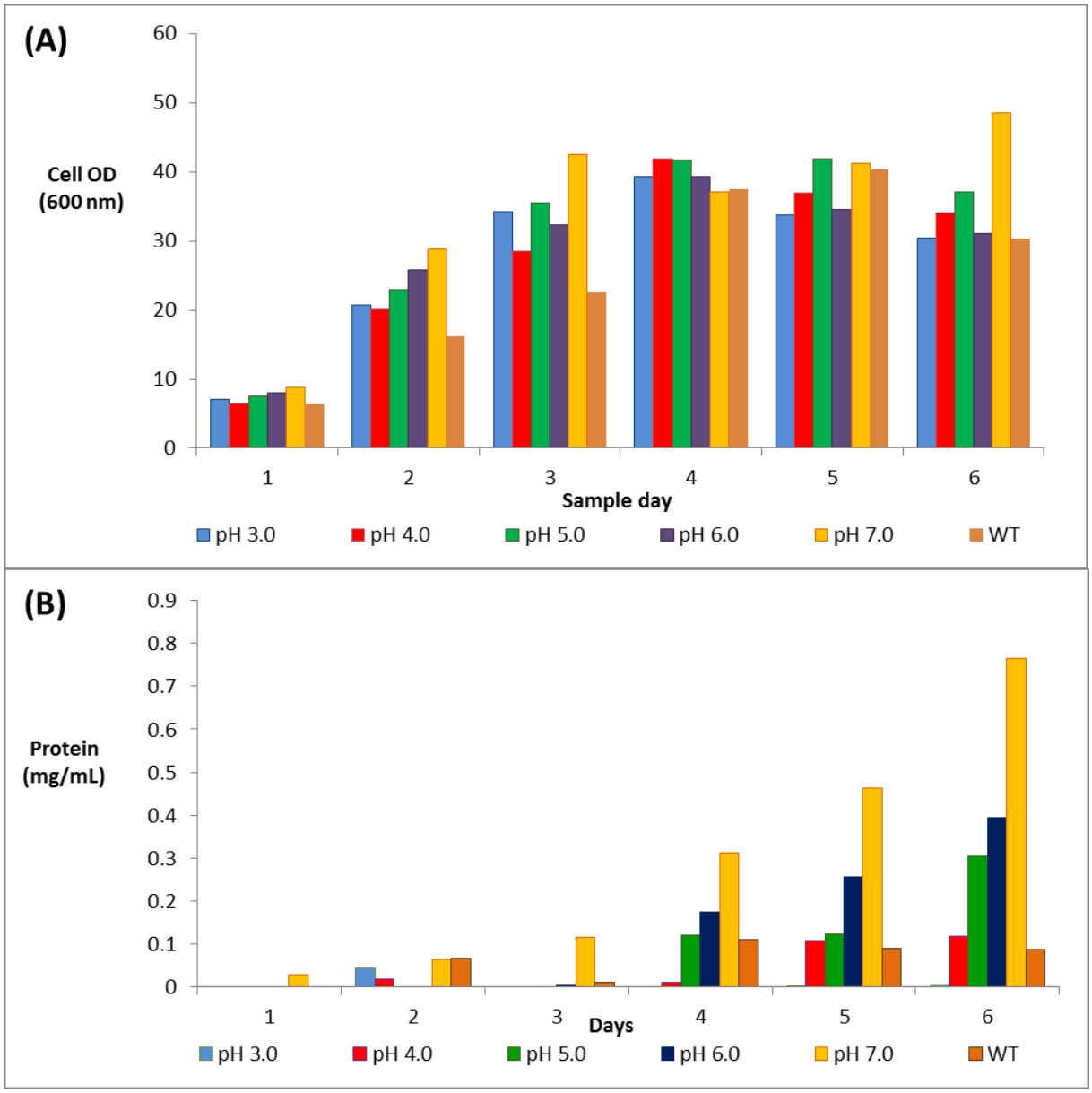


Figure 4.7: The effect of pH on cell density and secreted protein concentration. The optical density at 600 nm **(A)**, or the supernatant protein concentration **(B)** of uPHM pKOV410 *Yarrowia lipolytica* fermented in YPD 2% with 20 mM citrate at pH 3.0, 4.0, 5.0, 6.0, or 7.0 over six days of culture. (n=1)

The neutral pH of 7.0 was the optimal pH for maximal cell density, which reached 48.6 units at day 6 (Figure 4.7 A), and the highest protein secretion at 0.76 mg/mL (Figure 4.7 B). In contrast, media with a pH of 3.0 did not produce detectable protein concentrations for the transformed strain or the untransformed wild type control (Figure 4.7 B). The same direct relationship was not observed between pH and cell density in Figure 4.7 A. The effect of the addition of 20 mM citrate buffer on cell growth and secreted protein concentrations was next observed. The buffered samples were cultured simultaneously with the pH range cultures (Figure 4.8).

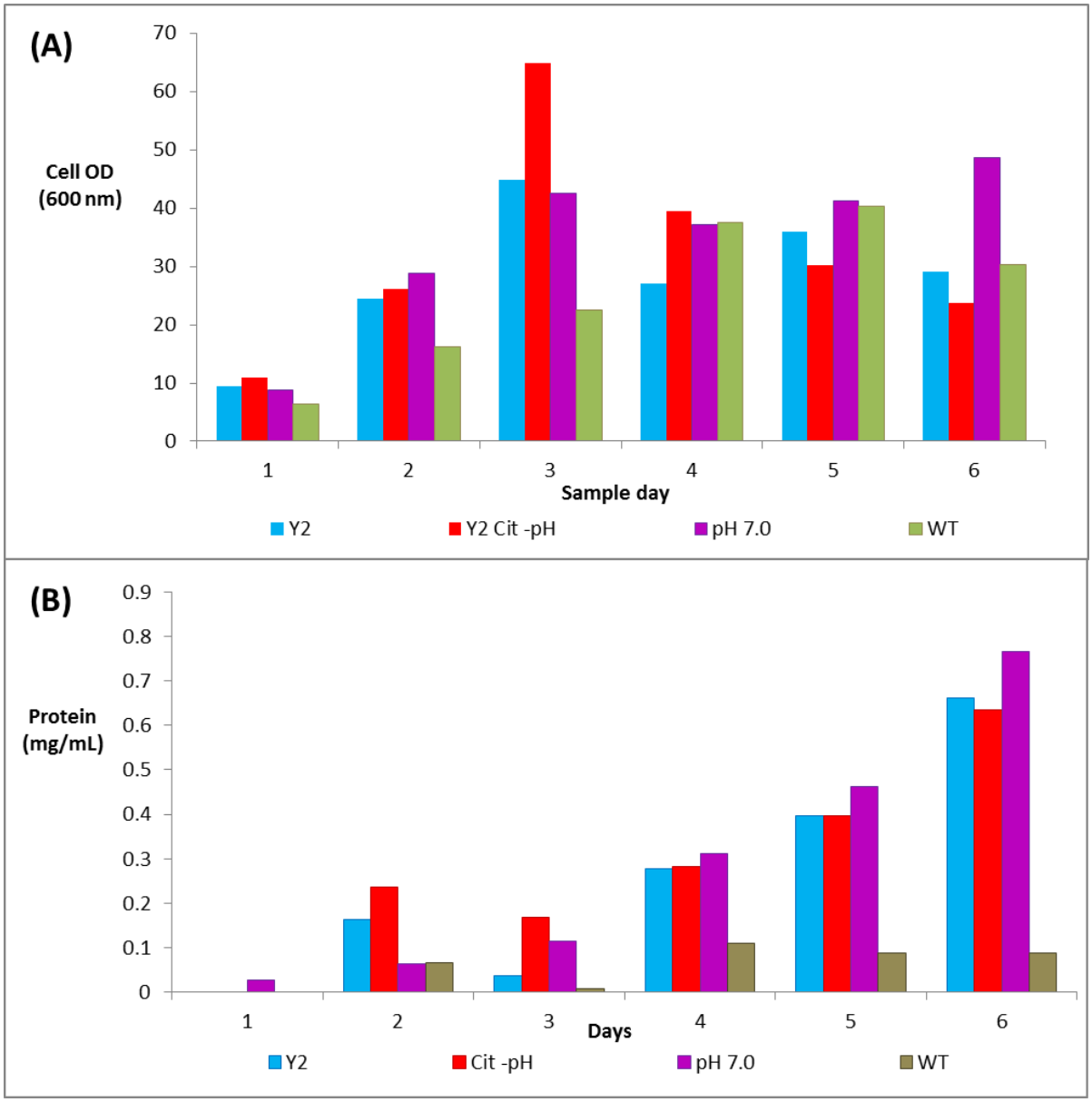


Figure 4.8: The effect of media on cell density and secreted protein concentration. The optical density at 600 nm **(A)**, or the supernatant protein concentration **(B)**, of uPHM pKOV410 *Yarrowia lipolytica* fermented in YPD 2%, YPD 2% with 20 mM citrate and no pH control medium, YPD 2% with 20 mM citrate and pH adjusted to 7.0, and untransformed wild type over six days (n=1).

The OD600 nm measurements were similar for all uPHM pKOV410 *Y. lipolytica* cultures despite the addition of 20 mM citrate as a buffering reagent and daily pH adjustments (Figure 4.8). This result mirrored the protein secretion levels, which were not apparently affected by 20 mM citrate or daily pH corrections. The pH 7.0 20 mM culture was the highest producer of secreted protein at 0.76 mg/mL, which was significantly higher than the baseline protein concentration secreted by the untransformed wild type (WT) control. The final culture

condition to be investigated was cultivation temperature, which was performed using cultures grown under optimised conditions in Y2 medium at pH 7.0.

4.3.6 Cultivation temperature optimisation for uPHM pKOV410 *Yarrowia lipolytica*

All cultures were inoculated and incubated at 28°C for 48 hours to compare secreted protein concentrations from an equivalent cell density in all cultures simultaneously. Triplicate sets of shaker flask cultures were then incubated at 10°C; 20°C; 28°C; and 32°C from day two. The protein secretion by each culture was measured, with the effect of different incubation temperatures observed from day three onwards (Figure 4.9).

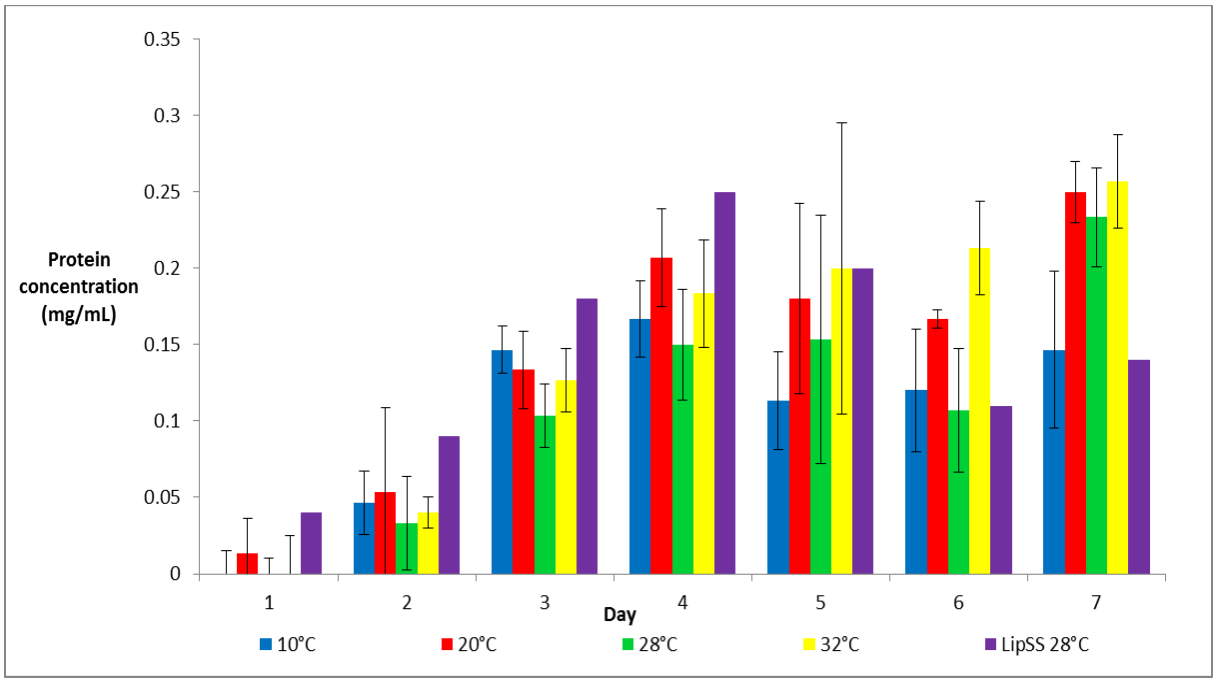


Figure 4.9: The supernatant protein concentration of uPHM pKOV410 *Yarrowia lipolytica* fermentation with 20 mM citrate at pH 7.0, grown at 10°C, 20°C, 28°C, and 32°C (n=3; SD: ±), and LipSS (n=1) over six days.

As expected with the *h4pd* promoter, appreciable protein concentration in the supernatant was not observed until day three (Figure 4.9), when the separate temperature regimes began. The secreted protein concentrations between samples did not significantly differ until day six, with the culture grown at 32°C producing more protein than the three grown at lower temperatures. By day 7, the cultures grown at 20°C, 28°C, and 32°C were producing an average of 0.24 mg/mL of supernatant protein, while the culture grown at 10°C produced less secreted protein at 0.155 mg/mL.

4.4 Discussion

The yeast screening PCR method's success in the present chapter was notable, as previously attempted methods were either less sensitive and inaccurate or not amenable to high-throughput screening of multiple colonies. The dilution of culture samples with 10 mM Li-acetate and 10 mM NaOH, heating, and centrifugation likely degraded and removed some of the yeast cell walls and sheared the gDNA, rendering the integrated pKOV410 site accessible to the DNA-dependant DNA-polymerase while reducing yeast-derived contaminants. When PCR was performed with purified pKOV410 pDNA in the presence of yeast cells, specific amplification of the gene cassette was possible, implying that its integration into the yeast genome rendered the cassette inaccessible to the polymerase.

A potentially superior screening method could be a reverse transcription PCR (RT-PCR) specific to the recombinant mRNA (Govender, 2015). The primary advantage of this alternative method is the confirmation of active transcription of the recombinant genes, with the option to quantify transcription levels in parallel. A disadvantage is that the current pre-treatment method may not be adaptable to mRNA, limiting its potential for colony screening. Additional reagents in solution would include RNase inhibitors and RNase free solutions, and the process would entail laborious RNA handling techniques.

An mRNA-based method's fundamental problem would be variable transcription levels of the recombinant gene, which may be influenced by growth media and cell growth stage. The *hp4d* promoter would limit transcription to the early log phase, rendering the method useful for late-stage culture studies (Nicaud *et al.*, 2002). The ability to confirm the DNA sequence identity and integration was a higher priority than the adaptation of a novel RNA screening technique, which was out of the project scope. With the current PCR method, it was possible to detect uPHM integrated into host gDNA, which was essential for screening colonies for high uPHM production levels. With a suitable strain selected, the next step was the purification of the produced uPHM protein with industrially applicable methods.

The highest reduction in total volume was achieved with the TFF purification step. TFF can be scaled up to relevant industrial levels, meeting an important requirement for the present study. If a partial purification of uPHM for downstream applications with lower protein homogeneity levels is required, this step alone could be sufficient. The reduction in volume could be increased with larger systems as necessary. Additionally, the TFF cartridges were re-usable

after cleaning and are low cost with inexpensive buffering reagents used, helping to further the overall green chemistry objectives of the present study.

The nickel-affinity IMAC purification using the 6xHis-tag on uPHM successfully removed contaminating proteins and selectively eluted uPHM. The protein was expressed and purified in mg per litre concentrations, despite the potential inhibitory effects affinity tags may have on expression levels with Lip2 pKOV410 *Y. lipolytica* (Bulani, 2013; Wood, 2014; Gerber *et al.*, 2015). The relatively small size (0.84 kDa) C-terminal location of the 6xHis-tag resulted in similar amounts of secreted proteins to those of previous recombinant protein expression studies, suggesting that this affinity tag did not inhibit expression. The availability of commercial antibodies specific to the 6xHis tag epitope proved beneficial in the initial screening of transformant colonies for uPHM production (data not shown). The band containing uPHM was identifiable with standard SDS-PAGE methods, reducing the need for western blots.

The SDS-PAGE band containing uPHM was less distinct than other bands, potentially due to the hyperglycosylation of uPHM (Cereghino and Cregg, 2000; Madzak *et al.*, 2005). The variable length of polysaccharide moieties generated from this post-translational modification step increases the protein mass relative to the predicted mass. The differences in mass are further compounded by high levels of diversity in the glycosylation patterns and residues in fungal laccase glycosylation (Jolivet *et al.*, 2007; Çelik and Çalık, 2012). These factors would interfere with the migration of the proteins under electrophoresis, resulting in a less-defined banding pattern for uPHM. The limited coverage of the uPHM sequence with the peptide mass MS indicated likely interference from glycosylation. In particular, the sequence areas around adjacent potential glycosylation sites did not generate sequence data, likely due to interference. Stricter purification methods to avoid contaminating proteins and deglycosylation studies with uPHM would improve sequence identity coverage and understanding of the glycosylation patterns present on uPHM. It may be possible to reduce the hyperglycosylation present by altering the fermentation conditions. The protein production observed on days 2 and 3 confirmed the expression in the initial stages of late-phase growth, as expected with the *h4pd* promoter.

The effect of media composition on cell morphology was a critical indicator of recombinant protein production and secretion. The cultures with only yeast cells present, such as Y1 and Y2, had the highest protein concentrations, producing up to 0.63 mg/mL after seven days of

fermentation. Higher concentrations of secreted proteins were achieved by cells grown in Y1 and Y2 media, despite the lower late-stage cell densities compared to cells grown in mPx. A possible explanation for the lack of protein secretion by cells grown in mPx may be the limited oxygen exchange within the shaker flask at 200 rpm. Dissolved oxygen concentrations of 1.5 mg/L for prolonged periods, which could be caused by the high cell density of cells grown in mPx, and the relatively limited oxygen exchange available in the shaker flask, could explain the presence of hyphae. Fluted shaker flasks and increased shaking above 250 rpm may offer an increased gaseous exchange, but these measures can result in the generation of foam, which would be counterproductive for oxygen exchange. The lower cell densities of cells grown in P and mP media, which also contained hyphae, imply that the cell morphology was likely caused by the media composition and not oxygen limitations.

The importance of culture conditions, especially pH, was confirmed by observing cell growth under different pH conditions, which revealed a linear relationship between pH and protein secretion, with media at pH 7.0 supporting the most cell growth. Basic pH levels were not investigated, as these would likely denature any secreted protein. In addition, the pH buffering reagent 20 mM citrate did not inhibit protein secretion and, in conjunction with constant daily pH adjustments to pH 7.0, resulted in the highest secreted protein concentrations of 0.73 mg/mL.

A reduction of culture temperature to 10°C reduced protein secretion by 50% relative to a culture temperature of 28°C. The difference in secreted protein concentrations above 20°C was not significant by day 7. The amount of secreted protein fell within the standard deviation range for shaker flasks incubated at each temperature.

With the primary fermentation conditions and a purification method optimised for uPHM production in shaker flasks, the analysis of uPHM activity was possible.

4.5 References

Alessandra, P., Cinzia, P., Paola, G., Vincenza, F. and Sannia, G. (2010) 'Heterologous laccase production and its role in industrial applications', *Bioengineered Bugs*, 1(4), pp. 252–262. doi: 10.4161/bbug.1.4.11438.

Antošová, Z. and Sychrová, H. (2016) 'Yeast hosts for the production of recombinant laccases: A review', *Molecular Biotechnology*, 58(2), pp. 93–116. doi: 10.1007/s12033-015-9910-1.

Attenborough, R. M. F., Hayward, D., Kitahara, M., Miller, D. and Ball, E. (2012) 'A "neural" enzyme in nonbilaterian animals and algae: preneural origins for peptidylglycine α-amidating monooxygenase.', *Molecular Biology and Evolution*, 29(10), pp. 3095–109. doi: 10.1093/molbev/mss114.

Bankar, A. V., Kumar, A. R. and Zinjarde, S. S. (2009) 'Environmental and industrial applications of *Yarrowia lipolytica*', *Applied Microbiology and Biotechnology*, 84(5), pp. 847–865. doi: 10.1007/s00253-009-2156-8.

Barth, G. and Gaillardin, C. (1996) '*Yarrowia lipolytica*', in *Nonconventional Yeasts in Biotechnology*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 313–388. doi: 10.1007/978-3-642-79856-6_10.

Bellou, S., Makrim A., Triantaphyllidou, I., Papanikolaou, S. and Aggelis, G. (2014) 'Morphological and metabolic shifts of *Yarrowia lipolytica* induced by alteration of the dissolved oxygen concentration in the growth environment', *Microbiology*, 160(Pt_4), pp. 807–817. doi: 10.1099/mic.0.074302-0.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J., Molina-Jouve, C. and Nicaud, J. (2009) '*Yarrowia lipolytica* as a model for bio-oil production.', *Progress in Lipid Research*, 48(6), pp. 375–87. doi: 10.1016/j.plipres.2009.08.005.

Bio-Rad Laboratories, I. (2012) *A Guide to Polyacrylamide Gel Electrophoresis and Detection, Bio-Rad*. 6040.

Bohlin, C., Jönsson, L., Roth, R. and van Zyl, W. (2006) 'Heterologous expression of *Trametes versicolor* laccase in *Pichia pastoris* and *Aspergillus niger*.', *Applied Biochemistry and Biotechnology*, 129–132, pp. 195–214. doi: 10.1385/ABAB:129:1:195.

Bulani, S. I., Moleleki, L., Albertyn, J. and Moleleki, N. (2012) 'Development of a novel rDNA based plasmid for enhanced cell surface display on *Yarrowia lipolytica*.', *AMB Express*, 2(1), p. 27. doi: 10.1186/2191-0855-2-27.

Bulani, S. I. (2013) *Yarrowia lipolytica* heterologous protein expression. PhD thesis, University of Free State.

Çelik, E. and Çalık, P. (2012) 'Production of recombinant proteins by yeast cells', *Biotechnology Advances*, 30(5), pp. 1108–1118. doi: 10.1016/j.biotechadv.2011.09.011.

Cereghino, J. L. and Cregg, J. M. (2000) 'Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*', *FEMS Microbiology Reviews*, 24(1), pp. 45–66. doi: S0168-6445(99)00029-7 [pii].

Chen, D. C., Beckerich, J. M. and Gaillardin, C. (1997) ‘One-step transformation of the dimorphic yeast *Yarrowia lipolytica*.’, *Applied Microbiology and Biotechnology*, 48(2), pp. 232–5. doi: 10.1007/s002530051043.

Darvishi, F., Moradi, M., Madzak, C. and Jolivalt, C. (2017) ‘Production of laccase by recombinant *Yarrowia lipolytica* from molasses: Bioprocess development using statistical modeling and increase productivity in shake-flask and bioreactor cultures’, *Applied Biochemistry and Biotechnology*, 181(3), pp. 1228–1239. doi: 10.1007/s12010-016-2280-8.

Derewenda, Z. (2004) ‘The use of recombinant methods and molecular engineering in protein crystallization’, *Methods*, 34(3), pp. 354–363. doi: 10.1016/j.ymeth.2004.03.024.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuvéglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J., Beyne, E., Bleykasten, C., Boisramé, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G., Straub, M., Suleau, A., Swennen, D., Tekaia, F., Wésolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P. and Souciet, J. (2004) ‘Genome evolution in yeasts.’, *Nature*, 430(6995), pp. 35–44. doi: 10.1038/nature02579.

Dulermo, T. and Nicaud, J. (2011) ‘Optimization of lipid synthesis and accretion’, *US Patent App. 13/806,332*. Available at: http://www.google.com/patents/US20130149754 (Accessed: 2 April 2014).

Dyballa, N. and Metzger, S. (2009) ‘Fast and sensitive colloidal Coomassie G-250 staining for proteins in polyacrylamide gels’, *Journal of Visualized Experiments*, (30), pp. 2–5. doi: 10.3791/1431.

Eipper, B. A., Quon, A. S., Mains, R. E., Boswell, J. S. and Blackburn, N. J. (1995) ‘The catalytic core of peptidylglycine alpha-hydroxylating monooxygenase: investigation by site-directed mutagenesis, Cu X-ray absorption spectroscopy, and electron paramagnetic resonance.’, *Biochemistry*, 34(9), pp. 2857–65. Available at: http://www.ncbi.nlm.nih.gov/pubmed/7893699

Gaier, E. D., Miller, M., Ralle, M., Aryal, D., Wetsel, W., Mains, R. and Eipper, B. (2013) ‘Peptidylglycine α-amidating monooxygenase heterozygosity alters brain copper handling with region specificity.’, *Journal of neurochemistry*. doi: 10.1111/jnc.12438.

Gale, J. S., McIntosh, J. E. A. and McIntosh, R. P. (1988) ‘Peptidyl-glycine α-amidating mono-oxygenase activity towards a gonadotropin-releasing-hormone C-terminal peptide substrate, in subcellular fractions of sheep brain and pituitary’, *Biochemical Journal*, 251(1), pp. 251–259. doi: 10.1042/bj2510251.

Gerber, I., Jordaan, J. and Crampton, M. (2015) *Efficient purification of polyhistidine-tagged proteins under denaturing conditions using MagReSyn® NTA: Inclusion bodies & hidden tags*. Available at: http://www.resynbio.com/index_htm_files/NTA_Appnote.pdf.

Gonzalez-Lopez, C. I., Szabo R., Blanchin-Roland, S. and Gaillardin, C. (2002) 'Genetic control of extracellular protease synthesis in the yeast *Yarrowia lipolytica*.', *Genetics*, 160(2), pp. 417–27. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1461987&tool=pmcentrez&rendertype=abstract.

Govender, S. (2015) Elucidating the role of growth rate on the production of a fusion protein under regulation of the hp4d promoter by *Yarrowia lipolytica*. MSc thesis. University of Stellenbosch.

Green, M. R. and Sambrook, J. (2012) *Molecular Cloning: A Laboratory Manual*. 4th edn, *Cold Spring Harbor Laboratory Press*. 4th edn. New York: John Inglis. Available at: www.cshlpress.org.

Groenewald, M., Boekhout, T., Neuvéglise, C., Gaillardin, C., van Dijck, P. and Wyss, M. (2014) '*Yarrowia lipolytica* : Safety assessment of an oleaginous yeast with a great industrial potential', *Critical Reviews in Microbiology*, 40(3), pp. 187–206. doi: 10.3109/1040841X.2013.770386.

Hanke, A. T. and Ottens, M. (2014) 'Purifying biopharmaceuticals: Knowledge-based chromatographic process development', *Trends in Biotechnology*, 32(4), pp. 210–220. doi: 10.1016/j.tibtech.2014.02.001.

Hofmeyer, T., Bulani, S., Grzeschik, J., Krah, S., Glotzbach, B., Uth, C., Avrutina, O., Brecht, M., Göringer, H., Van Zyl, P. and Kolmar, H. (2014) 'Protein production in *Yarrowia lipolytica* via fusion to the secreted lipase Lip2p.', *Molecular Biotechnology*, 56(1), pp. 79–90. doi: 10.1007/s12033-013-9684-2.

Jolivalt, C., Madzak, C., Brault, A., Caminade, E., Malosse, C. and Mougin, C. (2005) 'Expression of laccase IIIb from the white-rot fungus *Trametes versicolor* in the yeast *Yarrowia lipolytica* for environmental applications', *Applied Microbiology and Biotechnology*, 66(4), pp. 450–456. doi: 10.1007/s00253-004-1717-0.

Jolivet, P., Bordes, F., Fudalej, F., Cancino, M., Vignaud, C., Dossat, V., Burghoffer, C., Marty, A.,Chardot, T. and Nicaud, J. (2007) 'Analysis of *Yarrowia lipolytica* extracellular lipase Lip2p glycosylation', *FEMS Yeast Research*, 7(8), pp. 1317–1327. doi: 10.1111/j.1567-1364.2007.00293.x.

Kahari, D., Koorsen, G., Stoychev, S., Chipeta, Z., Labuschagne, M., Van Zyl, P., Nthangeni, B. and Zyl, V. (2007) 'Glycoprofiling of N-linked glycans of erythropoietin therapeutic protein expressed in *Yarrowia lipolytica*', in *CSIR innovations conference*, p. 6665. CSIR. Available at: http://researchspace.csir.co.za/dspace/bitstream/handle/10204/2866/Kahari_P_2008.pdf;sequence=1.

Kerscher, S., Durstewitz, G., Casaregola, S., Gaillardin, C. and Brandt, U. (2001) 'The complete mitochondrial genome of *Yarrowia lipolytica*', *Comparative and Functional Genomics*, 2(2), pp. 80–90. doi: 10.1002/cfg.72.

Kim, J. Y., Kim, Y.-G. and Lee, G. M. (2012) 'CHO cells in biotechnology for production of recombinant proteins: Current state and further potential', *Applied Microbiology and Biotechnology*, 93(3), pp. 917–930. doi: 10.1007/s00253-011-3758-5.

Kim, K. and Seong, B. L. (2001) 'Peptide amidation: Production of peptide hormones *in vivo* and *in vitro*', *Biotechnology and Bioprocess Engineering*, 6(4), pp. 244–251. doi: 10.1007/BF02931985.

Laemmli, U. K. (1970) 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4', *Nature*, 227(5259), pp. 680–685. doi: 10.1038/227680a0.

Liu, D., Coloe, S., Baird, R. and Pedersen, J. (2000) 'Rapid mini-preparation of fungal DNA for PCR', *Journal of Clinical Microbiology*, 38(1), p. 471.

Liu, H.-H., Ji, X.-J. and Huang, H. (2015) 'Biotechnological applications of *Yarrowia lipolytica*: Past, present and future.', *Biotechnology Advances*, 33(8), pp. 1522–1546. doi: 10.1016/j.biotechadv.2015.07.010.

Liu, L. and Alper, H. S. (2014) 'Draft genome sequence of the oleaginous yeast *Yarrowia lipolytica* PO1f, a commonly used metabolic engineering host', *Genome Announcements*, 2(4), pp. 2013–2014. doi: 10.1128/genomeA.00652-14.

Liu, L., Redden, H. and Alper, H. S. (2013) 'Frontiers of yeast metabolic engineering: Diversifying beyond ethanol and *Saccharomyces*', *Current Opinion in Biotechnology*, 24(6), pp. 1023–1030. doi: 10.1016/j.copbio.2013.03.005.

Madzak, C., Otterbein, L., Chamkha, M., Moukha, S., Asther, M., Gaillardin, C. and Beckerich, J. (2005) 'Heterologous production of a laccase from the basidiomycete *Pycnoporus cinnabarinus* in the dimorphic yeast *Yarrowia lipolytica*', *FEMS Yeast Research*, 5(6–7), pp. 635–646. doi: 10.1016/j.femsyr.2004.10.009.

Madzak, C. (2015) '*Yarrowia lipolytica*: recent achievements in heterologous protein expression and pathway engineering', *Applied Microbiology and Biotechnology*, 99(11), pp. 4559–4577. doi: 10.1007/s00253-015-6624-z.

Madzak, C. and Beckerich, J. (2013) 'Heterologous protein expression and secretion in *Yarrowia lipolytica*', in Barth, G. (ed.) *Microbiology Monographs*. Berlin, Heidelberg: Springer Berlin Heidelberg (Microbiology Monographs), pp. 1–76. doi: 10.1007/978-3-642-38583-4_1.

Madzak, C., Gaillardin, C. and Beckerich, J.-M. (2004) 'Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: A review', *Journal of Biotechnology*, 109(1), pp. 63–81. Available at: http://www.sciencedirect.com/science/article/pii/S0168165604000264

El Meskini, R., Culotta, V., Mains, R. and Eipper, B. (2003) 'Supplying copper to the cuproenzyme peptidylglycine α-amidating monooxygenase', *Journal of Biological Chemistry*, 278(14), pp. 12278–12284. doi: 10.1074/jbc.M211413200.

Nicaud, J., Madzak, C., van den Broek, P., Gysler, C., Duboc, P., Niederberger, P. and Gaillardin C. (2002) 'Protein expression and secretion in the yeast *Yarrowia lipolytica*.', *FEMS yeast research*, 2(3), pp. 371–9. Available at: http://www.ncbi.nlm.nih.gov/pubmed/12702287.

Pardee, K., Slomovic, S., Nguyen, P., Lee, J., Donghia, N., Burrill, D., Ferrante, T., McSorley, F., Furuta, Y., Vernet, A., Lewandowski, M., Boddy, C., Joshi, N. and Collins, J. (2016) 'Portable, on-demand biomolecular manufacturing', *Cell*, 167(1), pp. 248-259.e12. doi: 10.1016/j.cell.2016.09.013.

Porro, D., Gasser, B., Fossati, T., Maurer, M., Branduardi, P., Sauer, M. and Mattanovich, D. (2011) 'Production of recombinant proteins and metabolites in yeasts: When are these systems better than bacterial production systems?', *Applied microbiology and biotechnology*, 89(4), pp. 939–48. doi: 10.1007/s00253-010-3019-z.

Riva, S. (2006) 'Laccases: Blue enzymes for green chemistry', *Trends in Biotechnology*, 24(5), pp. 219–226. doi: 10.1016/j.tibtech.2006.03.006.

Theerachat, M., Emond, S., Cambon, E., Bordes, F., Marty, A., Nicaud, J., Chulalaksananukul, W., Guieysse, D., Remaud-Siméon, M. and Morel, S. (2012) 'Engineering and production of laccase from *Trametes versicolor* in the yeast *Yarrowia lipolytica*', *Bioresource Technology*, 125, pp. 267–274. doi: 10.1016/j.biortech.2012.07.117.

Ul-Hasan, S., Burgess, D., Gajewiak, J., Li, Q., Hu, H., Yandell, M., Olivera, B. and Bandyopadhyay, P. (2013) 'Characterization of the peptidylglycine α-amidating monooxygenase (PAM) from the venom ducts of neogastropods, *Conus bullatus* and *Conus geographus*.', *Toxicon : official journal of the International Society on Toxinology*, 74, pp. 215–24. doi: 10.1016/j.toxicon.2013.08.054.

Wood, D. W. (2014) 'New trends and affinity tag designs for recombinant protein purification', *Current Opinion in Structural Biology*, 26(1), pp. 54–61. doi: 10.1016/j.sbi.2014.04.006.

Young, C. L., Britton, Z. T. and Robinson, A. S. (2012) 'Recombinant protein expression and purification: A comprehensive review of affinity tags and microbial applications', *Biotechnology Journal*, 7(5), pp. 620–634. doi: 10.1002/biot.201100155.

Chapter 5: Peptidylglycine alpha-amidating monooxygenase activity

5.1 Introduction

While other enzyme activity assays can simply measure substrate consumption or product formation using absorbance spectroscopy, measuring PAM activity is more complex and has necessitated the development of PAM-specific assays. Consequently, the characterisation and measurement of PAM activity have been the subject of constant research, with methodological iterations ranging over a multitude of techniques (Li *et al.*, 1994; Schomburg and Stephan, 1994; Carpenter, 2006; Alwan *et al.*, 2019). The key to these method refinements and discoveries was identifying the multiple constituents and conditions required for the solutions used to measure PAM activity.

5.1.1 Initial PAM discovery and characterisation

The confirmation of PAM enzyme catalysis was first documented by Bradbury *et al.*, 1982, with an assay using porcine pituitary homogenate and its fractions. The radioisotope-labelled peptide substrate ^{125}I-D-Tyr-Val-Gly and high performance liquid chromatography (HPLC) were critical to the success of this assay, together conferring sufficient sensitivity to detect the low levels of activity. The initial confirmation of PAM enzymatic activity also revealed its slow catalytic turnover, which is one of its essential characteristics and inherent disadvantages (Bradbury *et al.*, 1982). In this early study, only 38% of the substrate was converted into the amidated product over 16 hours. These initial studies had particularly low turnover rates due to suboptimal conditions and limited purification, which were refined in subsequent enzyme activity studies. Importantly, PAM's cellular topological placement in signalling pathways was established from the cellular organelles present, guiding the design of PAM activity conditions for future assays. The identification of optimal PAM conditions and co-substrates, namely copper and ascorbate, also helped address the low levels of activity.

Discovering the optimal pH for the activity solution was critical to understanding and obtaining higher PAM activity (Glembotski *et al.*, 1984). Subsequent studies showed that the PHM and PAL subunits have different pH optima. A broad range of pH optima has been measured with PAM variants, either recombinantly produced or purified from hosts, ranging from acidic at pH 4.5 for *Rattus norvegicus* to alkaline at pH 8.5 for *Bos taurus* (Schomburg and Stephan, 1994). The majority of PAMs and particularly PHMs studied to date, have acidic pH optima,

and this property is shared among novel PAMs (Schomburg and Stephan, 1994; Münch *et al.*, 2021). The PAM reaction mechanism was determined after the intermediate products of the C-terminal glycine hydroxylated residue catalysed by PHM were identified (Tajima *et al.*, 1990). The hydrolysis products were identified and measured with HPLC and MS, highlighting the importance of these techniques for detecting PAM products. MS analysis also revealed that alkaline pH incubations could substitute for the PAL lyase reaction, with hydroxylated substrates αHA and Phe-Gly-Phe-Gly-OH used to produce benzamide and Phe-Gly-Phe-NH_2, respectively (Katopodis *et al.*, 1990; Tajima *et al.*, 1990). The use of HPLC-MS also enabled the discovery that the substrate specificity of rhPAM is limited to polypeptides and analogues, as no activity with HA has been measured with the human isoform (Merkler *et al.*, 2008; Trendel *et al.*, 2008; Cao *et al.*, 2011; Handa *et al.*, 2012; Kline *et al.*, 2013).

A significant improvement to the reproducibility of PAM activity assays was the identification of PHM oxygen requirements, which led to the incorporation of catalase to provide O_2 into activity solutions for determining PHM activity (Bradbury and Smyth, 1987). Bovine catalase has constant activity in a pH range of 4.0 to 8.5 but is partly inhibited and degraded by the copper and ascorbic acid in the PHM activity solution (Chance, 1952; Orr, 1967). The limited catalase activity can be counteracted by including a substrate such as H_2O_2, a stabiliser such as ethanol, and increasing the catalase concentrations in PAM activity solutions (Shimoi *et al.*, 1992; Schomburg *et al.*, 1994). The constant oxygen generated from catalase provides an alternate method to measure PAM activity by monitoring oxygen consumption within a sealed reaction vessel (King *et al.*, 2000). The oxygen concentration in the chamber becomes a function of catalase activity less the quantity of oxygen consumed by PAM. Surfactants such as Triton X-100 are included with PAM activity solutions to bind to the surface of the reaction vessel to prevent protein adsorption (King *et al.*, 2000). Monitoring substrate consumption is low throughput and can be tedious.

The link between substrates, enzyme activity, and products can be compromised by contaminants or substrate losses unrelated to enzyme activity. An inherent weakness in these methods is the effect of varying levels of catalase activity due to acidic conditions and catalase inhibitors, which are necessary for measuring PHM activity. Unstable catalase activity would compromise the measurement of PAM activity. Measuring an end product of PAM, such as glyoxylate, which is produced in equimolar quantities to the amidated peptide product, is, therefore, a superior methodology.

Studies of PAM activity have monitored the production of glyoxylate as the endpoint using phenylhydrazine-based assays (Christman *et al.*, 1944; Katopodis *et al.*, 1991). Phenylhydrazine is a toxic and inflammable reagent frequently used as rocket fuel (Gholamian *et al.*, 2012). Besides the phenylhydrazine, the strongly acidic development solution requires ferricyanide, which is also a hazardous material. In addition to being toxic, the assay is also temperature-sensitive, making it both dangerous and laborious (Christman *et al.*, 1944; Katopodis and May, 1990).

5.1.2 PAM activity assay with DMPD

Absorbance assays utilising a reporter chromophore are a superior method for measuring enzyme activity without imposing additional enzyme constraints. An absorbance assay using the reductant N,N-dimethyl-1,4-phenylenediamine (DMPD) as a substitute for ascorbic acid enabled the measurement of PAM activity (Li *et al.*, 1994). The substitution for a chromophoric reductant was based on similar assays with dopamine β-monooxygenase, which has a similar structure and activity to PAM (Wimalasena and Wimalasena, 1991). The reduction of DMPD by PHM produces a purple chromophore with an absorbance coefficient of 7300 $M^{-1}.cm^{-1}$, measured at 515 nm (Li *et al.*, 1994). A significant problem with the DMPD PAM assay is the continuous oxidation of DMPD by high levels of dissolved oxygen and non-specific reduction by the copper ions in PHM (Li *et al.*, 1994). While this assay showed initial promise, these issues have limited its use to studies involving oxygen consumption and X-ray fluorescence of PHM-bound copper ions, instead of absorbance measurements for activity assays as initially demonstrated (Chauhan *et al.*, 2016; Alwan *et al.*, 2019).

An absorbance-based assay for determining PAM activity, such as the 2-aminobenzaldehyde-glycine-glyoxylate (AGG) assay, would provide an inexpensive and more accessible method for studying PAM. UPLC-MS and AGG could provide definitive identification and measurement of glyoxylate production from novel PHM or PAM produced with recombinant techniques.

5.1.3 Aim

To demonstrate that commercial PAM and uPHM activity can be measured using the AGG assay and UPLC-MS, in order to determine the optimal conditions for enzyme activity.

5.1.4 Objectives

To use the AGG assay to measure PAM activity by measuring the amount of glyoxylate produced.

To characterise the temperature and pH optima of uPHM.

To demonstrate the feasibility of synthesising an alpha-amidated peptide pharmaceutical with rhPAM and uPHM using alkaline hydrolysis.

5.2 Material and Methods

5.2.1 Peptidylglycine alpha-amidating monooxygenase activity assay

Assay buffers were selected to prevent interference with redox reactions and minimise chelation with divalent metal ions. This avoids buffer interference by the Cu or Zn co-factors in PHM or PAL. All PHM activity assay solutions used a final concentration of 1.0 µM $CuCl_2$, with 1.0 µM $ZnCl_2$ included for rhPAM solutions. Conjugates of Cu and Zn with sulphate or other salts were avoided to minimise potential interference in the activity solution.

The following reagents were prepared fresh prior to performing the assay: 50 mM ascorbic acid, 0.01 mg/mL catalase, and 0.3% (v/v) H_2O_2. Oxygenation of the activity solution was provided by the addition of 0.5 µg/mL bovine liver catalase (Sigma; C 1345; EC 1.11.1.6). To limit the inhibition caused by Cu and ascorbic acid, 0.003% (v/v) H_2O_2 (Sigma; H1009-500ML) was included to provide a substrate for catalase in the assay reaction solution. A final concentration of 2.5 mM ascorbic acid was used per reaction.

A commercially available variant of recombinant *Homo sapiens* PAM (rhPAM) (RnDsystems 4837-AM-010; Lot number: PMG1020051; 10 µg) used. The rhPAM was supplied in a solution at 0.25 mg/mL and was the only commercially available enzyme with certified activity available for PAM at the time the study was conducted. The rhPAM was diluted per the manufacturer's instructions, and 100 ng of the enzyme was used per 250 µL reaction. The purified recombinant uPHM was expressed and purified as per section 4.2.4, with aliquots of 0.1 to 1.0 mg/mL of uPHM in the protein storage solution and variable quantities of uPHM used in each reaction.

PAM activity units were defined as µM acYVG/ ng PAM/ hour as used for rhPAM. PAL activity units were defined as µM αHA/ ng PAM/ hour as used for PAL activity from rhPAM. PHM activity units were defined as µM αHA/ µg PHM/ hour and were exclusively used for uPHM.

5.2.2 Peptidylglycine alpha-amidating monooxygenase activity assay

Activity assays were performed at 250 µL final volume in a 96-multiwell plate (Star labs, PlateOne®; S1837-9600). Plates were sealed with Parafilm® (Star lab group; I3080-5075) to limit evaporation during incubation. For AGG assays, the sample absorbance was measured at 540 nm, as per section 2.2.2. As the rhPAM had PAL activity, the alkaline hydrolysis step was not performed. The AGG assay, from section 2.2.2, was modified by the addition of 3 M

glycine pH 5.5 to bring the glyoxylate containing sample solution to 0.55 M glycine as the final concentration after the addition of 2-aminobenzaldehyde (2AB).

Prior to UPLC-MS analysis, samples were pooled in 1.5 mL reaction tubes and centrifuged at 20 000 x *g* for 10 minutes, at 4°C, to clarify the solution and pellet the denatured proteins. The supernatant solutions were transferred to glass vials, capped, and frozen for subsequent UPLC-MS analysis, per section 2.2.3. UPLC separation was performed with a Waters UPLC on a Waters HSS T3 C18 column coupled in-series to a Waters SYNAPT G1 HDMS mass spectrometer (MS).

5.2.3 DMPD-based assays of PHM activity

The DMPD-based PHM activity assay is easy to use and has potential general applicability in PAM activity studies. The conditions and concentrations used were based on the original research of Li *et al.*, (1994), but the reaction volume was reduced to 250 µL to enable the use of multiwell plates. The conditions for PHM used with DMPD were the same as those in section 5.2.2, modified by the replacement of ascorbic acid with 2.5 mM DMPD.2HCl (Sigma-Aldrich; 07750-25G). A lower purity spectrophotometric grade ≥ 97% DMPD (Sigma-Aldrich; 07750-25G) was initially used but was later replaced with ≥ 99% DMPD.2HCl (Sigma-Aldrich; 07750-25G) due to solvency problems. The rhPAM or uPHM activity solutions were incubated at 37°C, and the change in absorbance of the chromophore was measured at 515 nm (Hidex Sense Microplate Reader; Beta 425-301).

5.2.4 uPHM deglycosylation

The discrepancy between the predicted and observed masses of uPHM observed in section 4.3.2 could be due to hyperglycosylated expression. Therefore, the deglycosylation of purified uPHM was performed with peptide-N-glycosidase F (PNGase) (Sigma-Aldrich; P7367) from *Elizabethkingia miricola*. The lyophilized PNGase was compounded with buffer salts and was reconstituted with 100 µL deionised water to generate a 5.0 mM potassium phosphate stock, pH 7.5, before aliquotting and storage at -20ºC. PNGase activity was defined as the catalysis of 1 nanomole of denatured substrate protein in 1 minute at 37°C at pH 7.5. To 100 µg of uPHM in 100 µL protein storage solution (section 4.2.4.4), 1 U of PNGase was added, and the solution was incubated at 28°C for 2 hours. The deglycosylated uPHM solution was subsequently used for PHM activity assays (section 5.2.1).

5.3 Results

Previous publications have established that DMPD could measure PAM activity, and their protocols were adapted for use in the present study. The development of a novel assay to measure PAM activity was not in the project's initial aims.

5.3.1 DMPD-based assay to determine PAM activity

Multiple activity studies for rhPAM and purified uPHM were performed with DMPD as an alternate reducing agent. A spectrophotometric grade of ≥ 97% DMPD (Sigma-Aldrich; 07750-25G) was purchased and used for PAM activity assays. The stock solutions of 25 mM DMPD and the final 5.0 mM DMPD in the enzyme activity assay solutions encountered significant solvency issues, with insoluble DMPD particulates present even at final PAM activity concentrations. The insolubility DMPD was abrogated by procuring a higher purity product of ≥ 99% DMPD.2HCl (Sigma-Aldrich; 07750-25G), which was soluble as a 50 mM stock and in the final assay concentrations studied. Despite multiple attempts with both rhPAM and uPHM, no changes in absorbance beyond the background oxidation rate of DMPD were observed. HA and acYVG peptide substrates were tested with DMPD.2HCl, but the absence of quantifiable activity was confirmed by UPLC-MS analysis, with no peptide products detected. Despite multiple attempts to replicate the results with the available commercial recombinant human PAM and multiple peptide substrates, no significant activity was measured with DMPD based assays.

5.3.2 AGG assay to determine rhPAM activity

Based on the AGG assay for glyoxylate, described in section 2.2.2, the activity of rhPAM was determined according to the manufacturer’s instructions. The ability to measure rhPAM activity with the AGG assay would enable the verification of PAM activity assays. The AGG assay and rhPAM were used to establish the pH optima of the recombinant human PAM and PAL. The pH activity study was performed with acYVG and αHA, as these substrates have activity with different catalytic cores of PAM. The rhPAM and acYVG activity measurements are a function of PHM and PAL activity, as PHM hydroxylation is essential for the subsequent PAL activity. PAM activity was defined as one unit (U) equal to one µM of glyoxylate produced from the acYVG peptide substrate per ng of rhPAM per hour (Figure 5.1).

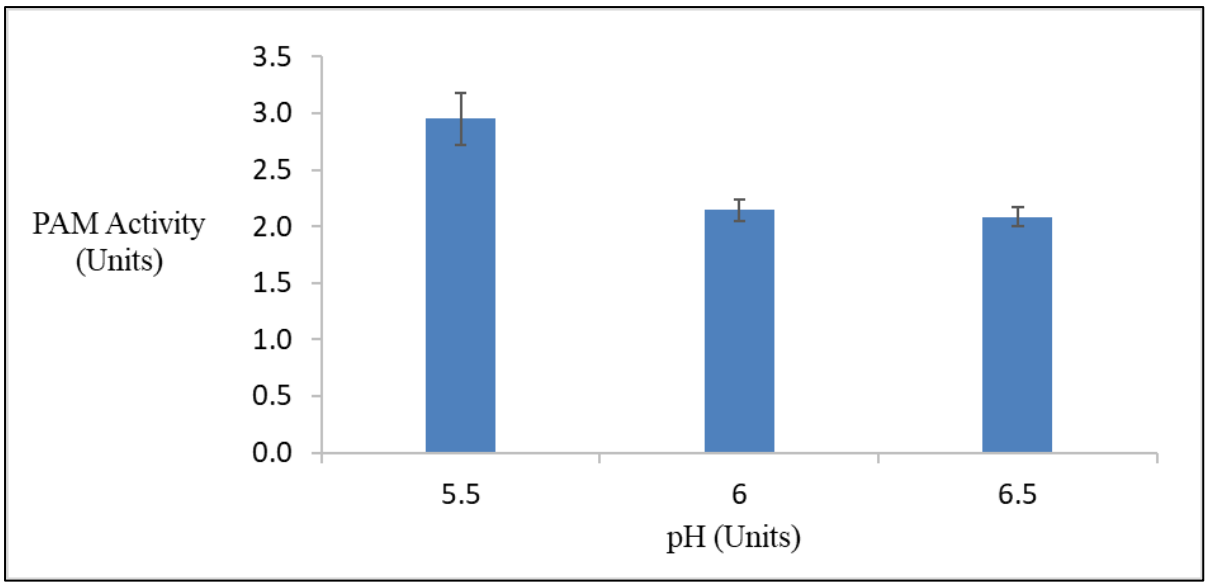


Figure 5.1: The pH range of rhPAM activity with acYVG in 100 mM MES, at pH 5.5, 6.0, and 6.5 at 37°C over 120 minutes (n=3; SD ±).

The pH range tested was limited by the buffering capacity of the MES buffer used, as recommended by the manufacturer, as other buffering reagents are documented to interfere with PAM activity. The activity of rhPAM increased with increasing acidity for the range tested with a maximum of 2.95 U reached at pH 5.5.

The PAL domain activity was independently measured with αHA, as the C-terminus hydroxyl modification was present on the HA peptide substrate. The PAL activity was defined as µM of glyoxylate produced from the αHA peptide substrate per ng of rhPAM per hour. (Figure 5.2). In contrast to the pH range of PHM, PAL activity increased with increasing alkalinity to a maximum activity of 5.63 U at pH 6.5.

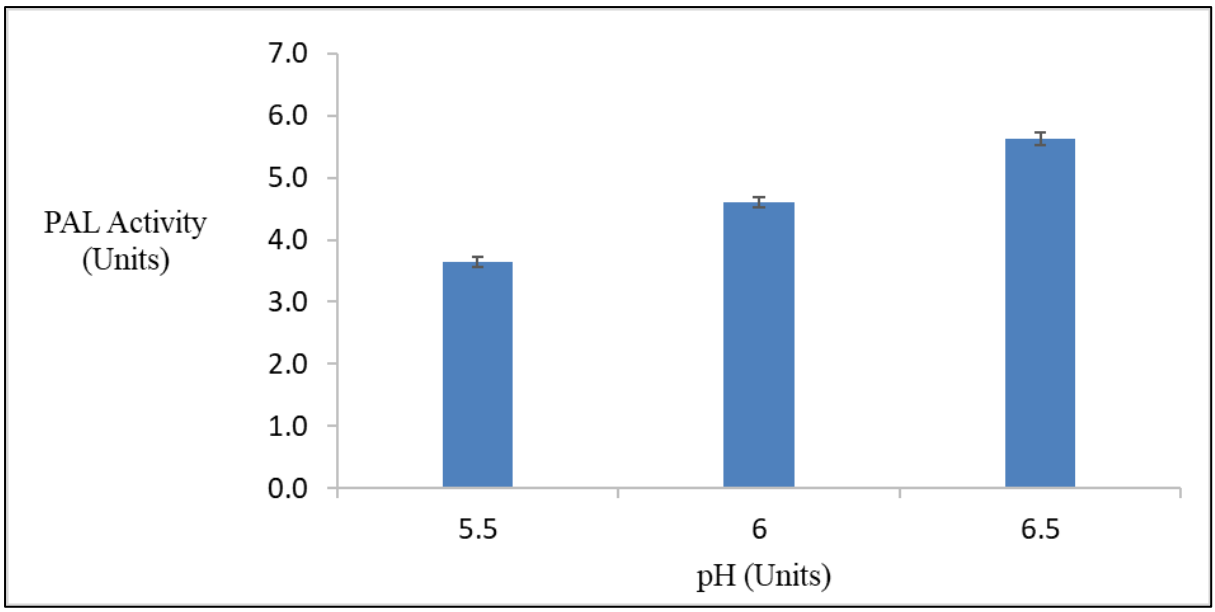


Figure 5.2: The pH range of rhPAM activity with αHA in 100 mM MES, at pH 5.5, 6.0, 6.5 at 37°C after 120 minutes. (n=3; SD ±).

With the PAL pH range established, the conditions used by King et al., 2000, specifically the inclusion of ethanol and Triton X-100, were next investigated. These non-essential reagents were precautionarily excluded from initial activity solutions in case they inhibited or interfered with the PAM activity. The surfactant properties of Triton X-100 and excess oxygenation from catalase created bubbles from the soluble protein in the activity solutions that interfered with absorbance readings. As ethanol increases the rate of AGG chromophore synthesis and subsequent absorbance, the AGG assay of rhPAM with ethanol and Triton X-100 was not performed. Instead, the activity of rhPAM and HA was assessed by measuring the relative peak area of benzamide (122.058 Da; 3.79 min) produced, as per section 2.2.3 (Figure 5.3).

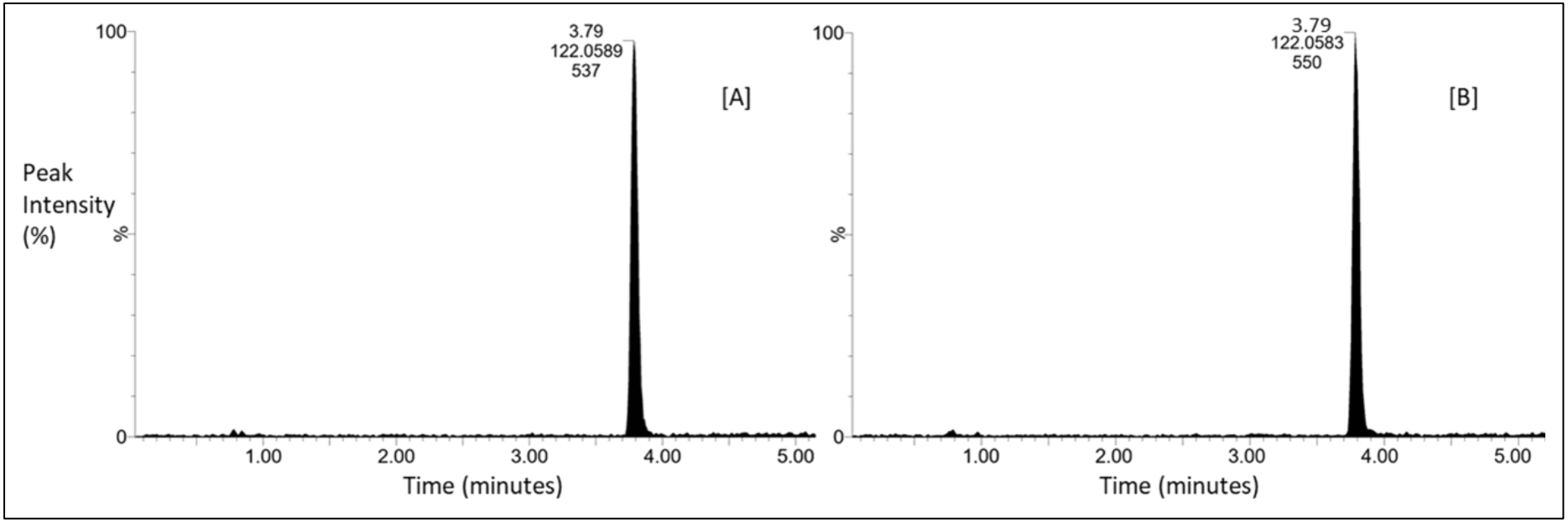


Figure 5.3: Chromatograms of the activity of rhPAM with 3.3 mM HA in 50 mM MES, at pH 6.0 and 37°C [A] with 1% (v/v) ethanol and [B] 0.001% (v/v) Triton X-100. The benzamide peaks (122.058 Da) at 3.79 min were integrated in ESI positive mode.

Based on this analysis, the relative percentage difference of peak area size was 2.36%, which did not merit the inclusion of ethanol or Triton X-100 in subsequent assays. As H_2O_2 is a substrate for catalase, it was used instead of ethanol to avoid any denaturing effects of ethanol.

Since the temperature optimum of rhPAM was assumed to be 37°C, this offered an opportunity to demonstrate that the AGG assay could be used with UPLC-MS to confirm this property. The human PAL could be tested independently of the PHM catalytic core by measuring the activity with αHA as the peptide substrate, according to the amount of AGG chromophore produced (235.07 Da; 0.95 min) (Figure 5.4).

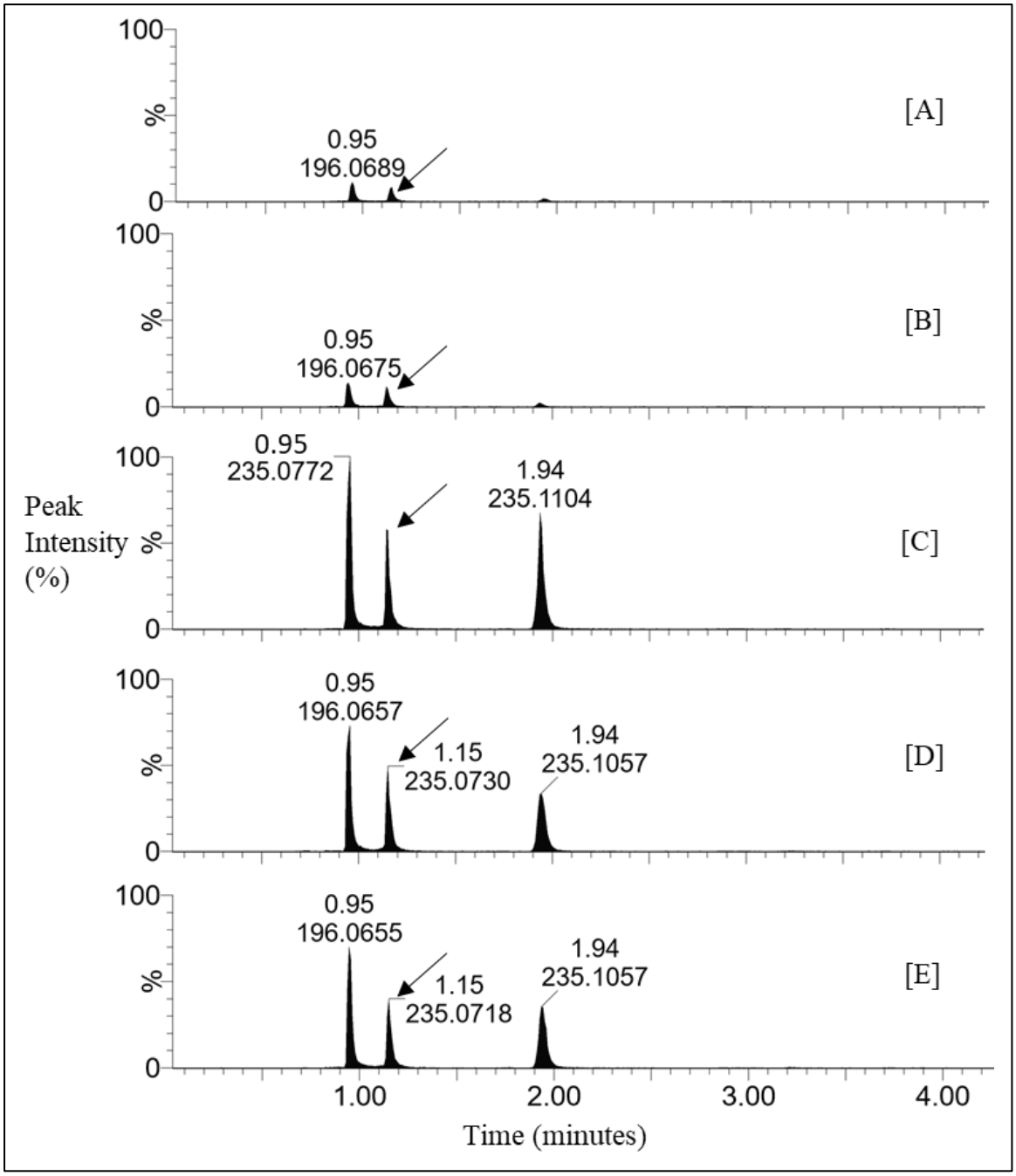


Figure 5.4: The temperature range of rhPAM activity with 3.3 mM αHA in 50 mM MES at pH 6.0 and 50°C [A], 45°C [B], 37°C [C], 28°C [D], and 20°C [E]. ESI negative mode stacked chromatograms of the AGG chromophore (235.07 Da) at 1.15 min are indicated with arrows.

The AGG chromophore synthesis reaction produced three peaks averaging 235.1 Da, in Figure 5.4 [C], at 0.95, 1.15, and 1.94 min. The presence of multiple buffers and reagents in the AGG assay results in peaks that could overlap with the analyte of interest. At lower concentrations, the initial AGG chromophore peak at 0.95 min (235.07 Da) often overlapped with buffer peaks, or other contaminating analytes. The second peak at 1.15 min (235.07 Da) did not have interfering peaks and offered a better resolution of the AGG chromophore. To minimise peak

overlap, only the AGG chromophore eluted at 1.15 min was used to measure glyoxylate concentrations and thus PAM activity.

The co-elution of analyte peaks was further mitigated by filtering the MS data to extract the peaks for the AGG chromophore or enzyme products. In Figure 5.4, the mass of 235.1 Da was extracted to simplify the chromatogram and accurately quantify rhPAM activity. The AGG synthesis reaction produces an AGG chromophore with a mean mass of 235.1 Da that elutes at multiple retention times, as seen in section 2.3.2 Figure 2.8, despite having the same bonds and chemical formulae.

The optimum temperature for PAL activity of the rhPAM enzyme was 37°C, according to the quantity of AGG chromophore produced at 1.15 min. The initial AGG chromophore eluted at 0.95 min, and was annotated as 196.06 Da in [A], [B], [C], and [D]. The different masses observed for the peak at 0.95 minutes were a mean of all the ion masses at 0.95 min by the MassLynx software. When the analytes detected at 0.95 min were examined in fragment ion mode, the predominant peak was identified as the AGG chromophore at 235.0734 Da. For the extracted mass peaks at 0.95 min, the variable ratio and concentration of ions masses and fragments changed the calculated average mass for the peak. The presence of buffers, substrates, and other interfering analytes at 0.95 min meant that the AGG chromophore could not be used to measure glyoxylate concentrations and PAM activity at this retention time.

5.3.3 Determination of uPHM activity by UPLC and AGG assay

The activity of uPHM was not detectable with the AGG assay due to its low specific activity and was therefore instead measured using UPLC. Unlike rhPAM, 10 µg of uPHM was used per reaction. In parallel to the determination of pH optimum, the temperature optimum of uPHM was determined with acYVG at pH 5.0 based on previous protocols developed for rhPAM activity assays. The higher temperature optimum required for uPHM would not be compatible with the bovine catalase, which degrades with prolonged incubation at temperatures above 37°C. The temperature stability of catalase, therefore, limited uPHM activity assays to 37°C. When the assay was performed at 45°C, fresh catalase was added every 30 minutes for the duration of the activity assay (Figure 5.5).

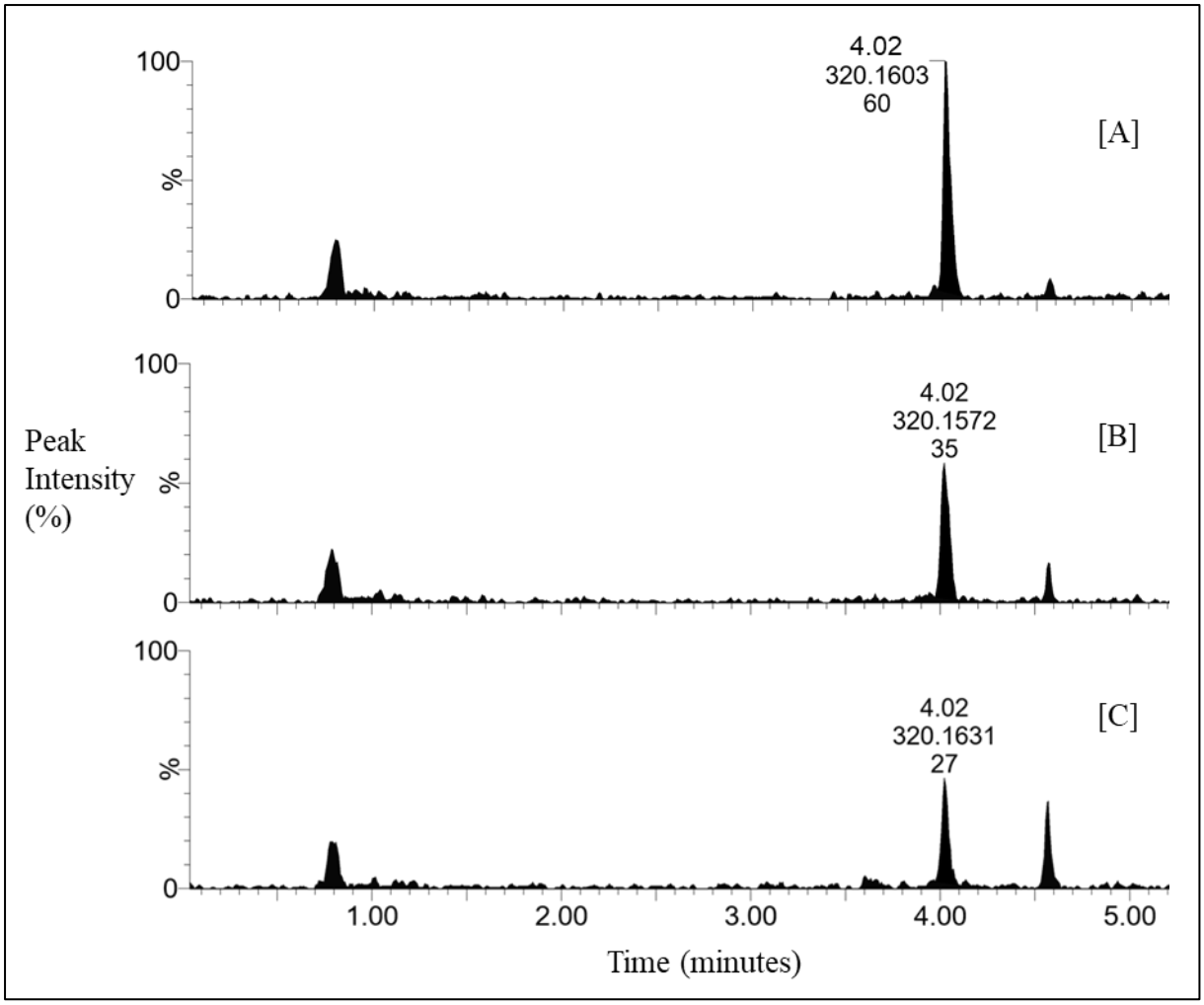


Figure 5.5: The temperature range of uPHM with 1 mM acYVG in 50 mM succinate, at 45°C [A], 37°C [B], and 20°C [C] at pH 5.0, for 60 min. ESI negative mode of the stacked and integrated chromatograms of acYV-NH_2 (320.16 Da) at 4.02 min are indicated.

As the source organism, *Cyphellophora europaea*, is a mesophilic fungus, the optimum temperature for all its endogenous enzymes was expected to be 37°C. However, the largest peak area for acYV-NH2 was observed at 45°C activity assay. As glycosylation can confer temperature stability to an enzyme, the effect of deglycosylation was assayed, along with an untreated uPHM for comparison (Figure 5.6).

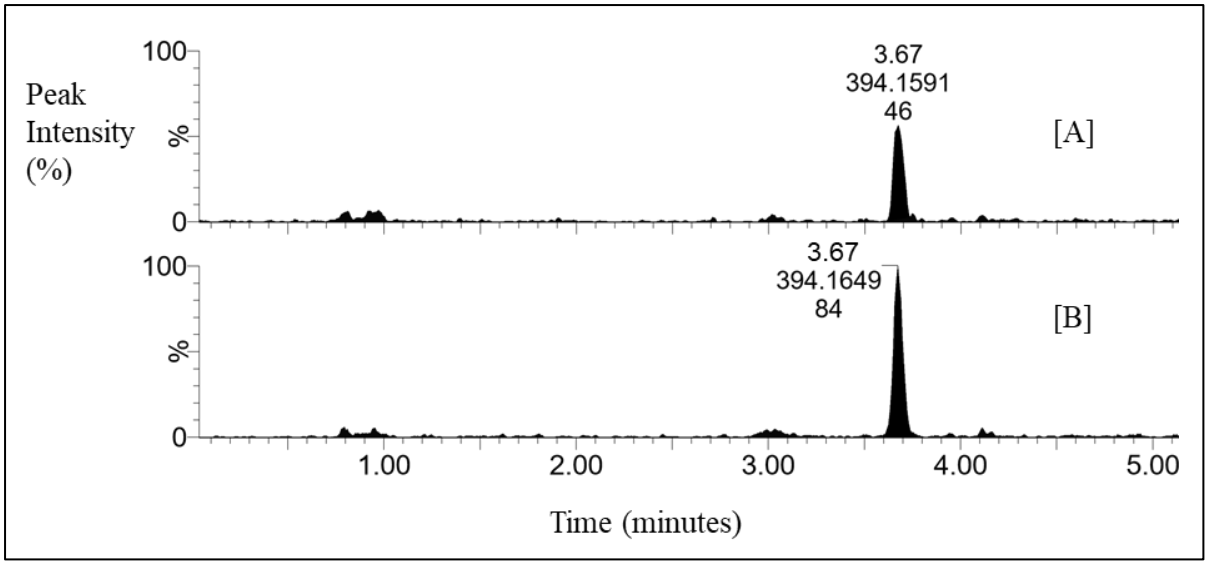


Figure 5.6: The effect of deglycosylation on uPHM with 1 mM acYVG in 50 mM succinate pH 5.0 for hyperglycosylated uPHM at 45°C [A], and deglycosylated uPHM at 37°C [B]. ESI negative mode of the stacked and integrated chromatograms of acYVG-OH (394.16 Da) at 3.67 min is indicated.

At the time that the deglycosylation experiment was performed, the final procedure for alkaline hydrolysis was not developed, and the hydroxylated product was monitored as a reference for activity. A sample of deglycosylated uPHM was incubated at 45°C, but no hydroxylated or amidated peptide product was detected. The larger peak measured for the deglycosylated uPHM assay, despite the lower temperature of 37°C, indicated that the glycosylation on uPHM both prevented it from denaturing at 45°C, but also likely inhibited activity of the glycosylated uPHM.

The activity assays were subsequently performed at 37°C to limit catalase degradation and prevent potential uPHM degradation. The conditions for uPHM could be optimised further by determining the optimum pH range for this enzyme. The activity for uPHM was measured by determining the peak area of the acYVG-OH (394.16 Da; 3.72 min) product from a range of uPHM activity solutions (Figure 5.7).

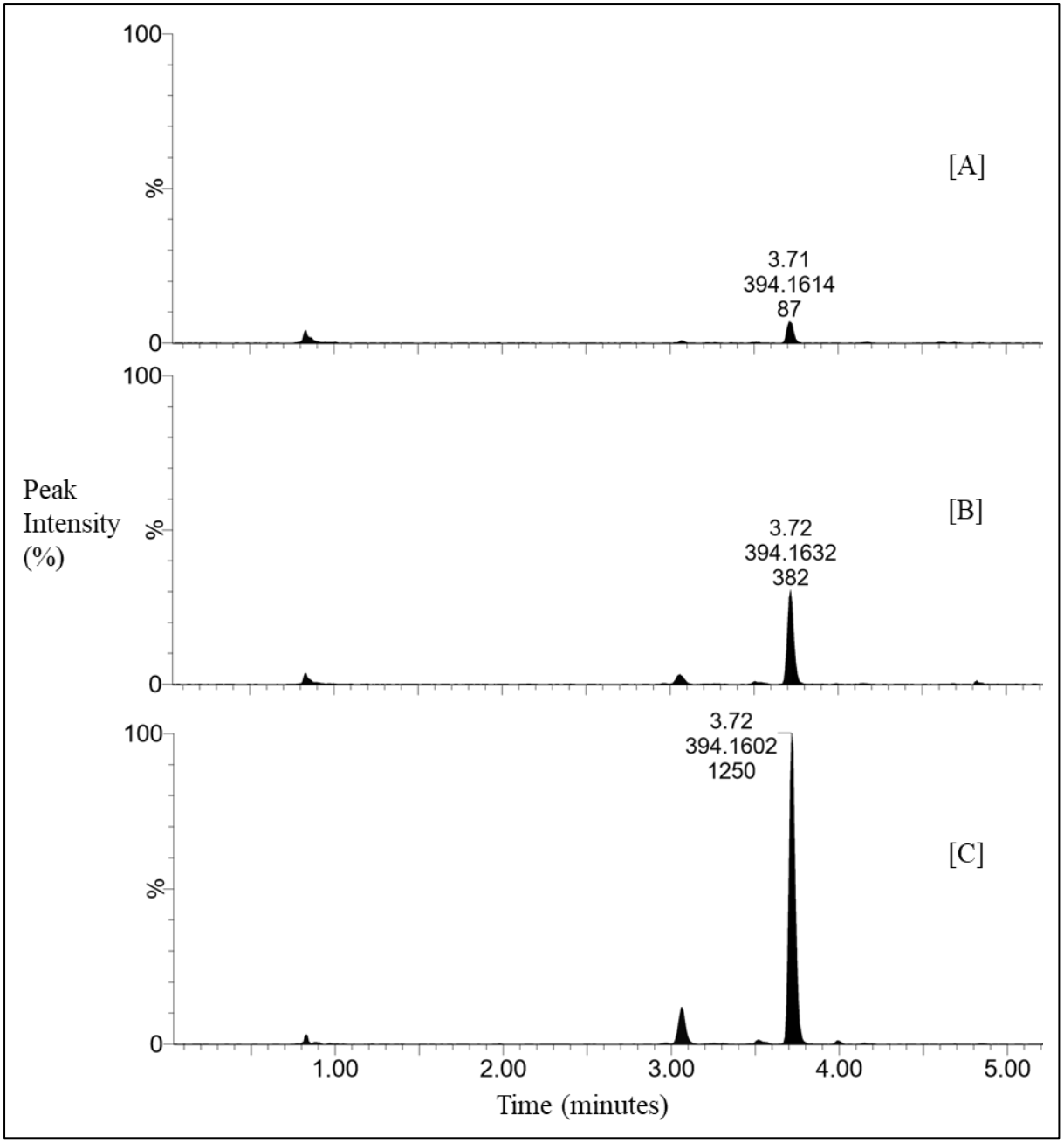


Figure 5.7: The pH range of uPHM activity with 1 mM acYVG in 100 mM succinate, at pH 6.0 [A], 5.0 [B]; 4.0 [C] at 37°C. Chromatograms in ESI negative mode of stacked and integrated peaks of acYVG-OH (394.16 Da) at 3.72 min.

The maximum activity for uPHM was observed at pH 4.0. The initial iteration of the AGG assay, with alkaline hydrolysis and conversion of the hydroxylated peptide to the amidated product, was limited because incubation at 37°C with only 150 mM glycine added was unlikely to result in a pH of 10.0. The resulting solution was closer to a neutral pH, and the hydrolysis was negligible, and no significant absorbance readings were measured with the AGG assay. The samples' non-integrated chromatograms showed minor quantities of AGG chromophore and acYV-NH_2, implying the alkaline hydrolysis step had not converted all the acYVG-OH

produced. The alkaline glycine concentration was increased to 220 mM and the pH adjusted to 10.5 with 5 µL of 3 M NaOH to ensure alkaline conditions for hydrolysis. The addition of acidic glycine and 5 µL of 1.5 M H_2SO_4 resulted in a final glycine concentration of 500 mM at pH 5.0 for the AGG chromophore synthesis reaction.

The acidic pH optimum of uPHM required careful selection of an appropriate buffer system. The buffer reagent had to have minimal interference with the redox reaction of PHM and minimal metal chelation properties. While having other prerequisite properties, the previously used buffers, MES and MOPS, could not effectively buffer in the desired pH range of 4.0 to 5.0. A set of buffers that could be suitable for uPHM activity were therefore tested. As phosphate works well with the catalytic cores of PAM and PHM and given its common usage, it was included in the buffers studied (Figure 5.8).

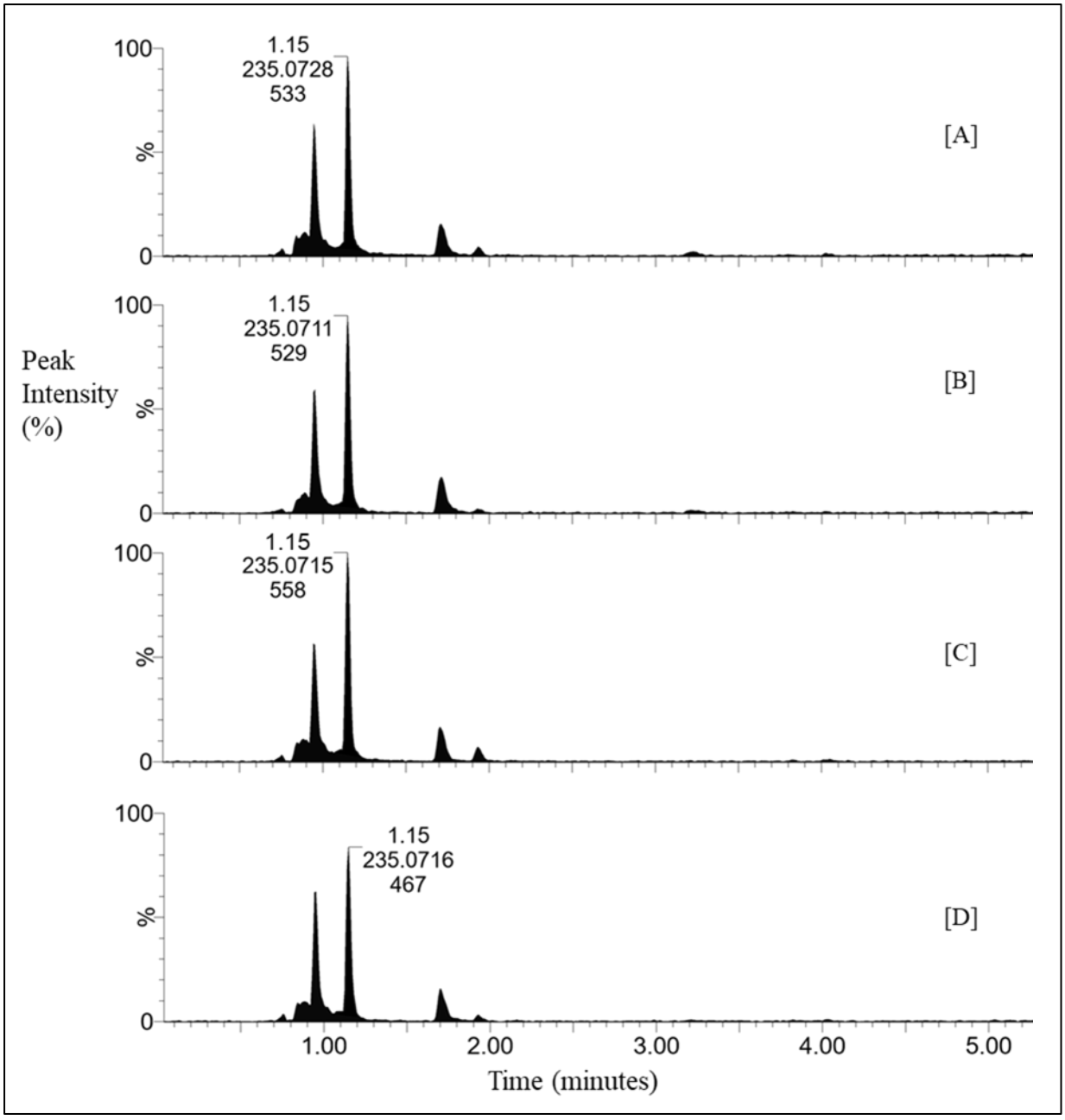


Figure 5.8: The buffer specificity of uPHM with 1 mM acYVG in 30 mM, pH 4.0 [A] succinate, phosphate [B]; citrate [C]; and acetate [D] buffers at 37°C. Chromatograms in ESI negative mode with the stacked and integrated chromatograms of the AGG chromophore (235.07 Da) at 1.15 min are indicated.

There were minimal differences in the observed peak area for each different buffer tested, except for acetate buffer, which indicates that these buffers had minimal effects on uPHM activity. As succinate buffer was compatible with uPHM activity and had minimal metal chelation properties, it was used in subsequent studies.

With the essential characteristics of uPHM activity measured, the ability of uPHM to produce alpha-amidated peptides could be assessed under optimal conditions. Due to the cost and low

molar quantity of substrate hydroxylated glycine at the peptide C-terminus relative to the high molar mass of exenatide (4186.63 Da), a lower molar concentration of 0.25 mM was used for the peptide substrate. The final protein concentration of 0.418 mg/mL of exenatide per reaction, in addition to the other proteins and reagents in solution, represented the limits of protein solubility with the present protocol. The high molar mass ratio for glycine-residue substrate sites to the high mass of substrate peptides meant that less than 0.25 mM of AGG chromophore was likely to be synthesised. Since this represents the lower limit of detection for the AGG assay, the solution was analysed with UPLC-MS to measure the AGG chromophore accurately. Both uPHM and rhPAM activity assays were performed with the C-terminal glycine-extended exenatide to demonstrate that alpha-amidated pharmaceutical peptide production is possible. The products from each reaction were measured with the AGG assay and UPLC-MS (Figure 5.9).

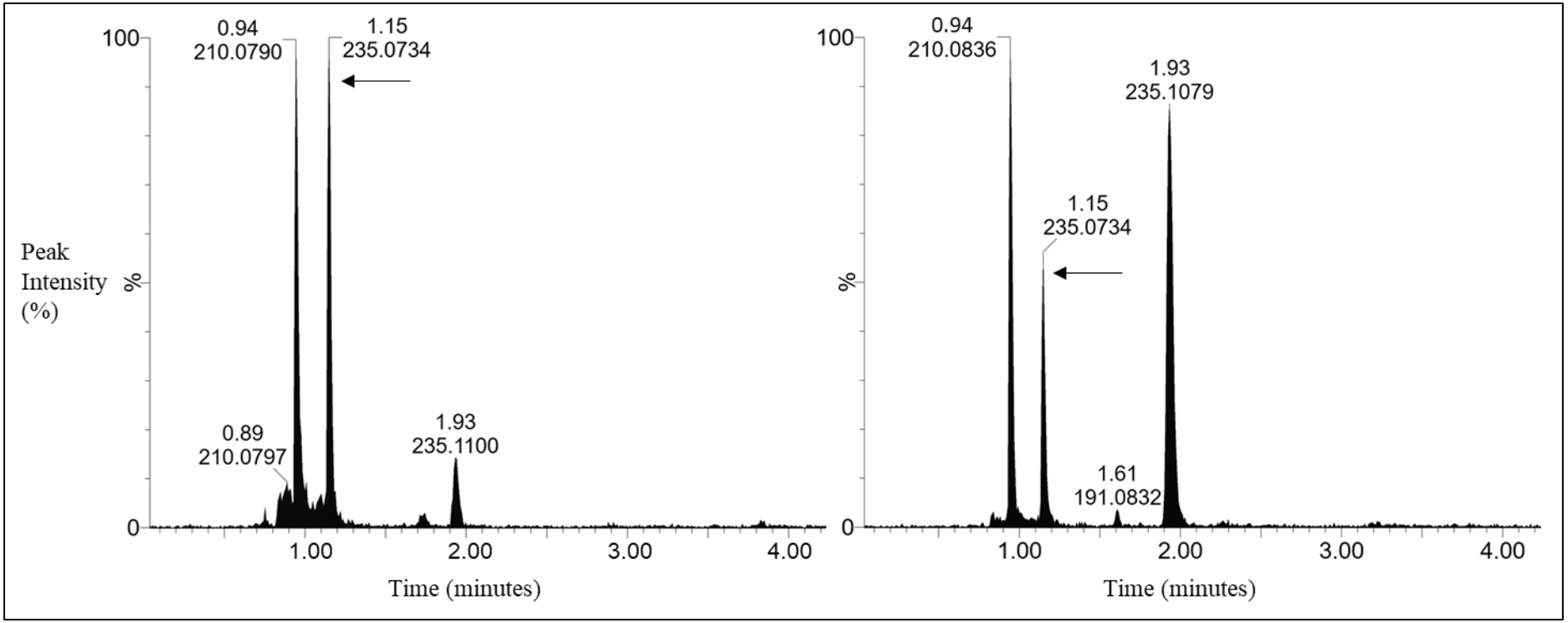


Figure 5.9: Activity at 37°C with 0.25 mM exenatide with a C-terminal glycine, with uPHM in 32 mM succinate, pH 4.0 [left], and rhPAM in 32 mM MES, pH 6.0 [right]. The chromatogram in ESI negative mode of the integrated peaks of the AGG chromophore (235.07 Da) at 1.15 min is indicated.

The synthesis of the AGG chromophore from glyoxylate by rhPAM and uPHM from the C-terminus glycine-extended exenatide demonstrates that uPHM can amidate peptides at rates similar to those of rhPAM. Performing the assay with uPHM required the alkaline hydrolysis step but demonstrated that uPHM was able to generate the final bio-active peptide at equivalent concentrations to rhPAM.

5.4 Discussion

The ability to measure PHM activity via absorbance measurements with DMPD would provide a high throughput method for kinetics studies. However, the minor increases in absorbance in the presence of PAM or uPHM may have been due to non-specific redox reactions between DMPD and copper ions at the active site. The reaction solution saturation with O_2 from catalase activity and high concentrations of Cu would have increased the auto-oxidation rate of DMPD, explain the increasing absorbance readings measured. The development of the DMPD assay should be attempted in future but was superseded in the present study by the reliability of the AGG assay and the sensitivity of UPLC-MS.

No activity was measurable by UPLC-MS or AGG assays with HA as the peptide substrate for uPHM. This apparent inactivity was unexpected as HA is a standard and inexpensive PHM substrate. A function of the strict PHM substrate specificity for uPHM is potentially screening substrates for subsequent PAL activity, as C-terminal glycine α-carbon hydroxylation is a prerequisite for PAL activity. Conversely, rhPAM appeared to have a lower substrate specificity for small peptides or amino acid analogues, as seen by its activity with both HA and αHA. A potential explanation for the observed stricter PHM substrate specificity is the proposed extracellular location of uPHM, predicted by its wild type gene containing a DOMON domain. Any PHM with a lower substrate specificity would interact with a wider variety of substrates and conditions than the controlled environment required for intracellular enzymes.

A potential metabolic explanation for the strict substrate specificity of PHM is the conservation of multiple co-factors required for PHM activity and preservation of its limited activity only for specific peptide substrates. The existence of numerous co-factors for PHM means that non-specific PHM activity would carry a higher biological cost for uPHM as an extracellular protein, as cells would have to provide the necessary co-factors into their immediate environment to enable PHM activity. These costs would place an evolutionary pressure on the PHM gene for greater substrate specificity to limit wasted activity. The intracellular location of rhPAM places different restraints on the enzyme, and the small peptides analogues it interacts with form an essential class of cellular signalling agents.

The lower activity of PHM relative to PAL in rhPAM has been observed with recombinantly expressed PHM and PAL catalytic cores for all pH conditions tested. At acidic pHs of 5.5 or lower, the topological location in late-stage Golgi vesicles would imply that PAL would also have an acidic pH optimum. The higher turnover rate of PAL compensates for the non-optimal

pH conditions and could explain the different pH optima for PHM and PAL. The turnover rate of multiple isoforms of PAL is higher than that of PHM, owing to the latter's requirement of co-enzymes and regeneration between active states (Eipper *et al.*, 1991; Chufán *et al.*, 2009). The pH values tested with MES and rhPAM in the present study were limited compared to other buffer systems due to the use of MES buffer. The buffer reagent used and activity assay parameters were according to the manufacturer's instructions. The maximum PAL activity of rhPAM at pH 6.5 would represent the actual pH optimum tested for PAL, but the pH 5.5 optimum observed with PAM represents a combination of the pH optima of PHM and PAL. Determining the true pH optima for PHM is further complicated by the differences in turnover rate between PAL and PHM, with PAL having a faster turnover rate than PHM.

The pH optima of PHM and PAL could be substrate-dependent, as they each possess distinct substrate-binding potential and may display different enzyme kinetics with different substrates. The requirement for PHM to hydroxylate the substrate, which then must diffuse to the active site of PAL before initiating PAL activity, further complicates these measurements. The higher activity recorded for the PAL domain with αHA compared to rhPAM with acYVG must be due to differences in activity levels, as the quantities of enzymes and substrates were kept constant. As PAM activity is the sum of PHM and PAL activity, the lower level of activity for PAM measured confirms that PHM activity is the rate-limiting step in the PAM reaction, with lower activity of PHM observed relative to that of PAL.

The minimal activity differences observed between different PAM activity solutions containing Triton X-100 or ethanol established that these additional reagents were unnecessary. Triton was included to stop non-specific absorption to the oxygen chamber vessel walls and was superfluous for the absorbance assays or UPLC-MS analysis. The catalase substrate ethanol was substituted for H_2O_2 at a lower molar concentration to avoid potential ethanol denaturation of uPHM.

An advantage of the AGG assay was its dual-use for initial absorbance measurements and subsequent UPLC-MS analysis if greater sensitivity was required to detect sub-µM glyoxylate concentrations. The AGG chromophore synthesis enabled the measurement of glyoxylate produced by PAM activity, which was otherwise not detectable by UPLC-MS. The higher mass of the exenatide and related polypeptides >20 amino acids in length would require longer run times for UPLC-MS to detect products and substrates, which could reduce an assay's throughput. The temperature range of rhPAM and the optimum at 37°C were expected, but the

results confirmed the accuracy of the AGG assay and UPLC-MS protocol for determining uPHM activity. The use of these techniques for measuring the activity of enzymes and applications that require the measurement of glyoxylate would prove beneficial in many fields.

The AGG assay was initially intended to be a medium-throughput absorbance assay for PHM activity, but the low specific activity of uPHM frustrated efforts to it as such. The AGG assay can provide a higher throughput measurement of PAM activity, as demonstrated by successfully measuring rhPAM in the present study. The limited sensitivity of the AGG assay meant that uPHM characterisation was scaled back and performed with UPLC-MS. Despite these setbacks, the characterisations of essential enzyme properties were performed, and the pH optimum of uPHM was determined to be 4.0. The acidic uPHM activity pH optimum was problematic for constant catalase activity and required pH adjustments for the alkaline hydrolysis of glyoxylate. While laborious, the pH adjustments were effective, provided a strong base was added, and the reaction solution had lower buffer concentrations. However, these alterations required measuring the activity assay and AGG solution pH at each stage of the AGG assay. The catalase was able to oxygenate the solution at pH 4.0 by including H_2O_2 to counteract lower activity levels. The temperature optimum of 45°C for uPHM meant that fresh catalase had to be added every 30 minutes, as it degraded rapidly at this temperature.

Further studies with uPHM at higher temperatures will only be possible if a thermostable catalase is used or if an alternative oxygenation method is developed. At 45°C, the uPHM could be an appropriate industrial enzyme, but the higher temperature stability may be due to its glycosylation, as the removal of glycosylation did increase activity, but lowered the temperature stability of the enzyme. Future studies may include a deglycosylation step during purification or mutation of the *uPHM* sequence to remove surface glycosylation sites. The inclusion of deglycosylation will depend on the desired reaction conditions and will be particularly useful if a higher turnover but less temperature-stable PHM is desired. The amino acid sequence alteration could be deleterious to uPHM expression as glycosylation can assist in correct protein folding and processing during translation.

Since citrate chelates divalent metal cations, it was unexpected that citrate buffer would support the highest relative activity of uPHM. The metal binding affinity of the active site residues of uPHM should have a higher metal binding coefficient than citrate. uPHM would be amenable to use with all the buffers test at its pH optimum of pH 4.0. The production of amidated

exenatide with uPHM at similar levels to rhPAM proves that it can be used for the production of amidated peptide pharmaceuticals.

5.5 References

Alwan, K. B., Welch, E., Arias, R., Gambill, B. and Blackburn, N. (2019) 'Rational design of a histidine–methionine site modeling the M-center of copper monooxygenases in a small metallochaperone scaffold', *Biochemistry*, 58(28), pp. 3097–3108. doi: 10.1021/acs.biochem.9b00312.

Bradbury, A. F., Finnie, M. D. A. and Smyth, D. G. (1982) 'Mechanism of C-terminal amide formation by pituitary enzymes', *Nature*, 298(5875), pp. 686–688. doi: 10.1038/298686a0.

Bradbury, A. F. and Smyth, D. G. (1987) 'Enzyme-catalysed peptide amidation. Isolation of a stable intermediate formed by reaction of the amidating enzyme with an imino acid', *European Journal of Biochemistry*, 169(3), pp. 579–584. doi: 10.1111/j.1432-1033.1987.tb13648.x.

Cao, F., Gamble, A., Kim, H., Onagi, H., Gresser, M., Kerr, J., and Easton, C. (2011) 'Potent and selective inhibitors of human peptidylglycine α-amidating monooxygenase', *MedChemComm*, 2(8), p. 760. doi: 10.1039/c1md00079a.

Carpenter, S. (2006) Enzyme linked spectroscopic assays for glyoxylate: The use of peptidylglycine alpha-amidating monoxygenase for the discovery of novel alpha-amidated hormones. PhD thesis. University of South Florida. Available at: http://scholarcommons.usf.edu/etd/2472/

Chance, B. (1952) 'Effect of pH upon the reaction kinetics of the enzyme-substrate compounds of catalase', *Journal of Biological Chemistry*, 194(2), pp. 471–481. doi: 10.1016/S0021-9258(18)55799-9.

Chauhan, S., Hosseinzadeh, P., Lu, Y. and Blackburn, N. (2016) 'Stopped-flow studies of the reduction of the copper centers suggest a bifurcated electron transfer pathway in peptidylglycine monooxygenase', *Biochemistry*, 55(13), pp. 2008–2021. doi: 10.1021/acs.biochem.6b00061.

Christman, A. A., Foster, P. W. and Esterer, M. B. (1944) 'The allantoin content of blood', *Journal of Biological Chemistry*, 155(1), pp. 161–171. doi: 10.1016/S0021-9258(18)43183-3.

Chufán, E. E., De, M., Eipper, B., Mains, R., and Amzel, L. (2009) 'Amidation of bioactive peptides: the structure of the lyase domain of the amidating enzyme', *Structure*, 17(7), pp. 965–973. doi: 10.1016/j.str.2009.05.008.Amidation.

Eipper, B. A., Perkins, S., Husten, E., Johnson, R., Keutmann, H. and Mains, R. (1991) 'Peptidyl-alpha-hydroxyglycine alpha-amidating lyase. Purification, characterization, and expression.', *The Journal of Biological Chemistry*, 266(12), pp. 7827–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1902227.

Gholamian, F., Sheikh-Mohseni, M. A. and Naeimi, H. (2012) 'Simultaneous determination of phenylhydrazine and hydrazine by a nanostructured electrochemical sensor', *Materials Science and Engineering: C*, 32(8), pp. 2344–2348. doi: 10.1016/j.msec.2012.06.020.

Glembotski, C. C., Eipper, B. A. and Mains, R. E. (1984) ‘Characterization of a peptide alpha-amidation activity from rat anterior pituitary.’, *The Journal of Biological Chemistry*, 259(10), pp. 6385–6392. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6725255.

Handa, S., Spradling, T., Dempsey, D. and Merkler, D. (2012) ‘Production of the catalytic core of human peptidylglycine α-hydroxylating monooxygenase (hPHMcc) in *Escherichia coli*’, *Protein Expression and Purification*, 84(1), pp. 9–13. doi: 10.1016/j.pep.2012.04.012.

Katopodis, A., Ping, D., Smith, C. and May, S. (1991) ‘Functional and structural characterization of peptidylamidoglycolate lyase, the enzyme catalyzing the second step in peptide amidation’, *Biochemistry*, 30(25), pp. 6189–6194. doi: 10.1021/bi00239a016.

Katopodis, A. G. and May, S. W. (1990) ‘Novel substrates and inhibitors of peptidylglycine .alpha.-amidating monooxygenase’, *Biochemistry*, 29(19), pp. 4541–4548. doi: 10.1021/bi00471a006.

Katopodis, A. G., Ping, D. and May, S. W. (1990) ‘A novel enzyme from bovine neurointermediate pituitary catalyzes dealkylation of alpha-hydroxyglycine derivatives, thereby functioning sequentially with peptidylglycine .alpha.-amidating monooxygenase in peptide amidation’, *Biochemistry*, 29(26), pp. 6115–6120. doi: 10.1021/bi00478a001.

King L. 3rd, Barnes, S., Glufke, U., Henz, M. E., Kirk, M., Merkler, K. A., Vederas, J. C., Wilcox, B. J. and Merkler, D. J. (2000) ‘The enzymatic formation of novel bile acid primary amides.’, *Archives of Biochemistry and Biophysics*, 374(2), pp. 107–117. doi: 10.1006/abbi.1999.1611.

Kline, C. D., Mayfield, M. and Blackburn, N. J. (2013) ‘HHM Motif at the CuH-site of dependent conformational switch’, *Biochemistry*, 52(15), pp. 2586–2596. doi: 10.1021/bi4002248.

Li, C., Oldham, C. D. and May, S. W. (1994) ‘NN-Dimethyl-1,4-phenylenediamine as an alternative reductant for peptidylglycine α-amidating mono-oxygenase catalysis’, *Biochemical Journal*, 300(1), pp. 31–36. doi: 10.1042/bj3000031.

Merkler, D. J., Asser, A., Baumgart, L., Carballo, N., Carpenter, S., Chew, G., Cosner, C., Dusi, J., Galloway, L., Lowe, A., Lowe, E., King, L., Kendig, R., Kline, P., Malka, R., Merkler, K., McIntyre, N., Romero, M., Wilcox, B. and Owen, T. (2008) ‘Substituted hippurates and hippurate analogs as substrates and inhibitors of peptidylglycine α-hydroxylating monooxygenase (PHM)’, *Bioorganic & Medicinal Chemistry*, 16(23), pp. 10061–10074. doi: 10.1016/j.bmc.2008.10.013.

Münch, J., Püllmann, P., Zhang, W. and Weissenborn, M. (2021) ‘Enzymatic hydroxylations of sp3-carbons’, *ACS Catalysis*, pp. 9168–9203. doi: 10.1021/acscatal.1c00759.

Orr, C. W. M. (1967) ‘Studies on ascorbic acid. II. Physical changes in catalase following incubation with ascorbate or ascorbate and copper(II)’, *Biochemistry*, 6(10), pp. 3000–3006. doi: 10.1021/bi00862a005.

Schomburg, D., Salzmann, M. and Stephan, D. (1994) ‘Catalase’, in *Enzyme Handbook 7*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 741–746. doi: 10.1007/978-3-642-78521-4_141.

Schomburg, D. and Stephan, D. (1994) *Enzyme Handbook*. Edited by D. Schomburg and D. Stephan. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-57942-4.

Shimoi, H., Kawahara, T., Suzuki, K., Iwasaki, Y., Jeng, A. and Nishikawa, Y. (1992) ‘Characterization of a *Xenopus laevis* skin peptidylglycine alpha-hydroxylating monooxygenase expressed in insect-cell culture’, *European Journal of Biochemistry*, 209(1), pp. 189–194. doi: 10.1111/j.1432-1033.1992.tb17276.x.

Tajima, M., Iida, T., Yoshida, S., Komatsu, K., Namba, R., Yanagi, M., Noguchi, M. and Okamoto H. (1990) ‘The reaction product of peptidylglycine alpha-amidating enzyme is a hydroxyl derivative at alpha-carbon of the carboxyl-terminal glycine.’, *Journal of Biological Chemistry*, 265(17), pp. 9602–9605. doi: 10.1016/S0021-9258(19)38709-5.

Trendel, J. A., Ellis, N., Sarver, J., Klis, W., Dhananjeyan, M., Bykowski, C., Reese, M. and Erhardt, P. (2008) ‘Catalytically active peptidylglycine α-amidating monooxygenase in the media of androgen-independent prostate cancer cell lines’, *Journal of Biomolecular Screening*, 13(8), pp. 804–809. doi: 10.1177/1087057108321976.

Wimalasena, K. and Wimalasena, D. (1991) ‘Continuous spectrophotometric assays for dopamine β-monooxygenase based on two novel electron donors: N,N-dimethyl-1,4-phenylenediamine and 2-aminoascorbic acid’, *Analytical Biochemistry*, 197(2), pp. 353–361. doi: 10.1016/0003-2697(91)90404-H.

Chapter 6: Discussion, conclusions, and recommendations

6.1 Discussion

The objectives of the present study were to express PAM, or the essential PHM domain, to enable a viable method of biological peptide synthesis for α-amidated peptides within a green chemistry framework. The two major obstacles encountered were expressing PAM or PHM active protein and developing a suitable, reliable assay to measure enzyme activity. The AGG assay offered significant improvements in terms of ease of use and reliability for measuring glyoxylate generated from PAM or PAL activity. The added alkaline hydrolysis step expanded the assay to enable accurate measurement of PHM activity. The expression and confirmation of PHM activity from the novel uPHM was the other crucial objective but required multiple sequences from diverse source organisms and truncations thereof. The AGG assay provided critical measurements of uPHM activity, which proved that it could be used for producing α-amidated peptides.

The AGG chromophore enabled micromolar absorbance readings of PAM activity and was compatible with subsequent UPLC-MS analysis, thus providing increased sensitivity as needed. This combination of techniques could replace radioisotope-labelled peptide substrates since it outperforms them based on cost, safety and sensitivity. The higher throughput achieved using the chromophore AGG offers multiple routes for improving PAM assays and the AGG assay. The optimum conditions for the alkaline hydrolysis with hydroxylated peptide substrates will require further investigation. The current alkaline conditions of pH 10.0 at 37°C for 45 minutes were effective for the complete hydrolysis of the glyoxylate residue from αHA.

Nevertheless, a higher temperature of 50°C or 60°C with 30 minutes or longer incubation would likely ensure total conversion of longer glycine-extended peptide substrates to amidated peptide products. It is recommended that the C-terminus-: glycine extended, α-carbon hydroxylated, and amidated peptides are procured for all peptides of interest for studies relating to alkaline hydrolysis. The lack of uPHM activity with HA frustrated initial attempts with the AGG assay and HA.

As observed by the dual peaks present in the UPLC-MS chromatograms, the AGG chromophore isomers should be further studied. These peaks could be separated under the

chromatographic methods in the present study, and their different properties could be of interest, particularly for fluorescence and absorbance spectrums. The AGG assays with the readily expressible PHM in *Yarrowia lipolytica* also creates a platform for improving PHM activity.

An improvement to the current methods would be to add a linker region with a protease recognition site before the His-tag purification sequence. The linker would enable the removal of the purification tag, with the uPHM selectively eluted from a nickel affinity column without high concentrations of imidazole that could chelate the active site copper ions. This change to the purification procedure would entail a longer immobilised metal affinity chromatography step. Still, it would reduce the multiple rounds of tangential flow filtration required to remove imidazole and exchange buffers. Introducing this change will likely decrease production time and costs if the enzyme is compatible with these purification techniques. A commonly used protease recognition site would also enable inexpensive commercial antibodies to detect the recombinant protein-expressing colonies and monitor the enzyme during the purification process. Alternatively, it could be worthwhile to remove the purification tag and additional truncations of the *uPHM* sequence, which could potentially increase expression levels.

Improvements to the current *uPHM* construct would involve studies with site-directed mutagenesis of the active site residues and removal of the glycosylation sites to enhance activity. While this version of *uPHM* could be expressed, the sequence could undergo directed evolution using a cloning plasmid in *Escherichia coli*, with subsequent expression and evaluation of the *uPHM* in *Y. lipolytica*. To replicate the activity of PAM, *PAL* could be included after the *uPHM*, preferably with a suitable linker region allowing protease cleavage for enzyme synergy studies. As the uPHM is reliably expressible, appropriate genetic modifications and sequence additions could enable the expression of heterologous PAL after the uPHM to create a fusion protein with complete PAM activity. The presence of the uPHM would help stabilise its expression in *Y. lipolytica* and create a chimaeric PAM. The multiple attempts to express PAM in the present study can attest to the difficulty of recombinantly expressing a heterologous protein not endogenous to a biological kingdom.

While the uPHM truncated construct activity has been confirmed as PHM, the role of the native gene within *Cyphellophora europaea* remains unknown. As the original gene likely was a type I membrane protein with multiple domains, this protein probably had additional functionality. PAM from *Chlamydomonas reinhardtii* is active in ciliogenesis and environmental sensing.

The original *C. europaea* uPHM may be involved in one of these alternate roles in addition to its role as a protease for post-translational modification peptides. There may be an unknown uPHM-only isoform, expressed by *C. europaea*, that fulfils this function.

The problems relating to recombinant expression are highlighted by the failure to express 13 other constructs in the present study. Despite the successful previous *Rattus norvegicus* and *Conus bullatus* PAM heterologous recombinant expression, and the confirmed identity of *Caenorhabditis elegans* PAM, only the uPHM fungal homolog was successfully expressed in the present study. This highlights that the probability of expression of any single construct is low, despite uniform expression methods. Multiple PAM, PHM, and PAL sequences should be identified, modelled, and expressed in various yeast and fungal hosts to mitigate these risks and investigate low sequence identity PAMs in future studies.

I-TASSER *in silico* protein modelling accurately predicted the *uPHM* class, structure and function in the present study, despite the low sequence identity of 16.8%. Any heterologous protein expression attempts in the future would benefit from uniform expression systems and hosts, such as *Aspergillus niger*, *Picha pastoris*, or *Saccharomyces cerevisiae*, which should ideally be performed simultaneously to evaluate protein expression. These should also include an RT-PCR method for transformed colonies to confirm recombinant gene expression and compare transcription rates. The delay in screening from the initial genetic yeast transformation to the detection of protein expression would be reduced by adopting this approach, and transformed colonies with non-transcribing genetic cassettes easily excludable.

An overall improvement to fermentation and purification techniques would be to scale up the volumes for both fermentation and purification. A batch-fed fermentation with uPHM and multiple litre purifications using high capacity equipment should provide > 100 mg of purified uPHM per run. The media for batch fed bio-reactors would require adaptation from the current methods as conditions and yeast metabolism will change with high cell density fermentation but will likely be a variant of the YPD media used in the present study. Continuous monitoring of the oxygen concentrations in the reactor would be beneficial as this may have altered the cell morphology in flask fermentation, thus causing the hyphae that contributed to high final dry cell mass without protein secretion. An oxygen saturated solution would also support higher cell density, with the potential for a 10-fold increase in protein titre, as seen with other multi-litre bioreactors fermentations of *Y. lipolytica*.

These modifications and the increased quantity of purified uPHM would, in turn, enable higher concentrations of uPHM to be used per assay reaction solution beyond the 10 µg used here. Higher total activity per reaction would allow the use of the AGG assay and absorbance readings alone, instead of relying on UPLC-MS data, as originally intended for a medium-throughput assay.

Further increases in activity may be possible at 45°C with uPHM, which is appealing for any industrial enzyme candidate, and would reduce running time and costs. A suitable acidic pH optimum and thermophilic catalase would allow the methods to remain primarily biological in nature, but adding proteins into the reaction solution would reduce total protein solubility levels. Preferably, the majority constituent of the solution would be peptide substrates, not catalase or other reagents, to simplify downstream peptide purification. An electrochemical method, with Cu and Zn as oxidative and reductive elements, would offer a cost-effective and reusable method that would not require the addition of catalase, thereby simplifying downstream purification and providing constant dissolved oxygen levels. However, the effect of the redox reaction on the activity solution and the peptide products would need to be investigated.

The use of larger multiwell plates, such as 384 well plates with a sealing lid, would enable the generation and screening of more data. This change in plate format would use less enzyme and reagents while simultaneously minimising evaporation during prolonged activity studies at 37°C or alkaline hydrolysis at higher temperatures. The reduction in total reaction volume for the activity assay solutions would also reduce the total enzyme quantities per well and reaction. The caveat to this approach for the AGG assay would be that the total absorbance generated would decrease with decreasing volume, likely negatively impacting sensitivity. However, 100 ng of rhPAM was used per 250 µL reaction solution, with between 5 and 10 µg for uPHM, so a reduction in volume would equate to a 60 to 80% reduction in the quantity of enzyme used per reaction. A reduction in volume to below 150 µL would enable the use of PCR strip tubes and thermocyclers, increasing the assays' ease of use and throughput.

A higher throughput system would allow for the screening of glyoxylate produced from PAM activity with α-amidated peptides in solution. A PAM and AGG system would be particularly valuable for identifying the presence and concentration of C-terminal glycine-extended peptides in biological solutions. Such a screening method could aid in the discovery of novel

peptides from tissue and environmental samples and enable the measurement of peptide expression levels in tissues and organs.

6.2 Conclusions

In the present study, α-amidated peptides were successfully produced using a recombinant PHM and standard laboratory techniques, which are likely to be economically feasible to produce at larger scales. This demonstrates that biological peptide synthesis with green chemistry standards can replace existing chemical techniques, effectively eliminating the use of organic solvents and reducing waste.

While multiple attempts were made to replicate the DMPD assay for measuring PAM activity, it was necessary to adapt the AGG assay to measure PAM and PAL activity. An alkaline hydrolysis step was added to enable the expansion of the assay to measure PHM activity. The AGG chromophore is not a direct product of the reaction and can be safely disposed of according to green chemistry goals.

In the present study, a novel fungal PHM was recombinantly expressed in *Y. lipolytica*. The quantities produced were increased by optimising the fermentation at pH 6.0 in citrate in YPD media, with maximum production achieved after 7 days at 28°C and 250 rpm in conical flasks. The uPHM produced had a pH optimum of 4.0 in 100 mM succinate with acYVG.

6.3 Future recommendations

A future study of the intracellular *Y. lipolytica* proteins would be beneficial to examine the total intracellular uPHM concentrations, determine potential degradation by cellular proteosomes, and establish whether these reduce enzyme activity or secreted protein concentrations. Further screening for *Y. lipolytica* transformants with higher protein secretion levels with a larger population of uPHM-transformed colonies is advisable for future research, as potentially higher levels of protein expression and secretion of uPHM might be possible. The issues relating to PAM activity assays and optimal conditions for activity need to be resolved for both the human recombinant PAM and the novel uPHM.

The fermentation process should be scaled up to multiple litres in a bioreactor using optimised medium. The larger fermentation volumes could be implemented in conjunction with a scale-up of purification with a larger TFF cartridge, increased quantities of MagResyn® Ni-NTA magnetic microbeads, or suitable chromatography columns.

It is recommended that the expression of the entire uPAM gene should be studied, and a thorough investigation into the functionality of the domains present in the native gene be attempted. Specifically, its association with cilia, environmental sensing and cellular topology should be studied in more depth. Uncovering the original uPAM biological functions would guide activity optimisation studies and reveal properties that can be exploited for industrial microbiology applications.

Studying the expression and function of novel genes is increasingly becoming the domain of intricate modelling algorithms. Deep learning protein modelling algorithms, such as Google Alphafold, have shown greater than 90% accuracy with structural predictions of novel proteins. The use of these algorithms will increase the confidence of enzyme function and classification prediction in the future.

The conditions for the alkaline hydrolysis of a hydroxylated peptide for all relevant peptide pharmaceuticals should be studied further to enable their subsequent production as biosimilars. Future studies of these peptides should acquire the C-terminal hydroxylated peptides, and alpha-amidated peptides to use as standards and further develop the current UPLC-MS method for detection of their specific peptides of interest. The current UPLC-MS and AGG methods for measuring PAM or PHM activity should form the basis for future research in these studies.

A future study that may be especially relevant for the industrial use of uPHM would be to measure its synergy with varying concentrations of PAL. *Escherichia coli* is capable of producing PAL recombinantly. Potentially limited quantities could be used with uPHM, as its higher turnover rate would enable smaller amounts of the enzyme to be used per reaction. An alternative to an enzyme catalyst for alkaline chemical hydrolysis may be beneficial when producing certain peptide pharmaceuticals. The effect of product inhibition for PHM could be studied with the inclusion of PAL.

To further reduce costs and limit waste generated during biological peptide production, the immobilisation of uPHM and PAM should be studied. As the uPHM was readily immobilised with the 6xHis purification tag on magnetic microbeads, a method developed with MagResyn® beads or a similar product would be advisable.

Appendix I: Record of all DNA and predicted protein sequences for PAM, PHM, PAL

Section A: PAM, PHM and PAL DNA sequences, and predicted peptide amino acid, sequences

The PAM and PHM sequences used from UniProt were the following: P14925 (AMD_RAT); P83388 (AMDL_CAEEL); G8EWC9 (G8EWC9_CONBU); W2RTG2 (W2RTG2_9EURO).

The chosen sequences were modified for expression in *Yarrowia lipolytica* with relevant codon optimisation and removal of variable sequences sections between isoforms in the source organism. Typically the sections encoding signal sequences, and transmembrane domains were removed from the primary sequence synthesised, to minimise total number of base pairs (bp).

Colour code and font style legend:

Table I.1: Topological feature legend for canonical PAM sequences of *Rattus norvegicus* , *Conus bullatus, and Caenorhabditis elegans* PAM.

Colour	Feature
Yellow	Signal peptide sequence
Green	PHM domain
Blue	PAL domain
Red	6 x His tag sequence and stop codons
Pink	Linker region or not classified
Bold	Artificially introduced restriction site
Underline	Sequence selected for primer design

All feature classifications are according to the UniProt review, as of 2014, and personal curated assignments, are based on MSA, for unknown or not described proteins contained below.

***Rattus norvegicus* PAM synthesized DNA sequence codon optimised for expression in *Yarrowia lipolytica*.**

```
      GGATCC
   1  TTCAAGGAAACCACTCGGTCTTTCTCTAACGAATGTCTGGGAACTATCGGACCTGTCACCCCTCTGGACGCCTCG
  76  GACTTTGCCCTCGACATCCGAATGCCCGGCGTGACCCCTAAGGAGTCGGATACTTACTTCTGCATGAGCATGCGA
 151  CTGCCCGTGGACGAGGAAGCCTTCGTCATTGATTTTAAGCCTCGCGCTTCGATGGACACCGTCCACCATATGCTG
 226  CTCTTCGGATGTAACATGCCCTCCTCTACTGGTAGCTACTGGTTTTGCGACGAGGGCACCTGTACTGATAAGGCC
 301  AACATCCTGTACGCTTGGGCTCGTAACGCTCCTCCTACCCGACTCCCTAAGGGAGTTGGTTTCCGCGTGGGCGGA
 376  GAGACTGGCTCGAAGTACTTTGTGCTGCAGGTCCACTACGGAGACATCAGCGCCTTCCGGGACAACCACAAGGAT
 451  TGCTCGGGAGTCAGCGTTCATCTCACCAGAGTGCCTCAGCCGCTGATTGCTGGAATGTACCTCATGATGTCCGTT
 526  GACACTGTGATCCCGCCCGGAGAGAAGGTGGTCAACGCCGATATTTCTTGTCAGTACAAGATGTACCCCATGCAC
 601  GTTTTCGCTTACAGAGTGCATACCCACCATCTGGGCAAGGTTGTGTCCGGATACCGAGTCCGCAACGGACAGTGG
 676  ACTCTGATCGGTCGACAGAACCCTCAGCTCCCTCAGGCTTTCTACCCTGTGGAGCACCCCGTGGACGTCACCTTC
 751  GGAGATATTCTCGCCGCTCGCTGCGTCTTTACTGGCGAGGGACGGACCGAAGCTACTCATATCGGTGGCACCTCG
 826  AGCGACGAGATGTGCAACCTTTACATTATGTACTACATGGAAGCCAAGTACGCTCTCTCGTTCATGACCTGTACT
 901  AAGAACGTGGCCCCGGACATGTTTCGAACCATCCCCGCCGAGGCTAACATCCCCATTCCTGTCAAGCCCGATATG
 976  GTTATGATGCACGGTCACCATAAGGAGGCCGAAAACAAGGAGAAGTCTGCTCTGATGCAGCAGCCTAAGCAGGGC
1051  GAGGAGGAGGTGCTGGAAGACTTCCATGTGGAAGAGGAACTGGATTGGCCTGGAGTCTACCTTCTGCCGGGACAG
1126  GTCTCGGGTGTTGCCCTTGACAGCAAGAACAACCTGGTTATTTTCCACCGTGGAGACCATGTGTGGGATGGTAAC
1201  TCCTTCGATTCTAAGTTTGTTTACCAGCAGCGAGGTCTCGGCCCCATCGAGGAAGACACCATCCTTGTCATTGAT
1276  CCTAACAACGCCGAGATTCTGCAGTCCTCTGGCAAGAACCTGTTTTACCTGCCCCACGGCCTCTCCATCGACACC
1351  GATGGAAACTACTGGGTGACTGACGTCGCTCTCCACCAGGTGTTTAAGCTTGATCCCCATTCTAAGGAGGGCCCT
1426  CTCCTTATTCTGGGACGATCCATGCAGCCCGGTTCTGACCAGAACCACTTTTGCCAGCCTACCGATGTTGCCGTG
1501  GAACCGTCCACTGGTGCTGTCTTTGTTTCCGACGGCTACTGTAACTCTCGCATCGTCCAGTTCTCCCCTTCTGGA
1576  AAGTTTGTTACCCAGTGGGGAGAGGAATCGAGCGGCTCCTCTCCCCGACCTGGACAGTTCTCGGTCCCTCACAGC
1651  CTGGCTCTCGTTCCTCATCTCGACCAGCTTTGCGTGGCTGATCGGGAGAACGGCAGAATCCAGTGTTTCAAGACC
1726  GACACTAAGGAGTTTGTGAGAGAAATTAAGCACGCCTCCTTCGGTCGTAACGTCTTTGCTATCTCTTACATTCCC
1801  GGCTTCCTCTTTGCCGTCAACGGAAAGCCGTACTTCGGTGACCAGGAGCCCGTCCAGGGATTCGTTATGAACTTT
1876  TCGAGCGGTGAAATCATTGATGTGTTCAAGCCTGTCCGGCAGCACTTTGACATGCCGCATGATATCGTGGCCTCT
1951  GAGGACGGCACCGTCTACATTGGAGATGCTCATACCAACACTGTCTGGAAGTTCACTCTGACCGAAAAGATGGAG
2026  CACCGTTCTGTTCACCACCACCACCACCACTAA
      CCTAGG
```

Figure I.1: The synthesized DNA sequence of rPAM, as originated from *Rattus norvegicus* PAM protein sequence.

```
   1  FKETTRSFSNECLGTIGPVTPLDASDFALDIRMPGVTPKESDTYFCMSMRLPVDEEAFVI
  61  DFKPRASMDTVHHMLLFGCNMPSSTGSYWFCDEGTCTDKANILYAWARNAPPTRLPKGVG
 121  FRVGGETGSKYFVLQVHYGDISAFRDNHKDCSGVSVHLTRVPQPLIAGMYLMMSVDTVIP
 181  PGEKVVNADISCQYKMYPMHVFAYRVHTHHLGKVVSGYRVRNGQWTLIGRQNPQLPQAFY
 241  PVEHPVDVTFGDILAARCVFTGEGRTEATHIGGTSSDEMCNLYIMYYMEAKYALSFMTCT
 301  KNVAPDMFRTIPAEANIPIPVKPDMVMMHGHHKEAENKEKSALMQQPKQGEEEVLEDFHV
 361  EEELDWPGVYLLPGQVSGVALDSKNNLVIFHRGDHVWDGNSFDSKFVYQQRGLGPIEEDT
 421  ILVIDPNNAEILQSSGKNLFYLPHGLSIDTDGNYWVTDVALHQVFKLDPHSKEGPLLILG
 481  RSMQPGSDQNHFCQPTDVAVEPSTGAVFVSDGYCNSRIVQFSPSGKFVTQWGEESSGSSP
 541  RPGQFSVPHSLALVPHLDQLCVADRENGRIQCFKTDTKEFVREIKHASFGRNVFAISYIP
 601  GFLFAVNGKPYFGDQEPVQGFVMNFSSGEIIDVFKPVRQHFDMPHDIVASEDGTVYIGDA
 661  HTNTVWKFTLTEKMEHRSVHHHHHH*
```

Figure I.2: The truncated protein sequence for rPAM.

Sequence length: tPHM: 1168 bp; tPAL: 869 bp; 6xHis + stop: 27 bp.

Theoretical pI: 5.71; theoretical molecular weight: 76728.86 Da.

Conus bullatus PAM DNA sequence

Figure I.3: The synthesized DNA sequence cPAM, as originated from *Conus bullatus* PAM protein sequence.

Figure 3 is the ordered protein-encoding DNA sequence for *C. bullatus* (Conbu), from Gene script, expressed as both translated product and encoding DNA. The sequence has been optimized for expression in *Yarrowia lipolytica* and has a total length of 2196 bp with a GC-content of 55.04 %. Sequence length: css 51; cPHM: 855; c-linker: 18; cPAL: 1239; 6xHis + stop: 27 bp.

Conus bullatus PAM protein sequence

```
  1  MRHYTHVAVV LLCGLGLSAS DAVNGSSTDG DNTTLYMDVL MKGAKPSKPD SYLCAGHDVK
 61  EDEAYVVKFE ANASADAVHH ILLYGCGDGP GSPDDVWKCE SICEGSQQIM FAWAKNAPPT
121  KLPEGVGMRI GKQLSIKTVV VQIHYAKSFS ENEEPDKSGI RLHLSKKKPQ YVAGVYIMVA
181  TFDIPPNETK YNVNISCKHD ENFSMFPFAY RTHAHNLGRV ITGYEYNGTY NLIGKGNPMW
241  PQAFYPTNTS VEVKPGDSLV ARCTFNSVGL NHTVSVGSTG DDEMCNFYIM FYTNNTVDVP
301  SGGCSGVDNA SLVDNMPEDS DVPLPPNPLL EAEAQGHHHA MSSSDEDVAV GSQTEVPLSN
361  PLEYQSGWPS EDLHLGQVGG VAADSDGNIY IFHRGDRIWN GSSFDGNNSF QFKDQPVQQE
421  TVIILSPDGH VIRKFGSGRF YMPHGIEVDQ EKNIWVTDVA LHQIFRIAPN GTLPNMELGV
481  RFVSGDDDNH FCKPTDVAVL RSGEFFVSDG YCNSRILKFS KEGELQMKWG QENSGTPPGP
541  SQFFIPHSLA LAEDLHMVCV ADRENGRIQC FDLDGKSPFI LKEPEFGPAV YAVEYSPLEG
601  GLLFAVNGGS SPVQGFTFSL PNKKLLRTWN LPSPQSLSTP HDVAVDTKGH SVYVGELGPS
661  EVRKFVPQVI NETASTIITT FSPPITTHTH THTHRGKHTI ISIIQLFIDV FVIRLFVFRI
721  LHHHHHH*
```

Figure I.4: The truncated protein sequence for cPAMss-.

Theoretical pI: 5.48; theoretical molecular weight: 79966.91 Da.

Utilising this whole DNA sequence for *C. bullatus*, the following truncated constructs were made with PCR and restriction endonulceases alterations: cPAMss-; cPHM+; cPHM-; cPAL.

Conus bullatus truncated protein sequence cPAMss-

```
  1  SASDAVNGSS TDGDNTTLYM DVLMKGAKPS KPDSYLCAGH DVKEDEAYVV KFEANASADA
 61  VHHILLYGCG DGPGSPDDVW KCESICEGSQ QIMFAWAKNA PPTKLPEGVG MRIGKQLSIK
121  TVVVQIHYAK SFSENEEPDK SGIRLHLSKK KPQYVAGVYI MVATFDIPPN ETKYNVNISC
181  KHDENFSMFP FAYRTHAHNL GRVITGYEYN GTYNLIGKGN PMWPQAFYPT NTSVEVKPGD
241  SLVARCTFNS VGLNHTVSVG STGDDEMCNF YIMFYTNNTV DVPSGGCSGV DNASLVDNMP
301  EDSDVPLPPN PLLEAEAQGH HHAMSSSDED VAVGSQTEVP LSNPLEYQSG WPSEDLHLGQ
361  VGGVAADSDG NIYIFHRGDR IWNGSSFDGN NSFQFKDQPV QQETVIILSP DGHVIRKFGS
421  GRFYMPHGIE VDQEKNIWVT DVALHQIFRI APNGTLPNME LGVRFVSGDD DNHFCKPTDV
481  AVLRSGEFFV SDGYCNSRIL KFSKEGELQM KWGQENSGTP PGPSQFFIPH SLALAEDLHM
541  VCVADRENGR IQCFDLDGKS PFILKEPEFG PAVYAVEYSP LEGGLLFAVN GGSSPVQGFT
601  FSLPNKKLLR TWNLPSPQSL STPHDVAVDT KGHSVYVGEL GPSEVRKFVP QVINETASTI
661  ITTFSPPITT HTHTHTHRGK HTIISIIQLF IDVFVIRLFV FRILHHHHHH *
```

Figure I.5: The truncated protein sequence for cPAMss-.

Theoretical pI: 5.48; theoretical molecular weight: 78102.61 Da.

```
  1  SASDAVNGSS TDGDNTTLYM DVLMKGAKPS KPDSYLCAGH DVKEDEAYVV KFEANASADA
 61  VHHILLYGCG DGPGSPDDVW KCESICEGSQ QIMFAWAKNA PPTKLPEGVG MRIGKQLSIK
121  TVVVQIHYAK SFSENEEPDK SGIRLHLSKK KPQYVAGVYI MVATFDIPPN ETKYNVNISC
181  KHDENFSMFP FAYRTHAHNL GRVITGYEYN GTYNLIGKGN PMWPQAFYPT NTSVEVKPGD
241  SLVARCTFNS VGLNHTVSVG STGDDEMCNF YIMFYTNNTV DVPSG*
```

Figure I.6: The truncated protein sequence for cPHM-.

Theoretical pI: 5.56; theoretical molecular weight: 31226.06 Da.

```
  1  SASDAVNGSS TDGDNTTLYM DVLMKGAKPS KPDSYLCAGH DVKEDEAYVV KFEANASADA
 61  VHHILLYGCG DGPGSPDDVW KCESICEGSQ QIMFAWAKNA PPTKLPEGVG MRIGKQLSIK
121  TVVVQIHYAK SFSENEEPDK SGIRLHLSKK KPQYVAGVYI MVATFDIPPN ETKYNVNISC
181  KHDENFSMFP FAYRTHAHNL GRVITGYEYN GTYNLIGKGN PMWPQAFYPT NTSVEVKPGD
241  SLVARCTFNS VGLNHTVSVG STGDDEMCNF YIMFYTNNTV DVPSGGCSGV DHHHHHH*
```

Figure I.7: The truncated protein sequence for cPHM+.

Theoretical pI: 5.89; theoretical molecular weight: 32567.45 Da.

```
  1  NASLVDNMPE DSDVPLPPNP LLEAEAQGHH HAMSSSDEDV AVGSQTEVPL SNPLEYQSGW
 61  PSEDLHLGQV GGVAADSDGN IYIFHRGDRI WNGSSFDGNN SFQFKDQPVQ QETVIILSPD
121  GHVIRKFGSG RFYMPHGIEV DQEKNIWVTD VALHQIFRIA PNGTLPNMEL GVRFVSGDDD
181  NHFCKPTDVA VLRSGEFFVS DGYCNSRILK FSKEGELQMK WGQENSGTPP GPSQFFIPHS
241  LALAEDLHMV CVADRENGRI QCFDLDGKSP FILKEPEFGP AVYAVEYSPL EGGLLFAVNG
301  GSSPVQGFTF SLPNKKLLRT WNLPSPQSLS TPHDVAVDTK GHSVYVGELG PSEVRKFVPQ
361  VINETASTII TTFSPPITTH THTHTHRGKH TIISIIQLFI DVFVIRLFVF RILHHHHHH*
```

Figure I.8: The truncated protein sequence for cPAL.

Theoretical pI: 5.49; theoretical molecular weight: 46376.02 Da.

Caenorhabditis elegans translated protein and gene as ordered

The protein encoding DNA sequence for *C. elegans*, as ordered from Gene script, expressed as both translated product and encoding DNA.

Optimized Sequence (Optimized Sequence Length: 2022 bp, GC%: 55.26)

Caenorhabditis elegans PAM DNA sequence synthesized

```
     GGATCC
  1  ATGAACGACCGAATTTCCATTAACCTGATTTACCTGGTGCTGACTTTTTGCTGCGTGTCTGCCGCCACTGTCCGA
 76  ACCGCCAAGAACGACGATATCCAGAAGTTCACCATCCAGATGATTGGCTACTCCCCCCAGAAGACTGACGATTAC
151  GTTGCCGTGTCGATTGAGGCTACCCCCGGATACGTGGTCGCCTTCGAGCCTATGGCCCACGCTGACCGAGTGCAC
226  CATATGCTGCTCTACGGTTGTACTATGCCCGCTTCGGAGCAGGGCTTTTGGCGAGGAATGGAGACCTGCGGTTGG
301  GGTGGAGGTTCTTACATCCTGTACGCCTGGGCTCGAAACGCTCCCAACCTGGTCCTCCCTAAGGACGTGGCCTTC
376  TCTGTCGGCCACGAGCAGGATGGAATTAAGTACTTTGTGCTCCAGGTCCATTACGCCCAGCCCTTCGCTGGAGAG
451  GTTCACGACTTTTCTGGTGTGACTATGCATATCTCCCAGAAGAAGCCTATGAACCTGGCCGCTGTTATGCTCTTC
526  GTGTCTGGTACCCCTATTCCTCCTCAGCTGCCTGCTTTCCAGAACAACATCACCTGCATGTTTGAGTCTTCCACT
601  CCCATTCACCCTTTCGCCTTTCGAACCCACACTCATGCTATGGGCCGACTCGTGTCCGCCTTCTTTAAGCACGAC
676  GGCCATTGGACCAAGATCGGAAAGCGAAACCCCCAGTGGCCTCAGCTGTTCGAGGGAATCCCCTCCAAGCTCATG
751  ATTGGATCGGGTGACCAGATGTCGGCTTCTTGTCGATTTGACTCGATGGATAAGAACCGAACCGTGAACATGGGT
826  GCCATGGGCGTCGATGAGATGTGCAACTTCTACATGATGTTTCACTACGACGCTAAGCTGGATAACCCCTACCCT
901  CAGGGAGCCATCTGTGCTAAGGACTACCCCTCCAAGATGATTGACTACCCTAAGGATGGTTTCGAGCTGCTCCCC
976  TCTCGACCTGAGCTGGAGCACCATGCTCACCAGTCCAAGGTCCCCTTTGGAATCGTTCAGGAGGCCATTCATGAG
1051 AACCTGGGCGGAGTGAAGCTCGGTCAGGTCGCTGGCCTGGCTTTCAACAACGAGCAGCAGCTGCTCGTTTTTCAG
1126 CGAGCTGGTCGAGTGTGGGACGCTTCCACTTTCGATAACTACAACATCCTGCTCGACAAGAAGCCCATTGCCGAT
1201 CCTGTTATCCTGGTGATTTCCTACTCGGGAAACCAGACCAAGCTGGAGCGAAAGCTCGGTGGCGGACAGTTCTAC
1276 CTCCCCCACGGCATCTACGTCGACAAGGATGGATTTGTTTACACCACTGACGTGGGATCTCACACTGTCGCCAAG
1351 TGGAAGATCGAGGGTAACGAGCTGAAGAACATTTGGACCTCGGGAGAGCTGCTCATGCCCGGCTCTGACCAGCAC
1426 CATTACTGTAAGCCTACCGGCATTACTCGAGTCGAGGACCAGCTGTACGTTACCGATGGATACTGCAACTCCCGA
1501 GTTGTGGTCCTGGACCTCAACGGAAAGCGAATCCGACAGTTCGGACTGCCCGGAGAGGATGCTGGCCAGTTTAAC
1576 CTCCCTCACGACATTGTGTCTGATTCCGCTGGACGACTGCTCGTCACCGACCGAGAGAACGGTCGAGTTCAGCAC
1651 ATGACCACTCAGGGCCATGTCATCGAGGAGTTCAAGTCCACCATGTTTACTAACATTTACTCGGCCGCTTCTCAC
1726 GAGGACTACGTCTTCATGGTTCCCGGTCGACCTATCATGGGCCATGAGACTGAGGGAATTGCTGTCTTCGTTGGC
1801 CGATCGGGCACCGGACTGATCGAGTACGCCTTTGGTCCCACCACTAAGGGCAAGCGAGAGCAGATGGGACCCCAG
1876 TTCGGTCAGCCTCACTGTCTGCGAGTGTGCCCTGACGGTGGCCATATCTTCGTCGGAGATATTGCCGAGGGAAAG
1956 GCTCGACTCTGGCAGTTTAAGATTCGACACGACCAGAACCACCATCACCATCACCATTAA
     CCTAGG
```

Figure I.9: The synthesized DNA sequence EPAM, as originated from *Caenorhabditis elegans* PAM protein sequence.

Sequence length: ess 63; ePHM: 837; e-linker: 21; ePAL: 1068; 6xHis + stop: 27 bp.

Caenorhabditis elegans PAM protein sequence

```
  1  MNDRISINLI YLVLTFCCVS AATVRTAKND DIQKFTIQMI GYSPQKTDDY VAVSIEATPG
 61  YVVAFEPMAH ADRVHHMLLY GCTMPASEQG FWRGMETCGW GGGSYILYAW ARNAPNLVLP
121  KDVAFSVGHE QDGIKYFVLQ VHYAQPFAGE VHDFSGVTMH ISQKKPMNLA AVMLFVSGTP
181  IPPQLPAFQN NITCMFESST PIHPFAFRTH THAMGRLVSA FFKHDGHWTK IGKRNPQWPQ
241  LFEGIPSKLM IGSGDQMSAS CRFDSMDKNR TVNMGAMGVD EMCNFYMMFH YDAKLDNPYP
301  QGAICAKDYP SKMIDYPKDG FELLPSRPEL EHHAHQSKVP FGIVQEAIHE NLGGVKLGQV
361  AGLAFNNEQQ LLVFQRAGRV WDASTFDNYN ILLDKKPIAD PVILVISYSG NQTKLERKLG
421  GGQFYLPHGI YVDKDGFVYT TDVGSHTVAK WKIEGNELKN IWTSGELLMP GSDQHHYCKP
481  TGITRVEDQL YVTDGYCNSR VVVLDLNGKR IRQFGLPGED AGQFNLPHDI VSDSAGRLLV
541  TDRENGRVQH MTTQGHVIEE FKSTMFTNIY SAASHEDYVF MVPGRPIMGH ETEGIAVFVG
601  RSGTGLIEYA FGPTTKGKRE QMGPQFGQPH CLRVCPDGGH IFVGDIAEGK ARLWQFKIRH
661  DQNHHHHHH*
```

Figure I.10: The translated protein sequence of *Caenorhabditis elegans* PAM.

Theoretical pI: 6.51; theoretical molecular weight: 74919.33 Da.

```
  1  ATVRTAKNDD IQKFTIQMIG YSPQKTDDYV AVSIEATPGY VVAFEPMAHA DRVHHMLLYG
 61  CTMPASEQGF WRGMETCGWG GGSYILYAWA RNAPNLVLPK DVAFSVGHEQ DGIKYFVLQV
121  HYAQPFAGEV HDFSGVTMHI SQKKPMNLAA VMLFVSGTPI PPQLPAFQNN ITCMFESSTP
181  IHPFAFRTHT HAMGRLVSAF FKHDGHWTKI GKRNPQWPQL FEGIPSKLMI GSGDQMSASC
241  RFDSMDKNRT VNMGAMGVDE MCNFYMMFHY DAKLDNPYPQ GAICAKDYPS KMIDYPKDGF
301  ELLPSRPELE HHAHQSKVPF GIVQEAIHEN LGGVKLGQVA GLAFNNEQQL LVFQRAGRVW
361  DASTFDNYNI LLDKKPIADP VILVISYSGN QTKLERKLGG GQFYLPHGIY VDKDGFVYTT
421  DVGSHTVAKW KIEGNELKNI WTSGELLMPG SDQHHYCKPT GITRVEDQLY VTDGYCNSRV
481  VVLDLNGKRI RQFGLPGEDA GQFNLPHDIV SDSAGRLLVT DRENGRVQHM TTQGHVIEEF
541  KSTMFTNIYS AASHEDYVFM VPGRPIMGHE TEGIAVFVGR SGTGLIEYAF GPTTKGKREQ
601  MGPQFGQPHC LRVCPDGGHI FVGDIAEGKA RLWQFKIRHD QNHHHHHH*
```

Figure I.11: The translated protein sequence of *Caenorhabditis elegans* PAMss-.

Theoretical pI: 6.53; theoretical molecular weight: 72548.46 Da.

```
  1  ATVRTAKNDD IQKFTIQMIG YSPQKTDDYV AVSIEATPGY VVAFEPMAHA DRVHHMLLYG
 61  CTMPASEQGF WRGMETCGWG GGSYILYAWA RNAPNLVLPK DVAFSVGHEQ DGIKYFVLQV
121  HYAQPFAGEV HDFSGVTMHI SQKKPMNLAA VMLFVSGTPI PPQLPAFQNN ITCMFESSTP
181  IHPFAFRTHT HAMGRLVSAF FKHDGHWTKI GKRNPQWPQL FEGIPSKLMI GSGDQMSASC
241  RFDSMDKNRT VNMGAMGVDE MCNFYMMFHY DAKLDNPYP*
```

Figure I.12: The translated protein sequence of *Caenorhabditis elegans* PHM-.

Theoretical pI: 6.70; theoretical molecular weight: 31305.88 Da.

```
  1  ATVRTAKNDD   IQKFTIQMIG   YSPQKTDDYV   AVSIEATPGY   VVAFEPMAHA   DRVHHMLLYG
 61  CTMPASEQGF   WRGMETCGWG   GGSYILYAWA   RNAPNLVLPK   DVAFSVGHEQ   DGIKYFVLQV
121  HYAQPFAGEV   HDFSGVTMHI   SQKKPMNLAA   VMLFVSGTPI   PPQLPAFQNN   ITCMFESSTP
181  IHPFAFRTHT   HAMGRLVSAF   FKHDGHWTKI   GKRNPQWPQL   FEGIPSKLMI   GSGDQMSASC
241  RFDSMDKNRT VNMGAMGVDE MCNFYMMFHY DAKLDNPYPQ GAICAKHHHH HH*
```

Figure I.13: The translated protein sequence of *Caenorhabditis elegans* PHM+.

Theoretical pI: 7.11; theoretical molecular weight: 32800.54 Da.

```
  1  DYPSKMIDYP KDGFELLPSR PELEHHAHQS KVPFGIVQEA IHENLGGVKL GQVAGLAFNN
 61  EQQLLVFQRA GRVWDASTFD NYNILLDKKP IADPVILVIS YSGNQTKLER KLGGGQFYLP
121  HGIYVDKDGF VYTTDVGSHT VAKWKIEGNE LKNIWTSGEL LMPGSDQHHY CKPTGITRVE
181  DQLYVTDGYC NSRVVVLDLN GKRIRQFGLP GEDAGQFNLP HDIVSDSAGR LLVTDRENGR
241  VQHMTTQGHV IEEFKSTMFT NIYSAASHED YVFMVPGRPI MGHETEGIAV FVGRSGTGLI
301  EYAFGPTTKG KREQMGPQFG QPHCLRVCPD GGHIFVGDIA EGKARLWQFK IRHDQNHHHH
361  HH*
```

Figure I.14: The translated protein sequence of *Caenorhabditis elegans* PAL.

Theoretical pI: 6.34; theoretical molecular weight: 40588.78 Da.

Utilising this whole DNA sequence for *C. elegans*, the following truncated constructs were made with PCR and restriction endonulceases alterations: EPAMss-; EPHM+; EPHM-; EPAL.

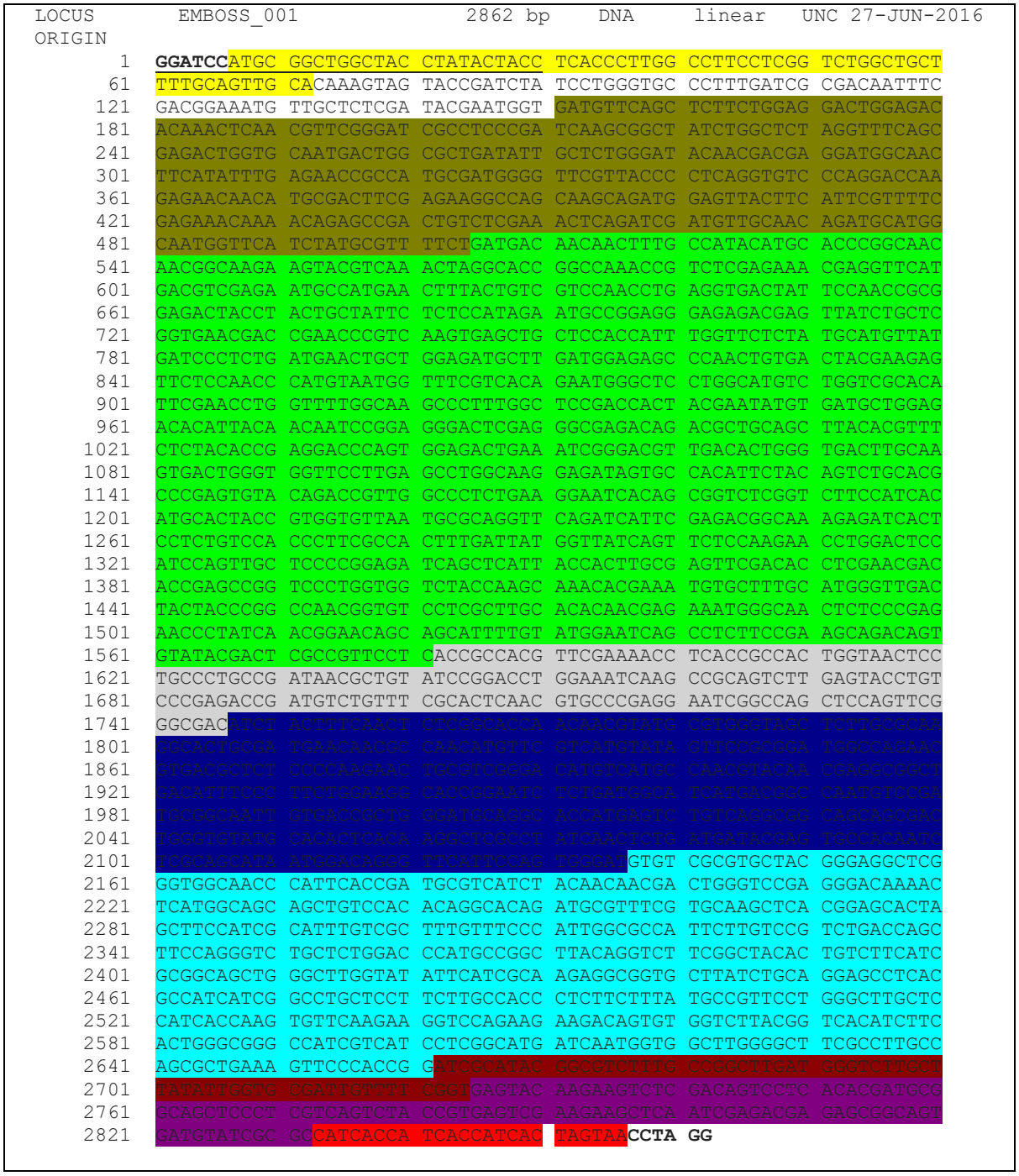
Potential PAM DNA sequence synthesized from *Cyphellophora europaea*

Due to the potential *C. europaea* gene not being described and the presence of non-canonical domains in its sequence, an alternative colour scheme is used, as per Table II.

Table I.2: Topological feature legend for canonical sequence *Cyphellophora europaea.*

Colour	**Sequence feature description [Acronym or abbreviation]**
Yellow	Signal peptide [SP]
Brown	DOMON domain [DOM]
Green	PHM region [PHM]
Grey	Probable linker region [LINK]
Navy blue	Cytochrome b561 (Possible PAL) [CYT]
Blue	Potential PAL region [PAL]
Maroon	Transmembrane domain [TRAN]
Purple	Possible cytoplasmic domain [CD]
Red	6xHis tag [HIS6] and Stop codon [STOP]
Underlined	Sequence for primer design
Bold	Artificial restriction site

Change the above table and colours to the table below.



```
LOCUS       EMBOSS_001              2862 bp    DNA     linear   UNC 27-JUN-2016
ORIGIN
        1  GGATCCATGC GGCTGGCTAC CTATACTACC TCACCCTTGG CCTTCCTCGG TCTGGCTGCT
       61  TTTGCAGTTG CACAAAGTAG TACCGATCTA TCCTGGGTGC CCTTTGATCG CGACAATTTC
      121  GACGGAAATG TTGCTCTCGA TACGAATGGT GATGTTCAGC TCTTCTGGAG GACTGGAGAC
      181  ACAAACTCAA CGTTCGGGAT CGCCTCCCGA TCAAGCGGCT ATCTGGCTCT AGGTTTCAGC
      241  GAGACTGGTG CAATGACTGG CGCTGATATT GCTCTGGGAT ACAACGACGA GGATGGCAAC
      301  TTCATATTTG AGAACCGCCA TGCGATGGGG TTCGTTACCC CTCAGGTGTC CCAGGACCAA
      361  GAGAACAACA TGCGACTTCG AGAAGGCCAG CAAGCAGATG GAGTTACTTC ATTCGTTTTC
      421  GAGAAACAAA ACAGAGCCGA CTGTCTCGAA ACTCAGATCG ATGTTGCAAC AGATGCATGG
      481  CAATGGTTCA TCTATGCGTT TTCTGATGAC AACAACTTTG CCATACATGC ACCCGGCAAC
      541  AACGGCAAGA AGTACGTCAA ACTAGGCACC GGCCAAACCG TCTCGAGAAA CGAGGTTCAT
      601  GACGTCGAGA ATGCCATGAA CTTTACTGTC GTCCAACCTG AGGTGACTAT TCCAACCGCG
      661  GAGACTACCT ACTGCTATTC TCTCCATAGA ATGCCGGAGG GAGAGACGAG TTATCTGCTC
      721  GGTGAACGAC CGAACCCGTC AAGTGAGCTG CTCCACCATT TGGTTCTCTA TGCATGTTAT
      781  GATCCCTCTG ATGAACTGCT GGAGATGCTT GATGGAGAGC CCAACTGTGA CTACGAAGAG
      841  TTCTCCAACC CATGTAATGG TTTCGTCACA GAATGGGCTC CTGGCATGTC TGGTCGCACA
      901  TTCGAACCTG GTTTTGGCAA GCCCTTTGGC TCCGACCACT ACGAATATGT GATGCTGGAG
      961  ACACATTACA ACAATCCGGA GGGACTCGAG GGCGAGACAG ACGCTGCAGC TTACACGTTT
     1021  CTCTACACCG AGGACCCAGT GGAGACTGAA ATCGGGACGT TGACACTGGG TGACTTGCAA
     1081  GTGACTGGGT GGTTCCTTGA GCCTGGCAAG GAGATAGTGC CACATTCTAC AGTCTGCACG
     1141  CCCGAGTGTA CAGACCGTTG GCCCTCTGAA GGAATCACAG CGGTCTCGGT CTTCCATCAC
     1201  ATGCACTACC GTGGTGTTAA TGCGCAGGTT CAGATCATTC GAGACGGCAA AGAGATCACT
     1261  CCTCTGTCCA CCCTTCGCCA CTTTGATTAT GGTTATCAGT TCTCCAAGAA CCTGGACTCC
     1321  ATCCAGTTGC TCCCCGGAGA TCAGCTCATT ACCACTTGCG AGTTCGACAC CTCGAACGAC
     1381  ACCGAGCCGG TCCCTGGTGG TCTACCAAGC AAACACGAAA TGTGCTTTGC ATGGGTTGAC
     1441  TACTACCCGG CCAACGGTGT CCTCGCTTGC ACACAACGAG AAATGGGCAA CTCTCCCGAG
     1501  AACCCTATCA ACGGAACAGC AGCATTTTGT ATGGAATCAG CCTCTTCCGA AGCAGACAGT
     1561  GTATACGACT CGCCGTTCCT CACCGCCACG TTCGAAAACC TCACCGCCAC TGGTAACTCC
     1621  TGCCCTGCCG ATAACGCTGT ATCCGGACCT GGAAATCAAG CCGCAGTCTT GAGTACCTGT
     1681  CCCGAGACCG ATGTCTGTTT CGCACTCAAC GTGCCCGAGG AATCGGCCAG CTCCAGTTCG
     1741  GGCGACATCT ACTTTCAACT CTGGGCACCA ACAACGTATG CGTGGGTACC TCTTGCGCAA
     1801  GGCACTGCGA TGAACAAGGC CAACATGTTC GTCATGTATA GTTCCGCGGA TGGCCAGAAC
     1861  GTGACGCTCT CCCCAAGAAC TGCGTCGGGA CATGTCATGC CAACGTACAA CGAGGGGGCT
     1921  GACATTTCCC TTCTGGAAGG GACGGGAATC TCTGATGGCA TCATGACGGC CAATGTCGCA
     1981  TGCGGCAATT GTGACCGCTG GGATGCAGGC ACCATGAGTC TGTCAGCGCG CAGCAGCGAC
     2041  TGGGTGTATG CACACTGACA AGGCTGGGCT ATCAACTCTG ATGATACGAG TGGCACAATC
     2101  TCGCAGCATA ATGCACAGGG TTCATTCCAG TGGGATGTGT CGCGTGCTAC GGGAGGCTCG
     2161  GGTGGCAACC CATTCACCGA TGCGTCATCT ACAACAACGA CTGGGTCCGA GGGACAAAAC
     2221  TCATGGCAGC AGCTGTCCAC ACAGGCACAG ATGCGTTTCG TGCAAGCTCA CGGAGCACTA
     2281  GCTTCCATCG CATTTGTCGC TTTGTTTCCC ATTGGCGCCA TTCTTGTCCG TCTGACCAGC
     2341  TTCCAGGGTC TGCTCTGGAC CCATGCCGGC TTACAGGTCT TCGGCTACAC TGTCTTCATC
     2401  GCGGCAGCTG GGCTTGGTAT ATTCATCGCA AGAGGCGGTG CTTATCTGCA GGAGCCTCAC
     2461  GCCATCATCG GCCTGCTCCT TCTTGCCACC CTCTTCTTTA TGCCGTTCCT GGGCTTGCTC
     2521  CATCACCAAG TGTTCAAGAA GGTCCAGAAG AAGACAGTGT GGTCTTACGG TCACATCTTC
     2581  ACTGGGCGGG CCATCGTCAT CCTCGGCATG ATCAATGGTG GCTTGGGGCT TCGCCTTGCC
     2641  AGCGCTGAAA GTTCCCACCG GATCGCATAC GGCGTCTTTG CCGGCTTGAT GGGTCTTGCT
     2701  TATATTGGTG CGATTGTTTT CGGTGAGTAC AAGAAGTCTC GACAGTCCTC ACACGATGCG
     2761  GCAGCTCCCT CGTCAGTCTA CCGTGAGTCG AAGAAGCTCA ATCGAGACGA GAGCGGCAGT
     2821  GATGTATCGC GCCATCACCA TCACCATCAC TAGTAACCTA GG
```

Figure I.15: The DNA sequence of the potential PAM from *Cyphellophora europaea*.

uPAM gene as synthesised by Genescript

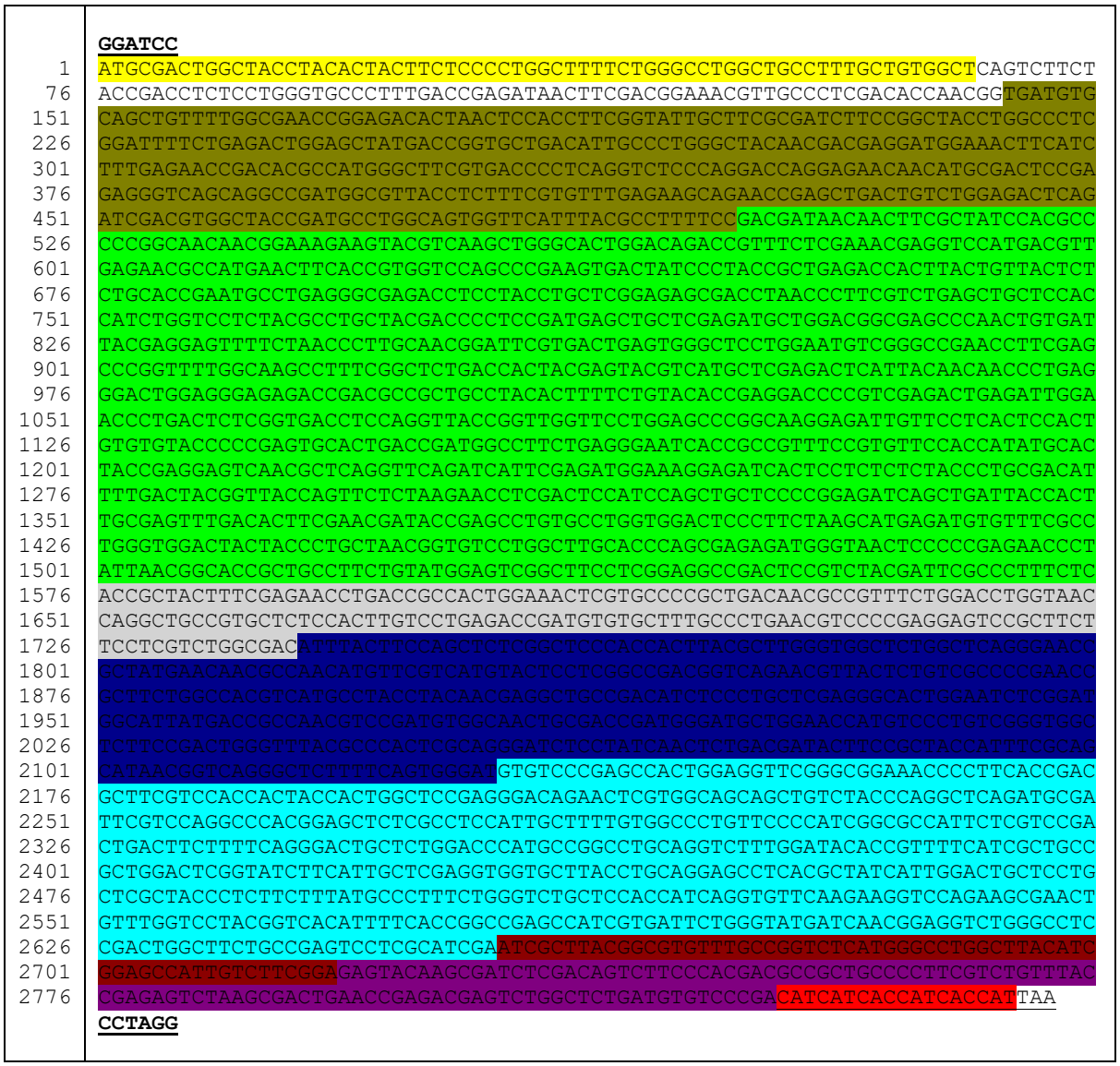


Figure I.16: The synthesized DNA sequence UPAM, as originated from *Cyphellophora europaea* PAM protein sequence.

uss 63;Unknown: 77; DOMON: 355; uPHM: 1077; u-linker: 165; Cytochrome b561: 390; uPAL:525; Transmembrane and cytoplasmic domains: 171; 6xHis + stop: 27

The corresponding translated protein sequence for UPAM:

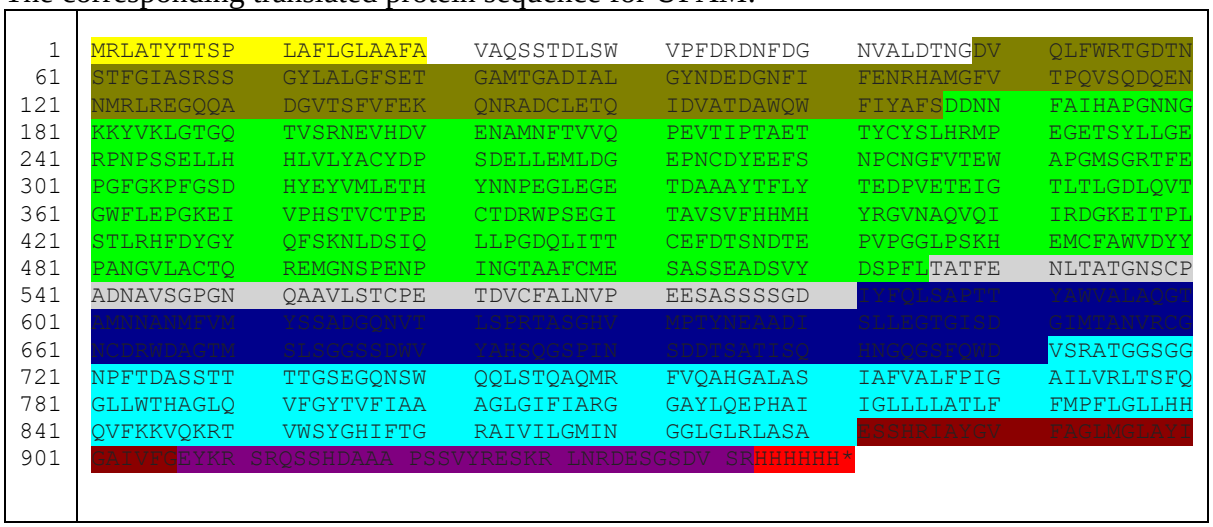


Figure I.17: The translated protein product for *Cyphellophora europaea* UPAM.
Theoretical pI: 4.71; theoretical molecular weight: 102990.14 Da.

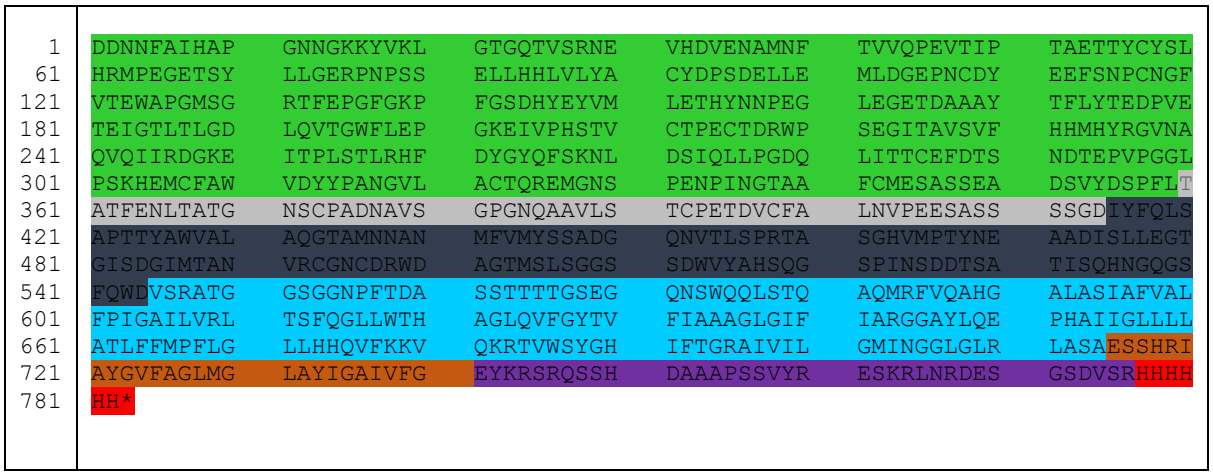


Figure I.18: The translated protein product for *Cyphellophora europaea* UPAM2.
Theoretical pI: 4.89; theoretical molecular weight: 84691.10 Da.

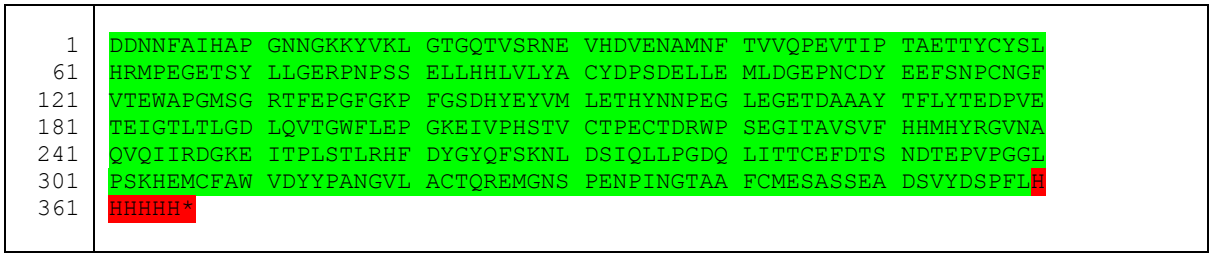


Figure I.19: The translated protein product for *Cyphellophora europaea* uPHM for gene.
Theoretical pI: 4.55; theoretical molecular weight: 40747.90 Da.

Appendix II: Primer sequences, properties, and PCR data.

PCR Information

All DNA melting temperatures (T_M), GC-content, primer-dimer, non-specific annealing, and secondary structures were checked with IDT oligo analyser (https://eu.idtdna.com/calc/analyzer) or NEB Tm Calculator (https://tmcalculator.neb.com). An additional 2 NT were added up stream of the restriction sites to allow for restriction endonuclease binding. The following PCR reagent concentrations, which approximate concentrations for the screening PCR method, were used with the IDT calculator to calculate the correct approximate T_M.

Oligo-nucleotide concentration:	0.25µM
Na+ concentration:	50 mM
Mg++ concentration:	1.5 mM
dNTPs concentration:	0.1 mM

Primer sites from pKOV410 for checking insert region

pKOVf

5`- GCAGTTCTGCAGAAGCGATTC**GGATCC** -3`
27 bp; T_M: 69.6 ºC GC: 55.6 %

pKOVr

5`- GTGATAAATAGCTTAGATACCACAGACAC -3`
29 bp; T_M: 62.8 ºC; GC: 37.9 %

pKOVHF

5`-GCCACAGATTTTCACTCCACACACCACATCACACATACAACCACACACATCCACAATG-3`
58 bp; T_M: 74.6ºC;GC-content: 46.6 %

pKOVHR

5`-GTAGATAGTTGAGGTAGAAGTTGTAAAGAGTGATAAATAGCTTAGATACCACAGACAC CCTAGG-3`
64 bp; T_M: 71.2ºC; GC-content: 39.1 %

Conus bullatus PAM primer sequences

In order to construct truncated version of the *Caenorhabditis elegans* and *Conus bullatus* genes down to their catalytic cores, it was necessary to introduce novel restriction sites as well as stop

codons to the sequences. An extra G/C bases pairs upstream from the restriction sites to allow for potential restriction of the PCR products.

Introduced restriction sites sequences were in bold, stop codons in yellow, and the primer gene binding sites were underlined.

C1

The first *C. bullatus* primer to remove the signal sequence on cPAM.

5`-GG**GGATCC**TCCGCTTCTGATGCTG -3`

8 +16 = 24 bp; T_{M1}: 51.9ºC; GC_1: 56.2%; T_{M2}: 63.4ºC; GC_2: 62.5%

C2

The second primer to isolate the cPHMcc when paired with C1.

5`-GG**CCTAGG**CTAACCAGAAGGGACGTC-3`

16+3+8 = 27 bp; T_{M1}: 50.5ºC; GC_1: 60%; T_{M2}: 64.1ºC; GC_2: 61.5%

C3

The third primer to allow the inclusion of the conserved cysteine within the cPHMcc. Primer is antisense relative to cPHM coding. It served to truncate an expanded cPAM sequence, including a sequence with a conserved cysteine at the end of the cPHMcc and had a stop codon and restriction site included.

5`-GG**CCTAGG**CTAGTGATGATGATGGTGGTGATCGACGCCGGAGCATCC -3`

18+18+3+8 = 47 bp; T_{M1}: 60.5ºC; GC_1: 66.7%; T_{M2}: 71.6ºC; GC_2: 59.6%

C4

The fourth primer for the construction of cPALcc.

5`-GG**GGATCC**AACGCCTCGCTGGTTG -3`

8 +16 = 24 bp; T_{M1}: 56.1ºC; GC_1: 62.5%; T_{M2}: 66.2ºC; GC_2: 66.7%

C5

The fifth primer to be used for pairing, alternatively with either, first or, fourth primers.

5`-GG**CCTAGG**CTAGTGATGGTGATGGTGATG-3`

18+3+8 = 29 bp; T_{M1}: 50.5ºC; GC_1: 50%; T_{M2}: 63.4ºC; GC_2: 55.2%

Table II.1: Truncated sequence design with primer combinations

Sequence Name	Primer Combination	Resultant truncated sequence modification
cPAMss-	C1 + C5	No signal sequence EPAM
cPHM-	C1 + C2	minimal EPHM truncation
cPHM+	C1 + C3	Full length EPHM, with a 6xHis tag.
cPAL	C4 + C5	PAL with his tag

Caenorhabditis elegans PAM primer sequences

E1

The first *C. elegans* primer to remove the signal sequence.
5`-GG**GGATCC**GCCACTGTCCGAACCGCC-3`
(8 + 18 = 26 bp); T_{M1}: 62.1°C; GC_1: 72.2 %; T_{M2}: 70.5°C; GC_2: 73.1 %

E2

The second primer to isolate the PHM. It served to truncate the EPAM sequence to the end of the ePHMcc and had a stop codon and restriction site included. Primer is antisense relative to ePHM coding.
5`-GG**CCTAGG**CTAAGGGTAGGGGTTATCCAG -3`
18+3+8 = 29 bp; T_{M1}: 52.7°C; GC_1: 55.6%; T_{M2}: 64.2°C; GC_2: 58.6 %

E3

The third primer to allow the inclusion of the conserved cysteine within the EPHMcc. Primer is antisense relative to ePHM coding. It served to truncate an expanded EPAM sequence, including a sequence with a conserved cysteine at the end of the ePHMcc and had a stop codon and restriction site included.
5`-GG**CCTAGG**CTAGTGATGATGATGGTGGTGCTTAGCACAGATGGCTCCCTG-3`
21+18+3+8 = 50bp; T_{M1}: 57.6°C; GC_1: 57.1%; T_{M2}: 70.8°C; GC_2: 56%

E4

The fourth primer to isolate the EPALcc.
(5`-CC**GGATCC**GACTACCCCTCCAAGATG -3`)
18 + 8 = 26 bp; T_{M1}: 51.5°C; GC_1: 55.6%; T_{M2}: 63.7°C; GC_2: 61.5%

E5

The fifth primer needed to remove the signal sequence or isolate EPALcc. The primer is antisense relative to ePAL coding, and had a stop codon and restriction site included.
(5`-GG**CCTAGG**CTAATGGTGATGGTGATGGTG-3`)
18+3+8 = 29 bp; T_{M1}: 52.1°C; GC_1: 50%; T_{M2}: 63.8°C; GC_2: 55.2%

Table II.2: Truncated sequence design with primer combinations

Sequence Name	Primer Combination	Resultant truncated sequence modification
EPAMss-	E1 + E5	No signal sequence EPAM
EPHM-	E1 + E2	minimal EPHM truncation
EPHM+	E1 + E3	Full length EPHM, with a 6xHis tag.
EPAL	E4 + E5	PAL with his tag

All the primer sequences and combinations were checked for any homo/hetero- dimers, hairpins, and non-specific binding in their respective isoform PAM gene. None were found with any of the primers here at or near their annealing temperatures.

Appendix III: Record of all plasmid DNA sequences and key features.

Plasmid Sequences

pUC57- Kanamycin Sequence

```
ORIGIN
        1 TCGCGCGTTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGCTCCCG GAGACGGTCA
       61 CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGGCGCG TCAGCGGGTG
      121 TTGGCGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGAGTGC
      181 ACCATATGCG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC
      241 ATTCGCCATT CAGGCTGCGC AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT
      301 TACGCCAGCT GGCGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT
      361 TTTCCCAGTC ACGACGTTGT AAAACGACGG CCAGTGAATT CGAGCTCGGT ACCTCGCGAA
      421 TGCATCTAGA TATCGGATCC CGGGCCCGTC GACTGCAGAG GCCTGCATGC AAGCTTGGCG
      481 TAATCATGGT CATAGCTGTT TCCTGTGTGA AATTGTTATC CGCTCACAAT TCCACACAAC
      541 ATACGAGCCG GAAGCATAAA GTGTAAAGCC TGGGGTGCCT AATGAGTGAG CTAACTCACA
      601 TTAATTGCGT TGCGCTCACT GCCCGCTTTC CAGTCGGGAA ACCTGTCGTG CCAGCTGCAT
      661 TAATGAATCG GCCAACGCGC GGGGAGAGGC GGTTTGCGTA TTGGGCGCTC TTCCGCTTCC
      721 TCGCTCACTG ACTCGCTGCG CTCGGTCGTT CGGCTGCGGC GAGCGGTATC AGCTCACTCA
      781 AAGGCGGTAA TACGGTTATC CACAGAATCA GGGGATAACG CAGGAAAGAA CATGTGAGCA
      841 AAAGGCCAGC AAAAGGCCAG GAACCGTAAA AAGGCCGCGT TGCTGGCGTT TTTCCATAGG
      901 CTCCGCCCCC CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCAGAGGTG GCGAAACCCG
      961 ACAGGACTAT AAAGATACCA GGCGTTTCCC CCTGGAAGCT CCCTCGTGCG CTCTCCTGTT
     1021 CCGACCCTGC CGCTTACCGG ATACCTGTCC GCCTTTCTCC CTTCGGGAAG CGTGGCGCTT
     1081 TCTCATAGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG TCGTTCGCTC CAAGCTGGGC
     1141 TGTGTGCACG AACCCCCCGT TCAGCCCGAC CGCTGCGCCT TATCCGGTAA CTATCGTCTT
     1201 GAGTCCAACC CGGTAAGACA CGACTTATCG CCACTGGCAG CAGCCACTGG TAACAGGATT
     1261 AGCAGAGCGA GGTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC TAACTACGGC
     1321 TACACTAGAA GAACAGTATT TGGTATCTGC GCTCTGCTGA AGCCAGTTAC CTTCGGAAAA
     1381 AGAGTTGGTA GCTCTTGATC CGGCAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTTGTT
     1441 TGCAAGCAGC AGATTACGCG CAGAAAAAAA GGATCTCAAG AAGATCCTTT GATCTTTTCT
     1501 ACGGGGTCTG ACGCTCAGTG GAACGAAAAC TCACGTTAAG GGATTTTGGT CATGAGATTA
     1561 TCAAAAAGGA TCTTCACCTA GATCCTTTTA AATTAAAAAT GAAGTTTTAA ATCAATCTAA
     1621 AGTATATATG AGTAAACTTG GTCTGACAGT TACCAATGCT TAATCAGTGA GGCACCTATC
     1681 TCAGCGATCT GTCTATTTCG TTCATCCATA GTTGCCTGAC TCCCCGTCGT GTAGATAACT
     1741 ACGATACGGG AGGGCTTACC ATCTGGCCCC AGTGCTGCAA TGATACCGCG AGACCCACGC
     1801 TCACCGGCTC CAGATTTATC AGCAATAAAC CAGCCAGCCG GAAGGGCCGA GCGCAGAAGT
     1861 GGTCCTGCAA CTTTATCCGC CTCCATCCAG TCTATTAATT GTTGCCGGGA AGCTAGAGTA
     1921 AGTAGTTCGC CAGTTAATAG TTTGCGCAAC GTTGTTGCCA TTGCTACAGG CATCGTGGTG
     1981 TCACGCTCGT CGTTTGGTAT GGCTTCATTC AGCTCCGGTT CCCAACGATC AAGGCGAGTT
     2041 ACATGATCCC CCATGTTGTG CAAAAAAGCG GTTAGCTCCT TCGGTCCTCC GATCGTTGTC
     2101 AGAAGTAAGT TGGCCGCAGT GTTATCACTC ATGGTTATGG CAGCACTGCA TAATTCTCTT
     2161 ACTGTCATGC CATCCGTAAG ATGCTTTTCT GTGACTGGTG AGTACTCAAC CAAGTCATTC
     2221 TGAGAATAGT GTATGCGGCG ACCGAGTTGC TCTTGCCCGG CGTCAATACG GGATAATACC
     2281 GCGCCACATA GCAGAACTTT AAAAGTGCTC ATCATTGGAA AACGTTCTTC GGGGCGAAAA
     2341 CTCTCAAGGA TCTTACCGCT GTTGAGATCC AGTTCGATGT AACCCACTCG TGCACCCAAC
     2401 TGATCTTCAG CATCTTTTAC TTTCACCAGC GTTTCTGGGT GAGCAAAAAC AGGAAGGCAA
     2461 AATGCCGCAA AAAAGGGAAT AAGGGCGACA CGGAAATGTT GAATACTCAT ACTCTTCCTT
     2521 TTTCAATATT ATTGAAGCAT TTATCAGGGT TATTGTCTCA TGAGCGGATA CATATTTGAA
     2581 TGTATTTAGA AAAATAAACA AATAGGGGTT CCGCGCACAT TTCCCCGAAA AGTGCCACCT
     2641 GACGTCTAAG AAACCATTAT TATCATGACA TTAACCTATA AAAATAGGCG TATCACGAGG
     2701 CCCTTTCGTC
//
```

Figure III.1: The DNA sequence of pUC-57-Kanamycin plasmid.

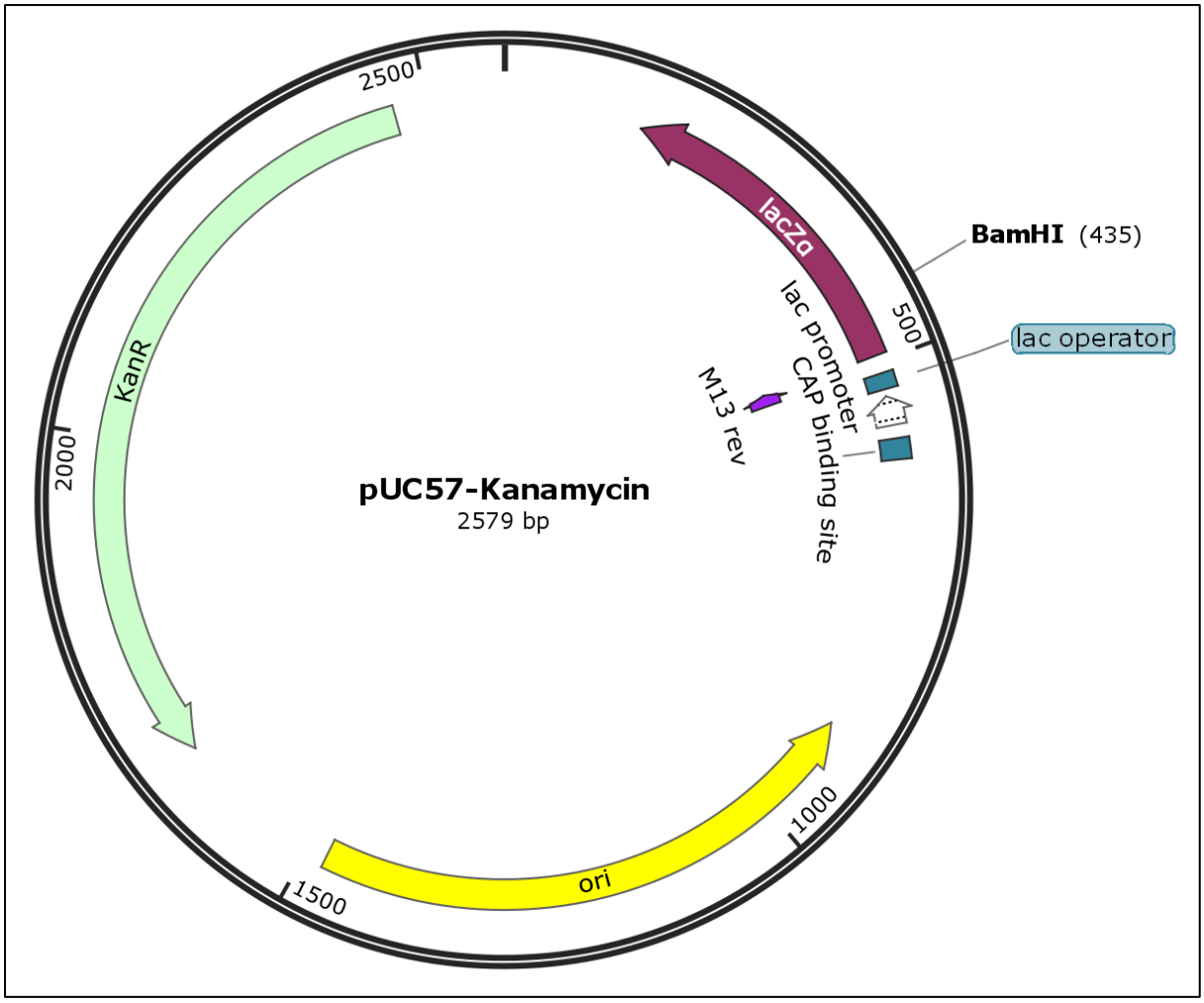


Figure III.2: Plasmid map of pUC57-Kanamycin blank with no gene inserted, annotated with primary plasmid features.

pKOV410 plasmid DNA sequence

Origin

GCTAGCTTATCGATACGCGTGCATGCTGAGGTGTCTCACAAGTGCCGTGCAGTCCCGCCCCCACTTGCTTCTCTTTGTGTGTAGTGTACGT
ACATTATCGAGACCGTTGTTCCCGCCCACCTCGATCCGGCATGCTGAGGTGTCTCACAAGTGCCGTGCAGTCCCGCCCCCACTTGCTTCTC
TTTGTGTGTAGTGTACGTACATTATCGAGACCGTTGTTCCCGCCCACCTCGATCCGGCATGCTGAGGTGTCTCACAAGTGCCGTGCAGTCC
CGCCCCCACTTGCTTCTCTTTGTGTGTAGTGTACGTACATTATCGAGACCGTTGTTCCCGCCCACCTCGATCCGGCATGCTGAGGTGTCTC
ACAAGTGCCGTGCAGTCCCGCCCCCACTTGCTTCTCTTTGTGTGTAGTGTACGTACATTATCGAGACCGTTGTTCCCGCCCACCTCGATCC
GGCATGCACTGATCACGGGCAAAAGTGCGTATATATACAAGAGCGTTTGCCAGCCACAGATTTTCACTCCACACCACATCACACATACA
ACCACACACATCCACAATGAAGCTTTCCACCATCCTTTTCACAGCCTGCGCTACCCTGGCTGCCGCCCTCCCTTCCCCCATCACTCCTTCT
GAGGCCGCAGTTCTGCAGAAGCGATTCGGATCCGGTACCTAGGGTGTCTGTGGTATCTAAGCTATTTATCACTCTTTACAACTTCTACCTC
AACTATCTACTTTAATAAATGAATATCGTTTATTCTCTATGATTACTGTATATGCGTTCCTCTAAGACAAATCGAATTCTTCTTTCCCCGC
TGATTCTGCCAAGCCCGTTCCCTTGGCTGTGGTTTCGCTAGATAGTAGATAGGGACAGTGGGAATCTCGTTAATCCATTCATGCGCGTCAC
TAATTAGATGACGAGGCATTTGGCTACGTTAAGAGAGTCATAGTTACTCCCGCCGTTTACCCGCGCTTGGTTCAATTTCTTCACTTTGACA
TTCAGAGCACTGGGCAGAAATCACTTTGCGTCAACATTGTTTTACAACCATCGCAAAGCTATGTTTTAATTAGACAGTCAGATTCCCCTTG
TCCGTACCAGTTCTGGGGCGGTTGATGATGTGGACGGAGCGACGCAGTGTCTGCCGCGGGTCCGGCTGCCAGTTGCCCAGCCGCCAGTCCC
TTGGGGTCTCCCACGCTTCCATCCACCAACCCCCAGAGCCAATCCTTATCCCGAAGTTACGGATCTATTTTGCCGACTTCCCTTATCTACA
TTATTCTATCAACTAGAGGCTGTTCACCTTGGAGACCTGCTGCGGTTATCAGTACGACTTGGCGTGAAAACTATTCCTTCCTGCGGGTTTT
CAAGGGGCGTCACAAGTGCATCGGAGGGGACAAAGCTGCCCCCCTCTTCCGGAACGATACCCTATCTCTAGACAAACTAATTCCAGGGTGT
TGACTCCGTTAAAAAGAAAAGATAACTCCTTCCAAGACTTGTGCCACCGTCCCCACATTCAGTTACGTTGCCGTGGAGAATCCACATCCAA
GTGCCGGAATTTCAACCGGCTTCCCTTTTGGAGAAGGCCCTTATCGGGCCGTACGCGACAGTTATGCCACCCCTTAGGATCGACTTACCCG
TGTCCAACTGCTGTTCACACGGAACCTTTCCCCACTTCAGTCTTCAAAGTTCTCATTTGAATATTTGCTACTACCACCAAGATCTGCACTA
GGGGCCGTTCGACGCAGCCTCACGGCCAACGCTTCTCAACAACCCCCACGCCTGCCTACTCGAGAGGAAACCTCCTCTGCGGCCGCCCGGG
GTGGGCGAAGAACTCCAGCATGAGATCCCCGCGCTGGAGGATCATCCAGCCGGCGTCCCGGAAAACGATTCCGAAGCCCAACCTTTCATAG
AAGGCGGCGGTGGAATCGAAATCTCGTGATGGCAGGTTGGGCGTCGCTTGGTCGGTCATTTCGAACCCCAGAGTCCCGCTCAGAAGAACTC
GTCAAGAAGGCGATAGAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGTCAGCCCATTCGCCGCCAAGC
TCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGCCACACCCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGC
CATTTTCCACCATGATATTCGGCAAGCAGGCATCGCCATGGGTCACGACGAGATCCTCGCCGTCGGGCATGCGCGCCTTGAGCCTGGCGAA
CAGTTCGGCTGGCGCGAGCCCCTGATGCTCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGAGTACGTGCTCGCTCGATG
CGATGTTTCGCTTGGTGGTCGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCCGCCGCATTGCATCAGCCATGATGGATACTTTCTCGG
CAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGCACTTCGCCCAATAGCAGCCAGTCCCTTCCCGCTTCAGTGACAACGTCGAGCAC
AGCTGCGCAAGGAACGCCCGTCGTGGCCAGCCACGATAGCCGCGCTGCCTCGTCCTGCAGTTCATTCAGGGCACCGGACAGGTCGGTCTTG
ACAAAAAGAACCGGGCGCCCCTGCGCTGACAGCCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTTGTGCCCAGTCATAGCCGAATA
GCCTCTCCACCCAAGCGGCCGGAGAACCTGCGTGCAATCCATCTTGTTCAATCATGCGAAACGATCCTCATCCTGTCTCTTGATCAGATCT
TGATCCCCTGCGCCATCAGATCCTTGGCGGCAAGAAAGCCATCCAGTTTACTTTGCAGGGCTTCCCAACCTTACCAGAGGGCGCCCCAGCT
GGCAATTCCGGTTCGCTTGCTGTCCATAAAACCGCCCAGTCTAGCTATCGCCATGTAAGCCCACTGCAAGCTACCTGCTTTCTCTTTGCGC
TTGCGTTTTCCCTTGTCCAGATAGCCCAGTAGCTGACATTCATCCGGGGTCAGCACCGTTTCTGCGGACTGGCTTTCTACGTGTTCCGCTT
CCTTTAGCAGCCCTTGCGCCCTGAGTGCTTGCGGCAGCGTGAAGCTAGCTTATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAAT
ACCGCATCAGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGC
GGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGAC
TATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCT
CCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCAC
GAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAG
CAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG
GACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGT
AGCGGCGGTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTTACTGAACGGTGAT
CCCCACCGGAATTGCGGCCGCCGGTAGAGTATAGGTAACACGGTTAAGCGCCATCCATTTTCAGGGCTAGTTCATTCGGCCGGTGAGTTGT
GACACACTCCTTAGCGGATTCCGACTTCCATGGCCACCGTCCGGCTGTCTAGATAAACTAACACCTTTTGTGGTGTCTGATGAGCGTGTAT
TCCGGCACCTTAACTCCACGTTCGGTTCATCCCGCATCGCCAGTTCTGCTTACCAAAAATGGCCCACTACAAGCTCTTCATTCGTAGCCAC

GGTTCGATTAAGTAACCAAGGCTTCTTACATATTTAAAGTTTGAGAATAAGCCGAGGGCATTTCGCCCCCAATACTTCTAATCATTCGCTT
TACCTCATAAAACTGATATGAGCTTCTGCTATCCTGAGGGAAACTTCGGCAGGAACCAGCTACTAGATGGTTCGATTAGTCTTTCGCCCCT
ATACCCAGATCCTACGATCGATTTGCACGTCAGAACCGCTACGAGCCTCCATCAGAGTTTCCTCTGACTTCACCCTATCCAGGCATAGTTC
ACCATCTTTCGGGTCCCAATGGATGTGTTTACTCAAAACTGTCATAATAATCAAGTTCGGTCGATTCTGAACCAATCCATTCACTTTCATT
GCGCGCTGGGGTTTGCCACCCTAACACTCACATATCCATTAGACTCCTTGGTCCGTGTTTCAAGACGGGTGAAATGGGTGGATTATGTCGT
CGGTGGCAGTGTGGAGGGTTAGGGGAGAACGCCCGAAAGGCGCTCCCATTTGTAACCCTCGTCTCGCTATCGATGACTCGGCGTCGGCAGT
ACACCGCCCACGAGGGGCGGCTGAAACCTCGGCCACTCTCCACTCATTTCCTTCCCTATCAACAATTTCACATACTATTTCACTCTCTTTT
CAAAGTTCTTTTCACCTTTCCTTCACAGTACTTGTTCGCTATCGGTCTCTCACCAGTATTTAGCTTTAGATGGAGTTTACCACCCACTTTG
AGCTGCATTCCCAAACAACTGTCGACAAAGGCCTGTTTCTCGGTGTACAGAGCTTGGTCCTCCTTGAAGTTGCGACACATGTCTTGATAGT
ATCTTGGCTTCTCTCTCTTGAGCTTTTCCATAACAAGTTCTTCTGCCTCCAGGAAGTCCATGGGTGGTTTGATCATGGTTTTGGTGTAGTG
GTAGTGCAGTGGTGGTATTGTGACTGGGGATGTAGTTGAGAATAAGTCATACACAAGTCAGCTTTCTTCGAGCCTCATATAAGTATAAGTA
GTTCAACGTATTAGCACTGTACCCAGCATCTCCGTATCGAGAAACACAACAACATGCCCCATTGGACAGACCATGCGGATACACAGGTTGT
GCAGTACCATACATACTCGATCAGACAGGTCGTCTGACCATCATACAAGCTGAACAGCGCTCCATACTTGCACGCTCTCTATATACACAGT
TAAATTACATATCCATAGTCTAACCTCTAACAGTTAATCTTCTGGTAAGCCTCCCAGCCAGCCTTCTGGTATCGCTTGGCCTCCTCAATAG
GATCTCGGTTCTGGCCGTACAGACCTCGGCCGACAATTATGATATCCGTTCCGGTAGACATGACATCCTCAACAGTTCGGTACTGCTGTCC
GAGAGCGTCTCCCTTGTCGTCAAGACCCACCCCGGGGGTCAGAATAAGCCAGTCCTCAGAGTCGCCCTTAGGTCGGTTCTGGGCAATGAAG
CCAACCACAAACTCGGGGTCGGATCGGGCAAGCTCAATGGTCTGCTTGGAGTACTCGCCAGTGGCCAGAGAGCCCTTGCAAGACAGCTCGG
CCAGCATGAGCAGACCTCTGGCCAGCTTCTCGTTGGGAGAGGGGACTAGGAACTCCTTGTACTGGGAGTTCTCGTAGTCAGAGACGTCCTC
CTTCTTCTGTTCAGAGACAGTTTCCTCGGCACCAGCTCGCAGGCCAGCAATGATTCCGGTTCCGGGTACACCGTGGGCGTTGGTGATATCG
GACCACTCGGCGATTCGGTGACACCGGTACTGGTGCTTGACAGTGTTGCCAATATCTGCGAACTTTCTGTCCTCGAACAGGAAGAAACCGT
GCTTAAGAGCAAGTTCCTTGAGGGGGAGCACAGTGCCGGCGTAGGTGAAGTCGTCAATGATGTCGATATGGGTCTTGATCATGCACACATA
AGGTCCGACCTTATCGGCAAGCTCAATGAGCTCCTTGGTGGTGGTAACATCCAGAGAAGCACACAGGTTGGTTTTCTTGGCTGCCACGAGC
TTGAGCACTCGAGCGGCAAAGGCGGACTTGTGGACGTTAGCTCGAGCTTCGTAGGAGGGCATTTTGGTATCAA

Figure III.3: The DNA sequence of pKOV410 plasmid.

Figure Legend:

ATG - Start codon for LipSS	**AAGCTT** – HindIII restriction site
CTGCAG – PstI restriction site	**GAAGCGATTC** – XmnI restriction site
GGATCC - BamHI restriction site	**CCTAGG** - AvrII (aka XmnJI) restriction site

Appendix IV: Record of all reagents and consumables

List of reagents and consmables

- NaCl (Merck; SAAR5822320EM)
- Agar Bacteriological (Biolab; HG000BX1)
- Phosphoric acid (Sigma-Aldrich; 345245)
- Absolute ethanol (Sigma-Aldrich; 32221-2.5L)
- Peptone Powder (Merck; SAAR4943300DN)
- YNB without amino acids (BD; 291940)
- Glass beads, acid washed (Sigma; G8772-100G)
- O'Generuler 1 kb (Thermo Scientific™; #SM1163)
- Pronasafe dye (Laboratorios CONDA; CK130)
- Chloroform (Merck; SAAR1595040LC)
- *Bam*HI restriction endonucleases (Thermo Scientific™; ER0051)
- *Avr*II restriction endonucleases (Thermo Scientific™; ER1561)
- *NotI* restriction endonucleases, fast digest (Thermo Scientific™; FD0593)
- *NotI* restriction endonucleases, (Thermo Scientific™; ER0591)
- Pyridine (Merck; 270970-2L)
- PageRuler Plus prestained protein ladder (ThermoFisher Scientific; 26619)
- Benzenesulphonyl chloride (Merck; 108138-5G)
- Organic solvent resistant 96-well plates (Star labs, PlateOne®; S1837-9600)
- Hippuric acid (Sigma-Aldrich; 112003-100G)
- 2-Aminobenzaldehyde (2AB) (Merck; A9628-1G),
- Glycine (Merck; G8898-1KG)
- Glyoxylate (Merck; G4502-1G)
- Formic acid (Merck, 533002)
- Sulphuric acid (Sigma-Aldrich, 258105-2.5L)
- NaOH (Sigma-Aldrich; S5881-1KG)
- Acetonitrile (Merck, 100029)
- Zyppy Plasmid Miniprep kit, (Zymo Research; D4036)
- Isolate II Plasmid mini kit (Bioline; BIO-52055)
- GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific; K0502).
- Electroporation cuvette, 0.2 mm gap (Sigma Aldrich; Z706086)
- CapeBio (CapeBio; Express DNA Ligation Kit)
- T4 DNA Ligase (Thermo Scientific™; EL0014).
- Kapa 2G (Kapabiosystems; KK5009)
- 2x Laemmli Sample Buffer (Bio-Rad; 161-0737)
- Ampliqon red (Ampliqon; A190303)

- Coomassie Brilliant Blue G 250 (Merck; 1154440025)
- Ammonium acetate (Merck; SAAR1122220EM)
- Trizma base (Tris) (Sigma-Aldrich; T1503-1KG),
- Acetic acid (Sigma-Aldrich; A6283-2.5L)
- EDTA (Sigma-Aldrich; EDS-1KG)
- NaCl (Merck; SAAR5822320EM)
- Li-acetate (Sigma-Aldrich; 517992-100G)
- Yeast extract (Biolab, HG000BX6)
- Peptone (Merck; SAAR4943300DN)
- D-glucose (Merck; SAAR2676020EM)
- Deep well 10 mL 24-well clear polypropylene plates (Corning, Axygen®; P-DW-10ML-24-C)
- Oxygen-permeable membrane to facilitate oxygen exchange (Corning, Axygen®; BF-400-S)
- Glycerol (Minema; G6020)
- Eppendorf reaction tubes, Brand® (1.5 mL) (Sigma-Aldrich; BR780540-1000EA)
- Bovine serum albumin (BSA) fraction V (Sigma-Aldrich; 10735086001)
- Bradford's reagent (Sigma-Aldrich; B6916-500ML)
- Tryptone (Sigma-Aldrich; T9410-1KG)
- KH_2PO_4 (Sigma-Aldrich; P9791-500G)
- $MgSO_4 \cdot 7\ H_2O$ (Sigma-Aldrich; 230391-1KG)
- $CuSO_4$ (Sigma-Aldrich; C1297-100G)
- Thiamine (Sigma-Aldrich; T4625-100G)
- Citrate (Sigma-Aldrich; C8532-5KG)
- Sugar beet molasses (Grafschafter®; Goldsaft 450 g)
- Ammonium chloride NH_4Cl (Sigma-Aldrich; A9434-500G)
- Ammonium hydroxide NH_4OH (Sigma-Aldrich; 05002-2.5L)
- Corning® 0.45 µm filter syringe (Sigma-Aldrich; CLS431220-50EA)
- 3-(N-Morpholino) propanesulphonic acid (MOPS) (Sigma-Aldrich; M1254-250G)
- Tangential flow filtration (TFF) system (Pall corporation, OA010C12)
- Nitrilotriacetic acid (NTA) residue-coated magnetic microspheres (MagReSyn®; MR-NTA005)
- Imidazole (Sigma-Aldrich; I239-100G)
- 15 mL Vivaspin® 6 centrifugal concentrators (Sartorius; VS0601).
- CuCl (Sigma-Aldrich; 224332-25G)
- Tris HCl (pH 6.8) (Sigma Aldrich; 10812846001)
- Sodium dodecyl sulfate (SDS) (Sigma-Aldrich; L3771-1KG)
- Bromophenol blue (Sigma-Aldrich; B0126-25G)
- β-mercaptoethanol (Sigma Aldrich; M6250-100ML)
- Degassed 30% acrylamide/bis (Bio-Rad; 1610158)
- Ammonium persulphate (APS) (Sigma-Aldrich; A3678-25G)

- N,N,N',N'-tetramethylethylenediamine (TEMED) (Sigma-Aldrich; T9281-25ML)
- Isopropanol (Propan-2-ol) (Sigma-Aldrich; 278475-2L)
- $ZnCl_2$ (Sigma-Aldrich; 208086-100G)
- Ascorbic acid (Sigma-Aldrich; A92902-100G)
- Catalase from bovine liver (Sigma-Aldrich; C9322-5G)
- H_2O_2 (30%) (Sigma; H1009-500ML)
- Bovine liver catalase (Sigma-Aldrich; C 1345; EC 1.11.1.6)
- Recombinant *Homo sapiens* PAM (rhPAM) (RnDsystems 4837-AM-010; Lot number: PMG1020051; 10 µg)
- Parafilm® (Star lab group; I3080-5075)
- Analytical grade (≥ 99%) DMPD.2HCl (Sigma-Aldrich; 07750-25G)
- Spectrophotometric grade (≥ 97%) DMPD (Sigma-Aldrich; 07750-25G)
- Peptide-N-Glycosidase F (PNGase) from *Elizabethkingia miricola*. (Sigma-Aldrich; P7367)

Appendix V: Protein concentration standard curve

BSA Standard curve

The procedure for the Bradfords protein assay was performed as per BioRad Quick Start™ Bradford Protein assay manual. A standard curve for protein concentrations ranging from 0.05 to 1.2 mg/mL of protein was established with dilutions of bovine serum albumin (BSA). Absorbance readings were measured at 15 after the addition of dye solution (BioRad, cat no.: 500-0205) to ensure a consistent measurement.

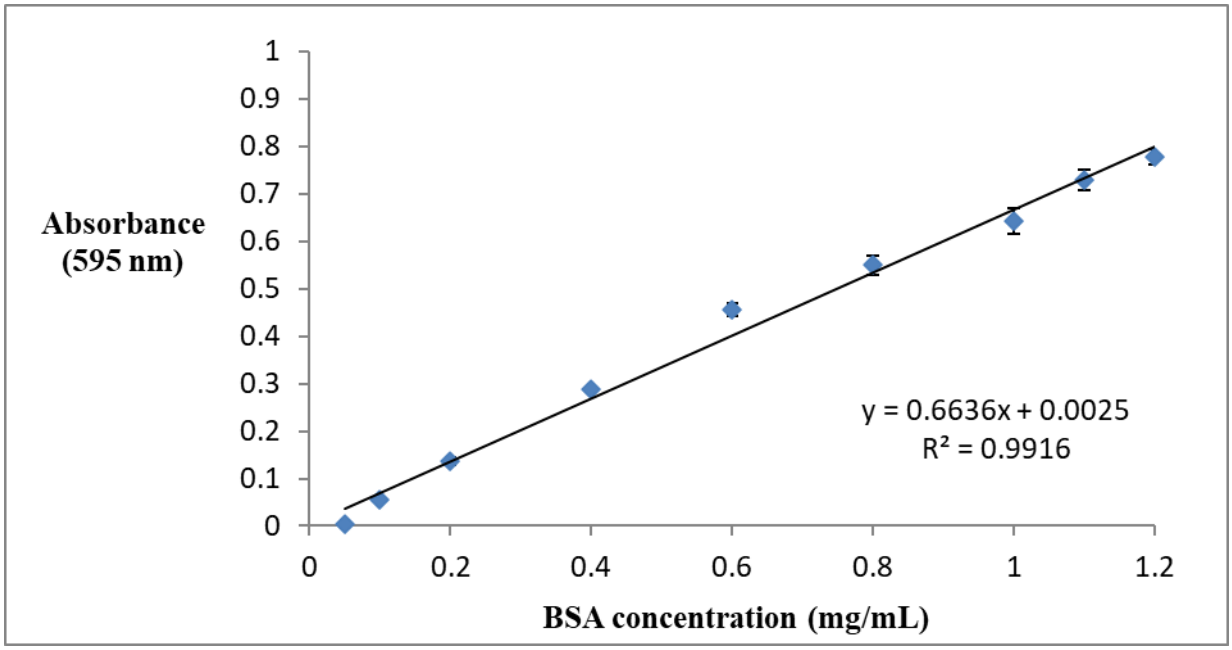


Figure V.1: A standard curve of BSA concentrations between 0.05 to 1.2 mg/mL measured with Bradford's protein concentration assay at 595 nm.