

https://dspace.org/

# Cloning and expression of an industrial enzyme in Pichia pastoris

**Browne, Lee Anne**

**2017**

http://hdl.handle.net/10539/23523

Browne, Lee Anne (2017) Cloning and expression of an industrial enzyme in Pichia pastoris, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/23523>

Downloaded from a DSpace repository. https://dspace.org/




# Cloning and Expression of an Industrial Enzyme in *Pichia pastoris*

**Lee Anne Browne**

A Dissertation submitted to the Faculty of Science, University of the Witwatersrand, in fulfilment of the requirements for the degree of Master of Science

Johannesburg, 2017

# Declaration

I, Lee Anne Browne (1271323), am a student registered for the degree of Masters in the academic year 2017.

I hereby declare the following:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above degree is my own unaided work except where explicitly indicated otherwise and acknowledged.
- I have not submitted this work before for any other degree or examination at this or any other University.
- The information used in the Thesis/Dissertation/Research Report has not been obtained by me while employed by, or working under the aegis of, any person or organisation other than the University.
- I have followed the required conventions in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.

Signed this 5th day of June 2017

## Abstract

*Pichia pastoris* is an established platform for the production of industrial enzymes. This non-fermentative methylotrophic yeast has many attractive features for the production of heterologous protein both in the laboratory and in industry. The PichiaPink™ multi-copy secreted expression system was selected for the heterologous production of the fluorinase from *Streptomyces cattleya.* Fluorinase enzymes are useful for the production of fluorinated compounds which are applied in the agrochemical and pharmaceutical industries. The gene was cloned into the pPinkα-HC vector and used to transform four host srains by electroporation. Protein production was induced with 0.5% methanol and expression and activity was analysed by SDS-PAGE and a HPLC activity assay. Construction of the pPinkα-HC-*fLA* expression plasmid and transformation of the host strains proved succesful. The PichiaPink™ integrants showed genetic instability as the expression cassette showed signs of gene excision, thereby reducing the gene copy number. The wild-type *strain1* efficiently secreted the foreign protein into the culture media, but the α-MF secretion signal was not processed correctly and secretion failed for the three protease knockout strains. However, the enzyme in both the secreted and intracellular protein fraction showed activity. Secretion methods need to be optimised and intracellular expression should be explored. The fluorinase enzyme was successfully cloned and expressed in four PichiaPink™ strains and a biologically active protein was produced.

To the ones I love, I dedicate this to you- I wouldn't be here without you.

To my parents, Chester and Belinda Browne-

thank you for always pushing me to try harder, to aim higher, to be better and for always making so many sacrifices for me.

I will always remember to "keep off the grass".

To my big sister, Kim Browne-

I have always looked up to you, thank you for keeping me out of the looney bin and keeping me going, I would be lost without you.

To my wonderful husband, Brandon Etty-

thank you for helping me through those long hours in the lab and for your patience and support through the tough times. For making me so happy that even the worse days are the best with you.

## Acknowledgements

Firstly, to my supervisor Prof. Karl Rumbold who doesn't just accept you into his lab, he makes you feel at home. This has been a wonderful opportunity, and I am thankful for his guidance.

I am thankful to the IMBL lab for always keeping things so fun and exciting; I enjoyed working with every single member of the team. I am especially thankful for their assistance and advice.

I will be forever grateful for the financial assistance received from the National Research Foundation and Postgraduate Merit Award (University of the Witwatersrand).

A big thanks to my friends and family for their support, it is greatly appreciated.

Above all, I know that nothing I achieve is without the Lord Jesus Christ guiding the way and for this I praise his holy name. "I can do all things through Christ who strengthens me"…Phillipians 4:13.

# Table of Contents

**Declaration** ..... **i**
**Abstract** ..... **ii**
**Dedication** ..... **iii**
**Acknowledgements** ..... **iv**
**Table of Contents** ..... **v**
**List of Figures** ..... **viii**
**List of Tables** ..... **x**
**Table of Abbreviations** ..... **xi**
**1. Literature Review** ..... **1**
**1.1 Recombinant DNA Technology** ..... 1
**1.2 Comparison of *P. pastoris* to other expression systems** ..... 1
*1.2.1 Bacterial hosts* ..... 1
*1.2.2 Yeast hosts* ..... 2
**1.3 *P. pastoris* as an expression system** ..... 3
*1.3.1 Methanol metabolism and respiratory growth* ..... 3
*1.3.2 AOX1 promoter* ..... 4
**1.4 Heterologous protein production using the *P. pastoris* system** ..... 4
*1.4.1 Selection of a host strain and expression vector* ..... 5
*1.4.2 Transformation and gene integration* ..... 6
*1.4.3 Limitations of P. pastoris* ..... 9
**1.5 An Overview of the PichiaPink™ expression system** ..... 10
*1.5.1 Selection in PichiaPink™* ..... 11
*1.5.2 Extracellular secretion and the α-mating factor secretion signal* ..... 12
*1.5.3 Protease knockout strains* ..... 13
**1.6 The Target Protein- fluorinase** ..... 13
**2. Aims and Objectives** ..... **15**
**2.1 Aims** ..... 15
**2.2 Specific objectives** ..... 15
**3. Materials and Methods** ..... **16**
**3.1 Strains, plasmids and general reagents** ..... 16

**3.2 Confirmation of the *fLA* gene in pUC-57** ..... 16

*3.2.1 Screening for positive clones using colony PCR* ..... 16

**3.3 Construction of recombinant PichiaPink™ expression vectors** ..... 17

*3.3.1 Primer design and PCR based cloning* ..... 17

*3.3.2 Screening for positive clones and sequencing* ..... 21

**3.4 Transformation of PichiaPink™ strains with linearized pPinkα-HC- *fLA*** ..... 21

*3.4.1 Preparation of transforming DNA* ..... 21

*3.4.2 Preparation of electrocompetent cells* ..... 22

*3.4.3 Transformation of* Pichia*Pink™ strains by electroporation and selection of transformants* ..... 23

**3.5 PCR analysis of PichiaPink™ integrants** ..... 24

**3.6 Expression of recombinant PichiaPink™ strains and optimisation** ..... 25

**3.7 SDS-PAGE analyses of secreted protein expression** ..... 25

**3.8 Comparison of cell disruption methods for protein extraction** ..... 26

*3.8.1 Bead milling* ..... 26

*3.8.2 Osmotic shock* ..... 26

*3.8.3 Sonication* ..... 26

*3.8.4 Freeze crushing* ..... 27

**3.9 Analysis of intracellular protein expression by SDS-PAGE** ..... 27

**3.10 Fluorinase enzyme activity assay using HPLC** ..... 27

**4. Results** ..... **29**

**4.1 Confirmation of the *fLA* gene in pUC-57** ..... 29

**4.2 Construction of recombinant pPinkα-HC-*fLA* vectors** ..... 31

*4.2.1 PCR Based Cloning* ..... 31

**4.3 Transformation of PichiaPink™ strains with linearized pPinkα-HC- *fLA*** ..... 35

*4.3.1 Preparation of transforming DNA* ..... 35

*4.3.2 Transformation of* Pichia*Pink™ strains by electroporation and selection of transformants* ..... 37

**4.4 PCR analysis of PichiaPink™ integrants** ..... 39

**4.5 Expression and analysis of secreted protein expression by SDS-PAGE** ..... 41

**4.6 Comparison of cell disruption methods for protein extraction** ..... 43

**4.7 Protein extraction and Analysis of intracellular protein expression by SDS-PAGE** ..... 43

**4.8 Fluorinase enzyme activity assay using HPLC** ..... 45

**5. Discussion** ..... **52**

**5.1 Construction of recombinant PichiaPink™ vectors (pPinkα-HC) with the gene of interest** ..... 52

**5.2 Transformation of the four PichiaPink™ strains with the expression construct** ..... 52

*5.2.1 Factors influencing the transformation efficiency* ..... 52

**5.3 Screening of PichiaPink™ clones for integration construct** ..... 54

*5.3.1 ADE2 complementation* ..... 54

*5.3.2 Integration by insertion and genetic instability* ..... 55

**5.4 Comparison of mechanical disruption methods for protein extraction** ..... 58

**5.5 Analysis of protein expression by SDS-PAGE** ..... 60

*5.5.1 Analysis of the secreted protein fraction by SDS-PAGE and the problems associated with the α-MF secretion signal* ..... 60

*5.5.2 Analysis of intracellular protein fraction by SDS-PAGE* ..... 62

**5.6 Fluorinase enzyme activity assay using HPLC** ..... 63

**6. Conclusion** ..... **65**

**7. References** ..... **66**

**A. Appendices** ..... **71**

**A.1 Recipes for media and general reagents** ..... 71

**A.2 Preparation of calcium chloride competent cells** ..... 74

**A.3 Preparation of stocks** ..... 75

**A.4 Calculating cell concentration** ..... 77

**A.5 Calculating the breaking efficiency** ..... 77

# LIST OF FIGURES

**Figure 1.1:** The main steps involved in producing a heterologous protein in *P. pastoris* ..... **5**

**Figure 1.2:** Homologous recombination in *P. pastoris* ..... **8**

**Figure 1.3:** Tandem multiple integration due to repeated insertion events ..... **9**

**Figure 1.4:** The pPinkα-HC plasmid for secreted protein expression ..... **11**

**Figure 1.5:** The fluorinase catalyses a substitution between SAM and fluoride ion to produce 5'FDA ..... **14**

**Figure 3.1:** Cloning and ligation of the gene of interest ..... **18**

**Figure 4.1:** Colony PCR for screening for the pUC-57-*fLA* construct ..... **30**

**Figure 4.2:** PCR based cloning for the construction of pPink α–HC-*fLA* ..... **32**

**Figure 4.3:** Colony PCR to screen for the pPink α–HC-*fLA* construct using gene specific primers ..... **34**

**Figure 4.4:** Linearization of the pPink α–HC parent plasmid (7.9 kb) and pPink α–HC-*fLA* construct (8.8 kb) at the *TRP2* region for integration ..... **36**

**Figure 4.5:** Following transformation of PichiaPink™ strains with the linearized expression construct (pPink α–HC-*fLA*) cells were plated on PAD selection agar ..... **38**

**Figure 4.6:** Colony PCR for direct testing of PichiaPink™ clones for insertion of the *fLA* gene with *AOX1* and *CYC1* primers ..... **40**

**Figure 4.7:** SDS-PAGE analysis of secreted protein expression. Expression induced with 0.5 % methanol and culture supernatant analysed on 16% SDS-PAGE gels ..... **42**

**Figure 4.8:** Comparison of the protein released after cell disruption, the S- soluble protein and P-pellet separated on 16 % SDS gels ..... **44**

**Figure 4.9:** SDS-PAGE analysis of intracellular protein expression ..... **44**

**Figure 4.10:** HPLC profile of 5'-FDA and SAM standards for the fluorinase enzyme activity assay ..... **46**

**Figure 4.11:** HPLC profile of the fluorinase enzyme assay testing the secreted protein fraction for activity ..... **47**

**Figure 4.12:** HPLC profile of the breaking buffer -to determine background peaks for intracellular enzyme activity assays ..... **48**

**Figure 4.13:** HPLC profile of the fluorinase enzyme assay testing the intracellular protein fraction for activity ..... **49**

**Figure 4.14:** HPLC profile of the fluorinase enzyme assay testing the intracellular protein fraction of strain 3 for activity ..... **50**

**Figure 4.15:** HPLC profile of the fluorinase enzyme assay testing the intracellular protein fraction of strain 4 for activity ..... **51**

**Figure 5.1:** Recombination between repeat sequences lead to deletions and insertion in chromosomal DNA ..... **56**

**Figure 5.2:** Illustrates the mechanism of the loop-effect that leads to deletion of whole or parts of the expression cassette in *P. pastoris* ..... **57**

**Figure 5.3:** Sequence of the alpha mating factor secretion signal, with the *Kex2* and Ste 13 signal cleavage sites ..... **61**

**Figure A.1:** The 900 bp nucleotide sequence of the *fLA* gene and the corresponding 299 amino acid sequence of the fluorinase from *Streptomyces cattleya* ..... **76**

**Figure A.2:** Sequence of the alpha mating factor secretion signal, the *Kex2* cleavage site is located after the final Arg amino acid PichiaPink™ pPinkα-HC vectors ..... **77**

# LIST OF TABLES

**Tables 3.1:** Protocol for restriction digestion of the gene insert and pPinkα-HC plasmid .... **19**

**Tables 3.2:** Reaction setup for the ligation of the digested DNA fragments, *fLA* gene and pPinkα-HC plasmid ..... **20**

**Tables 3.3:** Restriction digestion protocol for linearization of the recombinant and parent pPinkα-HC plasmid ..... **22**

**Tables 3.4:** Electroporation parameters used for transformation with the ECM® 830 electroporation system ..... **23**

**Tables 4.1:** Comparison of four cell disruption methods**,** showing the breaking efficiency and concentration of the soluble protein after disruption ..... **44**

# LIST OF ABBREVIATIONS

| | |
|---|---|
| 5-FDA | 5’-fluorodeoxyadenosine |
| 5-FDA synthase | 5’-fluoro-5’-deoxyadenosine synthase |
| *ADE* | phosphoribosylaminoimdazole carboxylase gene |
| Amp | ampicillin |
| $Amp^R$ | ampicillin resistance |
| AOX | alcohol oxidase |
| ARG4 | arginine 4 gene |
| BMGY | Buffered complex media containing glycerol |
| BMMY | Buffered complex media containing methanol |
| bp | base pair |
| C | capacitance |
| $dH_2O$ | distilled water |
| DNA | Deoxyribonucleic acid |
| DTT | dithiothreitol |
| EDTA | ethylene diamine tetraacetic acid |
| *ENO1* | enolase gene |
| ER | endoplasmic reticulum |
| etc | etcetera |
| F | farad |
| *fLA* | fluorinase gene |
| *FLD1* | formaldehyde dehydrogenase 1 gene |
| g | grams |
| *GAP* | glyceraldehyde phosphate gene |

| | |
|---|---|
| *GLUT1* | glucose transporter 1 gene |
| HCl | hydrochloric acid |
| HIS4 | histidine 4 gene |
| HPLC | high pressure liquid chromatography |
| *Kex2* | kexin protease |
| kDa | kilodaltons |
| L | litres |
| LB | Luria-Bertani |
| LiAc | lithium acetate |
| MCS | multiple cloning site |
| mg | milligrams |
| ml | millilitres |
| mRNA | messenger RNA |
| n | nano |
| NaCl | sodium chloride |
| OD | optical density |
| PAD | Pichia adenine dropout |
| PCR | polymerase chain reaction |
| *PEX8* | peroxisomal biogenesis factor 8 gene |
| pH | potential of hydrogen |
| PMSF | phenylmethylsulfonyl fluoride |
| SAM | S-adenosyl-L-Methionine |
| SDS-PAGE | sodium dodecyl sulphate- poly-acrylamide gel electrophoresis |
| SOC | super optimal broth with catabolite repression |

| | |
|---|---|
| *Ste13* | dipeptidyl peptidase IV gene |
| TAE | Tris base-acetic acid-EDTA |
| TEMED | tetramethyethlylenediamine |
| URA3 | uracil 3 gene |
| UV | ultraviolet |
| V | voltage |
| YNB | Yeast nitrogen base |
| YPD | yeast peptone dextrose |
| YPDS | yeast peptone dextrose sorbitol |
| *YPT1* | yeast protein two 1 gene |
| α | alpha |
| µ | micro |
| C° | degrees celsius |

# 1. Literature Review

## 1.1 Recombinant DNA Technology

One of the major uses for recombinant organisms is to manipulate them in order to produce a foreign gene and consequently express foreign protein. Many proteins are used in industry and are of high commercial value. However obtaining the protein directly from the host is not always viable (Li *et al,* 2011). For example, when the desired protein is produced at very low levels in the host, or when the protein is not always expressed readily in the host, heterologous expression would be more applicable. Producing proteins easily, inexpensively and completely active are some of the important factors relating to the production of proteins (Daly *et al,* 2005). The choice of expression system for the production of heterologous protein is a key factor. A plethora of expression systems exist, with some of the most commonly used being the bacterial *Escherichia coli* systems and fungal *Saccharomyces cerevisiae* system. This study will explore *Pichia pastoris* for the production of industrial enzymes due to its many advantages over traditional systems, and are explored below:

## 1.2 Comparison of *P. pastoris* to other expression systems

### *1.2.1 Bacterial hosts*

The first heterologous protein was expressed in *E. coli* in 1977 by Itakura *et al.,* (1997); over the past four decades prokaryotes such as *E. coli* have been one of the most well studied and commonly used expression systems in both the laboratory and industrial setting. However, there are many limitations associated with using prokaryotic systems that become especially evident when the protein to be expressed is that of a more complex protein. Prokaryotic systems are simple in nature and lack intracellular organelles, such as the golgi apparatus and endoplasmic reticulum present in eukaryotes, which function in modifying proteins (*Li et al.,* 2011). These modifications include phosphorylation, glycosylation, disulphide-bond formation and correct folding.

The inability of prokaryotes to correctly fold heterologous proteins can lead to the production of an inactive protein, as well as insoluble and misfolded inclusion bodies (Daly *et al.,* 2005). This is one of the major issues with prokaryotic systems such as *E. coli.* Consequently it becomes necessary to perform additional purification steps, including solubilising and re-folding of the protein which is both time consuming and an added expense, reducing commercial viability (Harbron, 2009).

Other limitations with *E. coli* hosts includes the occurrence of translational errors, as these systems do not recognise rare codons and lack the intracellular machinery for splicing introns when processing mRNA; leading to the production of non-functional proteins (Khow and Suntrarachun, 2012). Another occurrence with *E. coli*- produced proteins is the retention of the amino-terminal methionine; this causes instability and immunogenicity a major drawback (Daly *et al.,* 2005).

Due to the many limitations of prokaryotic expression systems, the system cannot be applied broadly to protein production, especially when the foreign protein is from a eukaryotic host, requires specialised modification, or even if solubility (inclusion body formation) is a problem and purification steps require minimisation (Daly *et al.,* 2005).

### *1.2.2 Yeast hosts*

Eukaryotic expression systems have been explored in the recent years with focus on mammalian and yeast systems. However, the choice of a mammalian system for the production of heterologous protein is not always viable as these systems are expensive to use as well as time consuming which is not attractive to industry. Specialised media and growth conditions are required; an overall sensitive system that is not easy to work with and prone to viral contamination (Cregg *et al.,* 2000; Macauley-Patrick *et al.,* 2005). Yeast systems like the *P. pastoris* system is considered easier, less expensive and faster to use in comparison to other non-yeast eukaryotic systems (Cregg *et al.,* 2000).

Yeasts, like prokaryotes are single celled organisms, grow rapidly and are easily manipulated, where they can be transformed with DNA. An advantage of this system is that it is eukaryotic, thus making it capable of implementing posttranslational modifications. These include glycosylation, proteolytic processing, disulphide bond formation and producing the correct folding patterns (Cregg, 2007).

*S. cerevisiae* was the first eukaryotic system used to produce heterologous proteins; this was mainly due to its safe history in producing commercial fermented products. *S. cerevisiae* is a well characterised system and has been applied to the production of a large array of proteins, however, it also has a number of limitations. A major problem with *S. cerevisiae* for the production of proteins was its primitive glycosylation pathway and tendency to produce hyperglycosylated proteins (Curran and Bugeja*,* 2009; Lin-Cereghino and Lin-Cereghino, 2007). This generated the need for an alternative yeast host and *P. pastoris* was explored (Curran and Bugeja*,* 2009). It grows to high cell densities producing proteins on the milligram-gram scale, it also produces shorter glycosyl chains producing a more authentic

protein (Ahmad *et al.*, 2014; Aw and Polizzi, 2013). In the past twenty years *P. pastoris* has become one of the more popular yeast expression systems.

## 1.3 *P. pastoris* as an expression system

*P. pastoris* was developed as an expression system in the 1980s and plays a critical role in biotechnology and the large-scale production of heterologous proteins. It is commonly applied to the production of proteins for molecular medicines, biopharmaceuticals and industrial enzymes used in biocatalysis (Ahmad *et al.*, 2014; Cregg *et al.*, 2000). *P. pastoris* has been used to express a wide variety of heterologous protein, both from prokaryotes and eukaryotes; including those from higher organisms. The system is useful for producing large-scale quantities of protein and this is useful for both laboratory research and in an industrial setting (Macauley-Patrick *et al.,* 2005). The expression system has many advantages for the production of heterologous proteins.

### *1.3.1 Methanol metabolism and respiratory growth*

Unlike *S. cerevisiae* which is fermentative yeast, *P. pastoris* is non-fermentative, thereby not fermenting carbon sources (Cregg, 2007). High cell densities are reached firstly, because carbon sources are converted to biomass, and secondly there is no build-up of ethanol, a toxic product of fermentation which hampers growth (Cereghino and Cregg 2000).

A feature somewhat unique to *P. pastoris* is that it is one of only 30 species of yeast that metabolise methanol. Methylotrophic yeast species are only found in two genera, namely *Pichia* and *Candida* (Cregg *et al.,* 2000). During the process of methanol metabolism the oxidation of methanol to formaldehyde produces hydrogen peroxide, catalysed by a peroxisomal matrix enzyme alcohol oxidase (AOX). Hydrogen peroxide can be toxic to cells, however the process of methanol metabolism takes place in the peroxisomes where it is removed, limiting toxicity to the cell (Cereghino and Cregg 2000; Macauley-Patrick *et al.,* 2005). The by-product hydrogen peroxide is then degraded to water and oxygen by catalase (Cat) a peroxisomal enzyme. The formaldehyde either generates energy for the cell through oxidation or is ultimately converted to glyceraldehyde-3-phosphate. The majority of foreign genes expressed in *P. pastoris* are transcribed under the control of the *AOX1* promoter and therefore a key feature is transcription in response to methanol (Lin-Cereghino *et al.,* 2006).

### *1.3.2 AOX1 promoter*

Ellis *et al.,* (1985) demonstrated that the *AOX* gene is controlled at the level of transcription and is a one of three methanol-regulatable genes. Cregg *et al.,* (1989) elucidated that *P. pastoris* has two genes that code for alcohol oxidase, these are the *AOX1* and *AOX2* genes, where the *AOX1* gene was responsible for most of the alcohol oxidase activity in the cell. Koch *et al.,* (2016) revealed that the *AOX* promoter is found in the nucleosome where it is unavailable to the transcription machinery. Upon induction with methanol then only will the AOX genes be released from the nucleosome and be available for transcription (Koch *et al.,* 2016). This is a tightly regulated system making the AOX1 gene highly favoured as a promoter for the production of heterologous proteins.

*P. pastoris* produces large amounts of the AOX1 enzyme (up to 30% of the total cellular protein) when induced with methanol as the sole carbon source (Couderc and Baratti, 1980), which may be compensation for the enzymes low affinity for oxygen (Koch *et al.,* 2016). When cells are grown in glucose or glycerol the *AOX1* gene is repressed and completely undetectable in the cellular protein content. The *AOX1* gene is regulated by a repression/depression mechanism and an induction mechanism, where methanol induces high levels of transcription (Cereghino and Cregg 2000).

The *AOX1* promoter is a strong and tightly regulated inducible promoter. This feature is advantageous being that the expression of heterologous proteins can be tightly controlled. Cells are grown on glucose or glycerol and protein expression is induced upon the addition of methanol. Growth on glycerol or glucose can be used to repress the promoter until high cell densities are reached. Thereafter expression can be induced with methanol; this is a benefit when expressing toxic protein, which may adversely affect the hosts metabolism or functioning (Daly *et al.,* 2005). The fact that the *AOX1* gene is tightly controlled- strong transcriptional activation, and is tightly regulated as it can only be induced with methanol, makes the *AOX1* gene the most suitable and favoured promoter to drive expression of foreign proteins in *P. pastoris.*

## 1.4 Heterologous protein production using the *P. pastoris* system

To express a foreign gene in *P. pastoris* three basic steps must be achieved, the foreign gene must be inserted into an expression vector, the expression construct must be integrated into the host genome and finally the strains must be examined for expression of the gene of interest (Macauley-Patrick,*et al.,*2005). A cloning strategy should be devised which includes selecting a host strain and an expression vector. These will be reviewed in

the following section. Figure 1.1 illustrates the steps involved in producing a heterologous protein in *P. pastoris,* from the cloning of the gene of interest into the expression vector to the expression of the protein. Considerations when selecting a host strain and expression vector is outlined in figure 1.1.

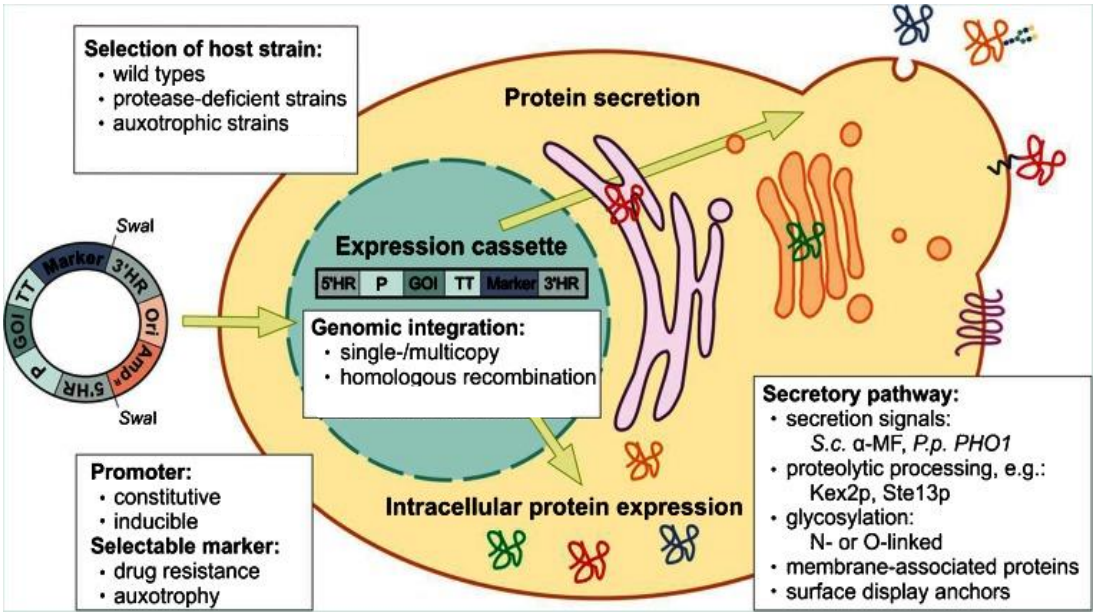



**Figure 1.1:** The main steps involved in producing a heterologous protein in *P. pastoris* as well as the main consideration when selecting a host strain and expression vector. (Adapted from Ahmad *et al.,* 2014)

### *1.4.1 Selection of a host strain and expression vector*

All *P. pastoris* strains are derived from NRRL-Y 11430 at the Northern Regional Research Laboratories, Peiria, IL (Cregg *et al.,* 2000). A large variety of *P. pastoris* vectors and strains are available. Selection of a particular strain and vector is ultimately dependant on the type of protein to be expressed.

*P. pastoris* vectors are constructed as *E.coli- P. pastoris* shuttle vectors (Cereghino and Cregg, 2000). This bifunctional system allows for replication in *E. coli* and maintenance in *P. pastoris* (Ahmad *et al*., 2014). The vector contains DNA sequences for expression in the yeast host system and bacterial sequences that allow for the gene to be cloned and replicated in *E. coli.* The bacterial backbone contains an origin of replication (ORI) to enable replication in *E. coli.* Bacterial sequences include a selectable marker, most commonly $Amp^R$

which codes for ampicillin resistance and is used for selection in early cloning steps (Curran and Bugeja, 2009).

The vector also contains sequences for selection in *P. pastoris,* selection may be based on auxotrophy or antibiotic resistance. Zeocin™, blasticidin S. and geoticin resistance are the antibiotic resistance genes that are available for *P. pastoris* (Ahmad *et al.,* 2014) Auxotrophically marked host strains are available with complementing gene containing vectors, most commonly *HIS4, ARG4, ADE1* and *URA3* (Ahmad *et al.,* 2014, Lin-Cereghino and Lin-Cereghino, 2007).

Expression vectors contain an expression cassette with a promoter region most commonly the *AOX1* promoter to drive mRNA production, followed by a multiple cloning site (MCS) for insertion of the foreign gene and a transcription termination region (Cregg and Creghino, 2000, Curran and Bugeja, 2009). As the *AOX1* promoter is not always suitable, a variety of promoters are available for *P. pastoris* which include both inducible and constitutive promoters. Alternative promoters include the *P. pastoris FLD1, GAP, PEX8, GLUT1, ENO1* and *YPT1* (Cregg and Creghino, 2000; Macauley-Patrick *et al.,* 2005). The expression cassette also contains unique sites for linearization, for integration of the expression cassette into the host genome. In *P. pastoris* these sites are most commonly the *AOX1* promoter or the locus for selection (Ahmad *et al*., 2014).

Vectors for secretion and intracellular expression are available for *P. pastoris* (Ahmad *et al.,* 2014). Extracellular expression is achieved by fusion of a secretion signal upstream of the MCS. The most common secretion signals used for *P. pastoris* are the acid phosphatase signal, native to *P. pastoris* and the α-mating factor and invertase signal derived from *S. cerevisiae* (Ahmad *et al.,* 2014). According to Cereghino and Cregg, (2000) *P. pastoris* secretes low levels of endogenous protein which allows for easy purification as the foreign protein will make up the higher portion of total protein in the medium. The option of secretion is usually only selected for proteins that are secreted in the native host (Cregg *et al,* 2000).

### 1.4.2 *Transformation and gene integration*

*P. pastoris* is transformed by integration of the gene of interest into the host genome. Integration occurs via homologous recombination, where the vector and host genome contain regions of homology. Orr-Weaver *et al.,* (1981) showed that ends of linear DNA (broken ends) are highly recombinogenic and interact directly with homologous chromosal DNA. Homologous recombination is important for genetic variation during sexual reproduction. It is also important for DNA repair synthesis, where DNA double strand breaks

initiate homologous recombination. This is the hallmark for integration of foreign DNA into a host genome by a linearized plasmid containing homologous ends (Aguilera *et al.,* 2000).

The first type of integration event occurs when the vector is digested in a unique site with a restriction enzyme, this is most commonly in the selection marker gene (*HIS4* or *ARG4 etc*,) or in the *AOX1* promoter region (Li *et al.,* 2007). This type of recombination leads to insertion of the foreign DNA into the host genome. The free linearized DNA stimulates the vector to recombine at the cut locus via a single crossover event, illustrated in Figure 1.2A-B (Daly *et al.,* 2005).

The second type of recombination occurs when the vector is digested in two unique sites and the free DNA ends are homologous to two regions in the host genome. This is most commonly achieved by digesting the vector at the 5' *AOX1* promoter and in a second *AOX1* gene (Li *et al.,* 2007). This leads to the deletion/replacement of a portion of the host genome with the gene of interest, illustrated in figure 1.2C (Daly *et al.,* 2005).

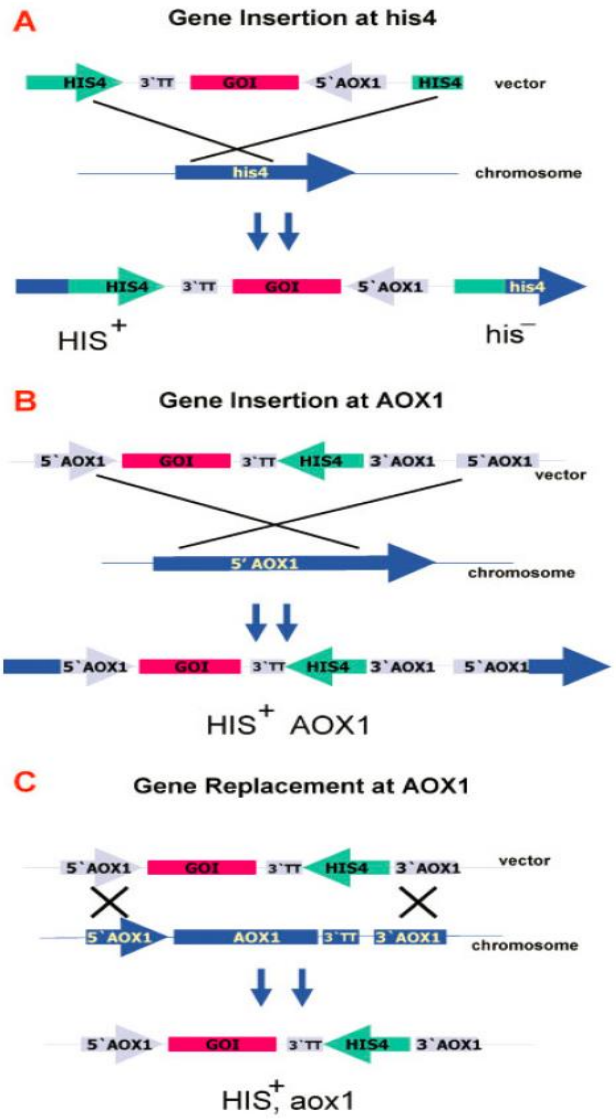



**Figure 1.2:** Homologous recombination in *P. pastoris.* Integration of the gene of interest via **A-B:** gene insertion or **C:** gene replacement (Adapted from Daly *et al.,* 2005).

Gene replacement leads to the production of single copy transformants, which are more genetically stable. On the other hand integration by insertion leads to less genetically stable transformants (Daly *et al.,* 2005). Integration by insertion can to lead multiple insertion events where multiple copies of the expression cassette are inserted with frequencies between 1-10%. Multiple copy transformants occur by tandem multiple integration events due to repeated recombination events (Daly *et al.,* 2005, Romanos *et al.,* 1992). After the first integration event between the vector and the host genome, second and third events can occur with additional free plasmids (Sunga *et al.,* 2008). The process of tandem multiple integration is illustrated in figure 1.3 below.

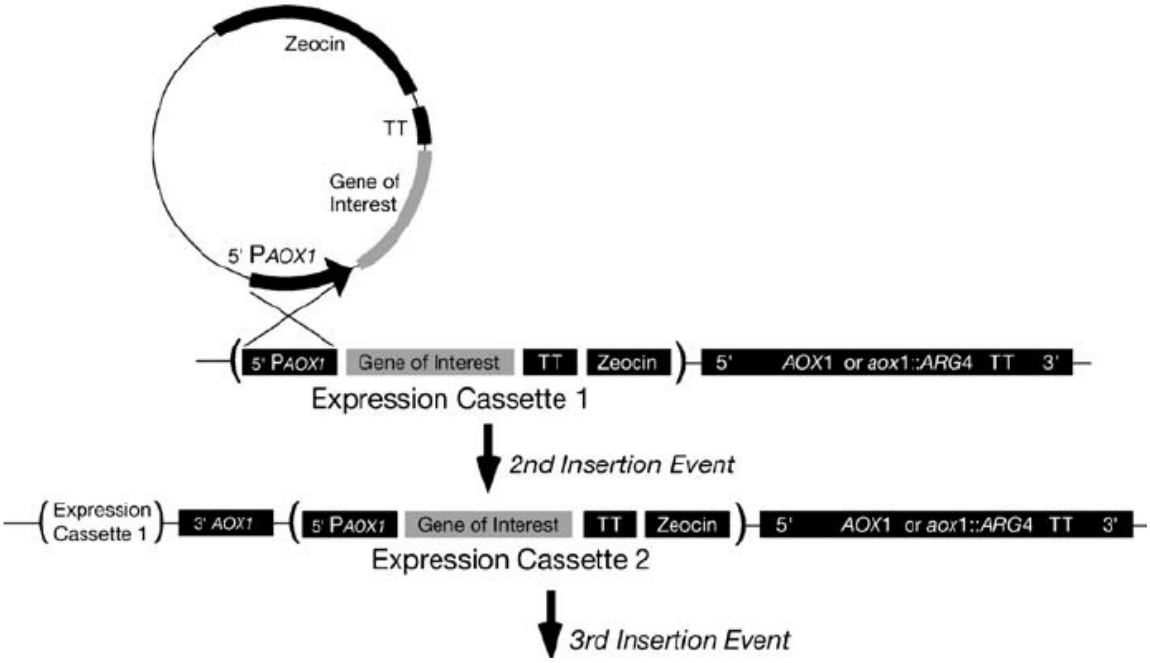



**Figure 1.3:** Tandem multiple integration due to repeated insertion events. The initial integration event occurs between the vector and the host genome, second, third and more events can occur with additional free plasmids, leading to the production of transformants containing multiple copies of the expression cassette (adapted from Li *et al.,* 2007).

Transformation of the expression cassette into the host genome can be achieved by, sphleroplast formation, lithium chloride treatment or by electroporation (Daly *et al,* 2005). According to Wang *et al.,* (2004) electroporation results in an increased uptake of plasmid DNA, compared to other transformation methods. The method of spheloroplast formation also yields high transformation efficiencies but is a long and cumbersome method when compared to electroporation (Wu *et al.,* 2004). The electroporation method will be used to transform the yeast cells in this study. Electroporation is used to transfer biological molecules across cell membranes (Richey *et al.,* 1989). The host cells and DNA are mixed in a conductive solution and held within an electric circuit. A high voltage electric pulse is then applied to the circuit containing the cell mixture. The pulse causes temporary pores to form in the cell membrane and a subsequent rise in the electric potential allows DNA molecules to pass through (Shigekawa and Dower, 1988).

### *1.4.3 Limitations of P. pastoris*

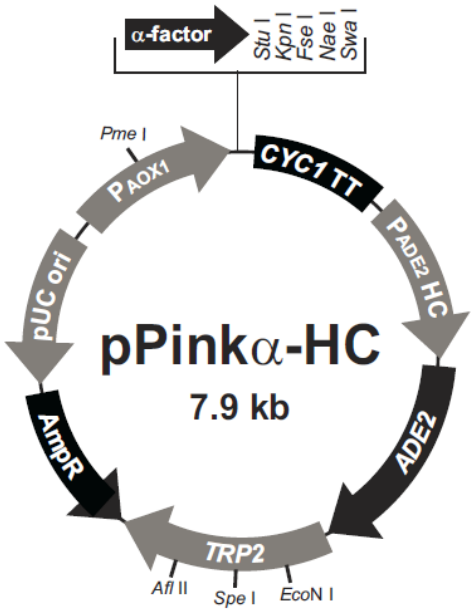
While there are many advantages of using *P. Pastoris* as an expression host, limitations and disadvantages exist as is the case with any biological system. The process of transformation is inefficient as DNA must enter the cell as well as integrate into the host genome, yielding low transformation efficiencies (Wu and Letchworth, 2004). When expressing a protein with

no commercially available antibody, screening for multi-copy clones is also difficult as a simple western blot cannot be done (Li *et al.,* 2011). For secreted protein expression, especially during large-scale production in a bioreactor, proteases are a major concern as they are secreted into the media and can lead to degradation of the foreign protein (Li *et al,* 2007).

In recent years, PichiaPink™ has been developed to overcome the problems that have been associated with *P. pastoris.* PichiaPink™ is a production strain of *P. pastoris* that has been developed by Invitrogen (Life Technologies Corporation, Carlsbad, CA, USA).

### 1.5 An Overview of the PichiaPink™ expression system

The PichiaPink™ expression system from Invitrogen is a production strain of *P. pastoris* and offers many advantages over other traditional *P. pastoris* systems. The system offers two types of copy vectors, a low and high copy. Additionally, it offers high copy vectors constructed with the commonly used α-mating factor secretion signal sequence derived from *S. cerevisiae* (figure 1.4). Four strains are available, one wild-type and three protease knock-out strains allowing you to screen different strains for the highest expression. The system provides a new efficient method to screen for multi-copy transformants based on colour formation. The PichiaPink™ strains are *ADE2* auxotrophs and selection is based on *ADE2* complementation.




**Comments for pPinkα-HC**
**7898 nucleotides**

5´ *AOX1* promoter region: bases 1-940
α-mating factor secretion signal: 941-1194
Multiple cloning site: bases 1196-1222
*CYC1* transcription termination region: bases 1228-1502
*ADE2* promoter region: bases 1517-1529
*ADE2* ORF: bases 1530-3215
*TRP2* gene: bases 3439-5250
pUC ori: 5619-6292 (c)*
Ampicillin (*bla*) resistance gene: bases 6437-7297 (c)

*(c): complementary strand

**Figure 1.4** The pPinkα-HC plasmid for secreted protein expression, contains α-MF secretion signal from *S. cerevisiae,* the ADE2 gene and its promoter for selection and a pUC19 bacterial backbone with $Amp^R$ (adapted from invitogen).

### *1.5.1 Selection in PichiaPink™*

*ADE2* complementation allows for visual selection of recombinant PichiaPink™ strains. The PichiaPink expression vectors contain the functional ADE2 gene and the full *ADE2* gene is knocked out in PichiaPink™ strains (Invitrogen). The *ADE2* gene encodes for phosphoribosylaminoimdazole carboxylase which functions in the biosynthesis of purine nucleotides (Li *et al.,* 2011). When the *ADE2* gene is mutated it leads to the colonies becoming red in colour, due to the accumulation of purine precursors in the vacuole. This phenotype can be used as a method for screening positive clones. PichiaPink™ strains require adenine in the media to grow as they are *ade*2 auxotrophs. The PichiaPink™ expression plasmids have the *ADE2* gene and when the PichiaPink™ strains are transformed with these expression plasmids the *ADE2* gene is re-introduced into the *P.*

*pastoris* chromosome and can therefore grow on media lacking adenine (Li *et al.,* 2011; Invitrogen).

After transformation the adenine biosynthetic pathway is re-established and the red phenotype is reverted back to the wild-type white phenotype. White colonies represent colonies that have been successfully transformed. This eliminates the need for other screening methods such as selection based on antibiotic resistance (Ahmed *et al.,* 2004; Invitrogen). Following transformation, both white and pink colonies will be observed, which can be related to their expression levels. This is because the colour of the colony is directly related to the copy number of the plasmid. In the pink colonies very little of the ADE2 gene is expressed and in white colonies a larger amount is expressed, indicating the white colonies have a higher copy number of the plasmid (Du *et al.,* 2012; Li *et al,* 2011).

The high and low copy vectors express the *ADE2* gene from different promoter lengths, which relates to the copy number of the gene. High copy vectors have a truncated *ADE2* promoter of 13 bp and the low copy vectors have a 82 bp ADE2 promoter (Du *et al*, 2012). For PichiaPink™ strains to grow on media lacking adenine, sufficient amounts of the *ADE2* gene must be expressed from the marker gene on the PichiaPink™ vector (Invitrogen). In the case of the low copy vector, lower copies of the integrated expression cassette are required, as the *ADE2* promoter is stronger. In the case of the high copy vector, higher copies of the integrated expression cassette are required for the PichiaPink™ strains to grow on media lacking adenine as the promoter strength of the high copy vector is weaker (Du *et al*, 2012).

### *1.5.2 Extracellular secretion and the α-mating factor secretion signal*

Selecting an expression vector for secreted protein production is attractive in industry as, product secreted to the culture media relieves the need for expensive and time consuming purification steps (Aw and Polizzi, 2013). The pPinkα-HC vector contains the *S. cerevisiae* α-mating factor secretion signal for targeted protein secretion. This is the most commonly used secretion signal used in the production of secreted proteins in *P. pastoris* (Daly *et al.,* 2005). The α-MF signal is composed of a pre-pro leader sequence, the pre-sequence is 19 amino acids and is followed by a 66 amino acid pro-sequence (Glick and Fitzgerald, 2014). Processing of the signal sequence and folding of the foreign protein involves several steps.
The pre-sequence functions to direct the protein into the secretory pathway, and is then removed during the translocation of the protein in the endoplasmic reticulum. Protein folding and other modifications begin in the endoplasmic reticulum (Glick and Fitzgerald, 2014). The

pro-protein is transported to the golgi apparatus. In the golgi the pro-sequence is removed by *Kex2*, an endo-peptidase that is produced in the golgi. The mature protein is then packaged in secretory vesicles and transported to the cell surface (Daly et al 2005; Lin-Cereghino *et al.,* 2013).

### *1.5.3 Protease knockout strains*

When the protein of interest is targeted for secretion, there is a risk of the protein being degraded by proteases. During the process of fermentation, proteases are secreted into the medium which leads to the subsequent degradation of the desired protein, thereby decreasing the overall yield of the expressed protein (Cregg, 2007). Similarly, when the protein is expressed intracellularly, during protein purification, the protein is at risk of degradation due to the release of proteases in the subsequent lyses steps (Macauley-Patrick *et al.,* 2005).

The PichiaPink$^{TM}$ expression system offers four strains of which, three are protease knock-out strains. PichiaPink$^{TM}$ strain 1 is a protease wild-type, PichiaPink$^{TM}$ strain 2 is *pep4 knock-out,* PichiaPink$^{TM}$ strain 3 is a *prb1 knock-out and* PichiaPink$^{TM}$ strain 4 is a *pep4 and prb1 knock-out* (Invitrogen). The *PEP4* gene codes for proteinase A which activates carboxypeptidase Y and proteinase B (encoded by *PRB1)* (Cregg *et al.,* 2000; Li *et al.,* 2007). The availability of more than one strain in the PichiaPink$^{TM}$ expression system provides the option of selecting a strain that produces the desired protein levels before large scale expression. This utility facilitates time saving and also reduces overall costs, which is advantageous.

## 1.6 The Target Protein- fluorinase

In 1986 *Streptomyces cattleya* a soil bacterium, was reported to produce fluoroacetate and fluorothreonine. This organism has been used to study fluorometabolite biosynthesis and fluorination enzymology (O' Hagan and Deng, 2015). The first enzyme committed to fluormetabolite production, 5'-fluoro-5'-deoxyadenosine (5'FDA) synthase (EC 2.5.1.63) also called fluorinase was first isolated from *S. cattleya* in 2003 (Schaffrath *et al.,* 2003).

The fluorinase enzyme catalyses the reaction between S-adenosyl-L-Methionine (SAM) and fluoride ion to produce 5'-fluorodeoxyadenosine (5'-FDA) and L-Methionine (Deng *et al.,* 2004; O' Hagan and Deng, 2015) (figure 1.5). The identified fluorinase was found to

participate in secondary metabolism by delivering fluorinated products when *S. cattleya was* grown in the presence of fluoride (O' Hagan and Deng, 2015).

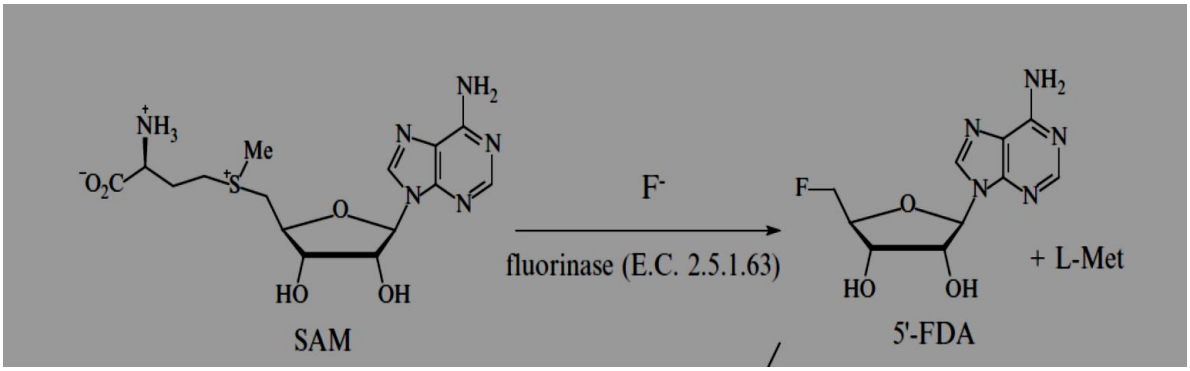



**Figure 1.5:** The fluorinase catalyses a substitution between SAM and fluoride ion to produce 5'FDA (Adapted from O' Hagan and Deng, 2015).

After isolation of the first fluorinase enzyme, four more fluorinase genes were found through genome mining, however, the number of fluorinase genes being identified remains relatively few (O' Hagan and Deng, 2014). This is due to the fact that natural sources of fluorinated compounds are uncommon in nature.

The field of synthesizing fluorinated organic compounds has increased over the years due to their significance in agrochemical and pharmaceutical industries. Fluorine substitutions are used in these fields as fluorine has a great effect on membrane permeability, metabolic stability and receptor-binding properties. For example fluorine substitution in medical applications increases stability, bioactivity, hydrophobicity and bioavailability of molecules which in turn improves therapeutic indices (Zhang *et al.,* 2012). According to Furuya *et al.,* (2011), approximately 20-30% of drugs on the market contain a fluorine molecule. Methods have been developed to synthesise fluorinated compounds, however, progress in determining biologically based methods are required (Zhang *et al.,* 2012).

# 2. Aims and Objectives

Identifying potential biocatalysts with significance to industry is one of the major focusses in our lab. Due to the number of processes and modifications that are essential to produce a biologically active protein, the final protein structure, function and activity is dependent on the choice of expression system used. Cost and time are also a major consideration when selecting an expression system for protein production. In most cases, proteins of commercial interest originate from eukaryotes like plants and fungi, and therefore the reliance on bacterial expression systems is not always feasible. The aim of this research was to explore *P. pastoris* as an expression host, where the fluorinase from *Streptomyces cattleya* was selected for expression.

## 2.1 Aim

To clone and express the fluorinase enzyme from *Streptomyces cattleya,* using the PichiaPink™ expression system.

## 2.2 Specific Objectives

1. Construction of recombinant PichiaPink™ vectors (pPinkα-HC) with the gene of interest
2. Transformation of the four PichiaPink™ strains with the expression construct
3. Test PichiaPink™ clones for insertion of the gene of interest
4. Pilot expression studies and optimisation
5. Analysis of protein expression by SDS-PAGE
6. Test activity of the recombinant protein using an enzyme activity assay and HPLC

# 3. Materials and Methods

## 3.1 Strains, Plasmids and General Reagents

Recipes for growth media, plasmid stocks, buffers and solutions are outlined in detail in the appendix. All protocols unless otherwise stated were obtained from the Invitrogen manual provided with the PichiaPink™ secreted protein vector kit (Invitrogen). For cloning and expression the PichiaPink™ expression system (Invitrogen) was used. This included four PichiaPink™ strains and the pPinkα-HC plasmid. For cloning and propagation *E. coli* JM109 cells were used. All DNA and protein quantifications were done using the Quibit®2.0 Fluorometer (Invitrogen).

## 3.2 Confirmation of the *fLA* gene in pUC-57

The full coding sequence of the *fLA* gene is given in the appendix (Figure A.1). The gene was synthesised in a pUC57 plasmid in a previous project. To ensure the plasmid was maintained correctly and to confirm the presence of the gene, chemically competent *E. coli* JM109 cells were transformed with the pUC 57- *fLA* construct. The cells were made competent per the calcium chloride method, the detailed procedure is outlined in appendix (A.2). Zero point five microliters of plasmid DNA were mixed with 50 µl of competent cells and incubated on ice for 20 min. The mixture was heat shocked at 42°C for 45 sec and then cooled on ice for 5 min. Five hundred microliters of SOC media was added and the cells were incubated at 37°C for 1 hr with shaking at 250 rpm. Cells were spread on LB agar plates supplemented with 100 µg/µl ampicillin. The transformation plates were incubated at 37°C overnight. Positive clones were screened for using colony PCR with primers specific to the *fLA* gene.

### *3.2.1 Screening for positive clones using colony PCR*

Using sterile tips a small portion of each colony was scraped off and suspended in 20 µL nuclease free $dH_2O$. The cells were heated at 95°C for 5 min to lyse cells and release plasmid DNA. The cell lysate was centrifuged at 10 000 *x*g for one minute using a MiniSpin Plus benchtop centrifuge (Eppendorf). The supernatant served as the DNA template in the colony PCR.

Approximately 814 bp of the *fLA* gene was amplified from the pUC 57- *fLA* construct using gene specific primers- SC: Forward CGTCCGATCATCGCCTTCAT and SC: Reverse CGTTACGAGCGATGGACAGA. PCR was performed in a final volume of 25 µl containing 1X *Pfu* Buffer, 200 µM dNTPs, 0.5 µM of each primer, 1% DMSO, 1.0 mM $MgCl_2$, 10 µl supernatant and 0.25 µl *Pfu* DNA polymerase (Thermoscientific). A touchdown PCR protocol

was used. Thermal Cycle 1: reactions were run for 30 cycles: denaturation was at 90°C for 30 sec, annealing at 66 °C-55°C for 30 sec and elongation at 72°C for 50 sec. Thermal Cycle 2: reactions were run for 25 cycles: denaturation was at 90°C for 30 sec, annealing at 58 °C for 30 sec and elongation at 72°C for 50 sec. With an initial denaturation at 95°C for 30 sec and a final extension at 72°C for 5 min.

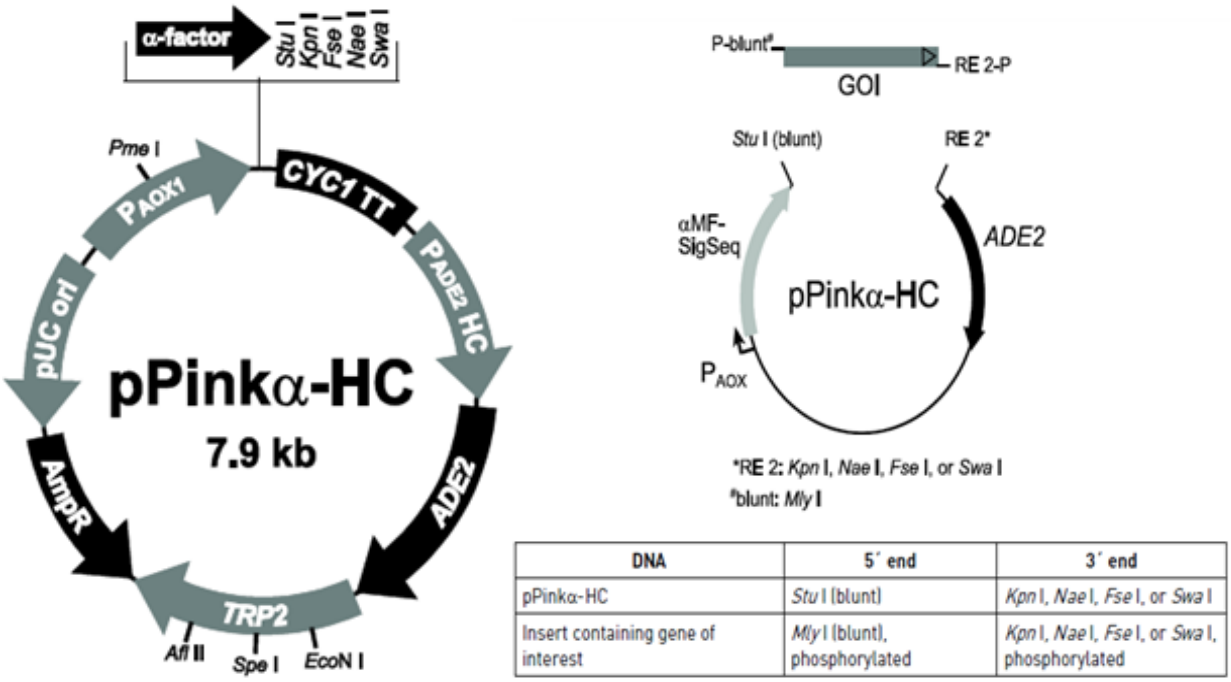
The PCR products were resolved on a 1% agarose- the detailed protocol is outlined in the appendix. Positive colonies containing the pUC 57- *fLA* construct were grown in LB media supplemented with 100 µg/µl ampicillin. The plasmid DNA was isolated using the Zyppy™ Plasmid Miniprep Kit (Zymo Research) according to the manufacturer's instructions.

### 3.3 Construction of recombinant PichiaPink™ expression vectors

In order to copy the *fLA* gene from the pUC 57 donor plasmid to the pPinkα-HC recipient plasmid a PCR based approach was used. PCR-based cloning is a method of copying a piece of DNA from a donor plasmid while simultaneously adding restriction sites to the copy of DNA so that it can be cloned into a recipient plasmid.

#### *3.3.1 Primer Design and PCR based cloning*

To clone the gene of interest upstream of the α-MF signal, a 5' blunt end and a 3' sticky end flanking the gene was required. Figure 3.1A shows the pPinkα-HC expression plasmid and the multiple cloning site. *Fse* I is the only enzyme that did not cut within the gene of interest. Therefore, a forward primer (Phos-FP) with no restriction site was designed that is complementary to the gene, a phosphorylated forward primer was selected. This would provide the 5' blunt end compatible to the plasmids *Stu* I site. A reverse primer (*Fse*I-RP) with an *Fse* I site was designed to create a 3' sticky end compatible to the *Fse* I site on the plasmid. Figure 3.1B outlines the cloning strategy for cloning the gene of interest into the pPinkα-HC plasmid.




| DNA | 5´ end | 3´ end |
|---|---|---|
| pPinkα-HC | *Stu* I (blunt) | *Kpn* I, *Nae* I, *Fse* I, or *Swa* I |
| Insert containing gene of interest | *Mly* I (blunt), phosphorylated | *Kpn* I, *Nae* I, *Fse* I, or *Swa* I, phosphorylated |

**Figure 3.1:** Cloning and ligation of the gene of interest **A.** pPinkα-HC expression plasmid for the extracellular secretion of proteins, **B.** cloning the gene of interest (GOI) into the pPinkα-HC expression plasmid by double restriction digest

The Phos-FP: 5' ATGGCGGCGAACTCCACGC and *Fse*I-RP-:3' ATAGGCCGGCCTTAACGTGCTTCAACACGAGC were synthesised by Inquaba Biotech (Pretoria, South Africa). The parameters for the PCR needed to be optimised as the reverse primer was long (31-32 bases) and caused non-specific binding. Also the GC content of both primers was above 50%, where the GC content of the forward primer was 71.43%. PCR parameters were optimised using several protocols and different DNA polymerases.

Phusion® High-fidelity DNA polymerase (New England Biolabs) was selected as it produces PCR products with blunt ends and a two-step thermocycling protocol recommended for primers with annealing temperatures greater than 72°C can be used. PCR was performed in a final volume of 20 µl containing 1X Phusion HF buffer, 200 µM dNTPs, 3% DMSO, 1.5 mM $MgCl_2$, 0.5 µM of each primer, 0.2 µl Phusion® High-fidelity DNA polymerase and approximately 10 ng of plasmid DNA. The reactions were run for 30 cycles: denaturation was at 95°C for 10 sec and elongation at 72°C for 15 sec. An initial denaturation at 95°C for 30 sec and a final extension at 72°C for 5 min.

The PCR products were resolved on a 1% agarose gel the detailed protocol is outlined in the appendix (A.1).

*PCR product purification*

*Fse* I is a low efficiency enzyme that is affected by salts or PCR components in the solution. The PCR products needed to be purified to remove and contaminants. The PCR product was purified using the Nucleospin® QuickPure PCR cleanup kit (Macherey-Nagal).

*Restriction Digestion of the fLA gene and pPinkα-HC plasmid*

To clone the gene of interest into the recipient plasmid, compatible ends are required. The PCR product was digested with *Fse* I to create a 3' sticky end. The pPinkα-HC plasmid was double digested with *Stu* I and *Fse* I to create both a 5' blunt end and a 3' sticky end, respectively. The restriction digestion was setup according to the table 3.1.

**Table 3.1:** Protocol for restriction digestion of the gene insert and pPinkα-HC plasmid

| **Restriction digest for PCR product** | | |
|---|---|---|
| Component | Volume (µl) | Final Concentration |
| Gene insert (103 ng /µl) | 10 µl | ~1 µg |
| 10 X Restriction Enzyme Buffer * | 1 µl | 1X |
| *Fse* I * | 1 µl | 10 units/µl |
| Sterile nuclease free water | to 10 µl | |
| **Restriction digest for pPinkα-HC plasmid** | | |
| Component | Volume (µl) | Final Concentration |
| Plasmid DNA (424 ng/µl) | 2.5 µl | ~1 µg |
| 10 X Restriction Enzyme Buffer * | 2.5 µl | 1X |
| *Stu* I * | 1 µl | 10 units/µl |
| *Fse* I * | 1 µl | 2 units/µl |
| Sterile nuclease free water | to 25 µl | |

*Cutsmart™ NEBuffer and Cutsmart™ enzymes (New England Biolabs)

To confirm each restriction enzyme was working efficiently, the pPinkα-HC plasmid was digested with *Stu* I and *Fse* I independently, as the size of the excised fragment between the two sites is too small to visualise on an agarose gel.

Each reaction was incubated at 37°C for 15 min, followed by deactivation at 65°C for 20 min. Following the digestion, the plasmid was dephosphorylated with 2 Units of Shrimp Alkaline Phosphatase (rSAP) in 1X rSAP buffer (New England Biolabs) to prevent self-ligation. The reaction was incubated at 37°C for 1 hr followed by heat deactivation at 65°C for 15 min.

The digested PCR product was phosphorylated with T4 Polynucleotide Kinase (New England Biolabs) in 1X T4 DNA ligase buffer a. This was incubated at 37°C for 1 hr followed by heat deactivation at 65°C for 20 min.

The digested plasmid including the uncut pPinkα-HC control plasmid, were resolved on a 0.8% agarose gel, the detailed protocol is outlined in the appendix. The digested plasmids were gel purified using the Zymoclean$^{TM}$ Gel DNA Recovery Kit (Zymo Reserach) to ensure the *Stu* I enzyme was removed. The digested PCR products were not resolved on an agarose gel as the size of the excised fragment was too small to visualise on the gel (6 bp).

*Ligation*

To clone the fluorinase gene into the pPinkα-HC plasmid a ligation reaction was setup. Based on the size of the insert (0.9 kB) and the plasmid (7.96 kB), the ratio of insert to vector was calculated according to: ng insert= (ng vector kbp insert)/ kbp vector. The reaction setup for the ligation reaction is outlines in table 3.2.

**Table 3.2:** Reaction setup for the ligation of the digested DNA fragments, *fLA* gene and pPinkα-HC plasmid.

| **Component** | **Volume (µl)** |
|---|---|
| 5X Ligase buffer * | 2 µl |
| T4 DNA Ligase * | 0.5 µl |
| Digested pPinkα-HC plasmid | 10 ng |
| Digested PCR product | 1.3 ng |
| Sterile nuclease free water | to 10 µl |

*Ligase buffer and T4 DNA Ligase (New England Biolabs)

Two reactions were set up, one was incubated at 25 °C for 2 hr and the other reaction was incubated at 16 °C overnight. Each ligation was used to transform chemically competent *E. coliI* JM109 cells. The protocol used is as previously described using the heat shock method. Cells were plated on LB-amp plates and incubated at 37 °C overnight.

*3.3.2 Screening for positive clones and sequencing*

Colonies from transformation plates were screened for the pPinkα-HC-*fLA* construct. Template DNA was prepared according to the method used in section 3.2.1. The Phos-FP and *Fse*I-RP gene specific primers were used to amplify the *fLA* gene (900 bp). The PCR was performed using the two step protocol outlined in section 3.3.1 and resolved on 1 % agarose gel of which the detailed protocol is outlined in the appendix (A.1).

Positive colonies containing the pPinkα-HC-*fLA* construct were grown in LB media supplemented with 100 µg/µl ampicillin. The plasmid DNA was isolated using the Zyppy™ Plasmid Miniprep Kit (Zymo Research) according to the manufacturer's instructions. Plasmid DNA was sequenced to confirm the sequence and ensure the gene was cloned in frame with the α-MF signal, pPink α–HC-*fLA* constructs were sequenced by Inquaba™ Biotec (Pretoria, South Africa). Plasmid specific primers (*AOX1* and *CYC1*) corresponding to the promoter and termination region were used. To ensure good coverage an additional primer complementary to the *fLA* gene was used (*fLA* internal). A single positive clone was selected and stocks of the plasmid DNA was prepared according to the protocol outlined in the appendix (A.3).

## 3.4 Transformation of PichiaPink™ strains with linearized pPinkα-HC- *fLA*

*3.4.1 Preparation of transforming DNA*

Linearized plasmid DNA is required for the transformation of *P. pastoris*; this is for efficient integration of the gene of interest into the host genome. The pPinkα-HC plasmid was linearized in the *TRP*2 region with *Bcu* I (*Spe* I). Forty micrograms of both the recombinant pPinkα-HC- *fLA* and pPinkα-HC parent plasmid were linearized. The protocol for the restriction digestion is outlined in table 3.3.

**Table 3.3:** Restriction digestion protocol for linearization of the recombinant and parent pPinkα-HC plasmid

| **Restriction digest for pPinkα-HC plasmid** | | | | | |
|---|---|---|---|---|---|
| **Component** | | | **Volume (µl)** | | **Final Concentration** |
| Plasmid DNA | 424 ng/µl [1] | 170 ng/µl [2] | 100 µl [1] | 250 µl [2] | ~42.5 µg |
| 10 X Buffer Tango | | | 20 µl | 35 µl | 1X |
| *Bcu*I (*Spe*I) * | | | 42.5 µl | 42.5 µl | 10 units/µl |
| Sterile nuclease free water | | | to 200 µl | to 350 µl | |

[1] Parent plasmid (pPinkα-HC)

[2] Recombinant construct (pPinkα-HC plasmid-*fLA*)

*Buffer Tango and *Bcu*I (*Spe*I) (Thermoscientific)

Each of the reactions was incubated at 37 ° for 16 hr. A small aliquot of each reaction was then analysed using horizontal gel electrophoresis, to determine if the plasmid DNA was completely linearized before deactivating the enzyme. The uncut and cut plasmid were resolved on a 0.8% agarose gel (A.1).

The linearized plasmid DNA was purified using standard protocol according to Cocolin *et al.,* (2000) with some modifications. Digested plasmid DNA was precipitated with 2.5 volumes of 100% ethanol and 1/10 volume of 3 M sodium acetate, followed by incubation at -20 °C for 20 min. The DNA was pelleted by centrifugation at 16 000 *xg* for 10 min at 4°C and washed with 80 % ethanol. The pellet was air-dried and resuspended in 40 µl sterile $dH_2O$.

*3.4.2 Preparation of electrocompetent cells*

To prepare electrocompetent PichiaPink™ strains for electroporation. Each of the PichiaPink™ strains were streaked on YPD agar plates and grown at 25 °C for 3-5 days, until distinct colonies formed. A single colony of each strain was used to inoculate 10 ml of YPD media and incubated shaking at 28 °C at 300 rpm for 1-2 days. The starter culture was used to inoculate 100 ml of YPD media to a final $OD_{600}$ of 0.2, and incubated shaking at 28 °C at 300 rpm for 1-2 days. Each strain was grown till the cell density reached an $OD_{600}$ of 1.3-1.6. Cells were pelleted by centrifugation at 1 500 x g for 5 min at 4 °C and washed with 250 ml sterile ice cold $dH_2O$. Cells were pelleted as in the previous step and washed in 50 ml sterile ice cold $dH_2O$. This was repeated with 10 ml of ice cold 1 M sorbitol and finally resuspended in 300 µl of ice cold 1 M sorbitol. Cells were kept on ice and used fresh.

### *3.4.3 Transformation of* PichiaPink$^{TM}$ *strains by electroporation and selection of transformants*

Each strain of PichiaPink$^{TM}$ was transformed with the linearized expression pPinkα-HC *-fLA* construct, linearized parent plasmid and with no template as a negative control. For each transformation 80 µl of the electrocompetent cells were mixed with 10 µl of plasmid DNA (for the negative control 10 µl of sterile $dH_2O$ was added). The cell mixture was transferred to a cold 0.2 cm electroporation cuvette (VWR) and incubated on ice for 5 min. Optimisation of electroporation parameters was required. The ECM® 830 electroporation system (BTX Harvard Apparatus) was used for pulsing the cells.

**Table 3.4:** Electroporation parameters used for transformation with the ECM® 830 electroporation system

| Parameter | Procedure 1 | Procedure 2 | Procedure 3 |
|---|---|---|---|
| **No. of Pulses** | 5 | 5 | 3 |
| **Interval** | 100 ms | 100 ms | 150 ms |
| **Pulse length** | 150 µs | 200 µs | 100 µs |
| **Voltage (HV)** | 2 000 V | 2 000 V | 1 500 V |
| **Polarity** | Unipolar | Unipolar | Unipolar |

Following electroporation, cells were incubated in YPDS recovery solution for 4 hr at 28 °C (optimised to 6 and 8 hr on the 2$^{nd}$ and 3$^{rd}$ attempt) and plated on *Pichia* adenine dropout (PAD) selection plates.

Following these attempts the Gene Pulser Xcell$^{TM}$ electroporation system (Biorad), with a built in protocol for *P. pastoris* was used. The conditions were C= 25 uF; PC= 200 ohm; V= 2 000V with an expected time constant of ~5 millisec. Immediately after pulsing 1 ml of YPDS recovery media was added to the transformation reaction. For each strain half of the transformation mixture was incubated at 28 °C and half at 35 °C, after 6 and 12 hr 250 µl of the cells were removed and spread on PAD selection plates. The plates were incubated at 25 °C for 5-10 days until distinct colonies formed.

### 3.5 PCR analysis of PichiaPink$^{Tm}$ integrants

The first approach for screening PichiaPink$^{Tm}$ integrants included isolating genomic DNA which was used as a template in a PCR with gene specific primers. A ZR Fungal/Bacterial DNA MiniPrep™ kit (Zymo Research) was used according to the manufacturer's instructions. The genomic DNA was used as a template in a PCR with the *AOX1* and *CYC1* primers.

A second approach was to use direct PCR. The colonies from each transformation plate were transferred to master plates using colony grids. Twenty white colonies for each strain were selected for screening.

A 96-well PCR plate a small amount of cells were scraped from the master plate and resuspended in 75 µl of sterile $dH_2O$. The cell suspension was mixed by pipetting up and down several times. Cells were lysed by five rounds of heat treatment in the microwave followed by freezing. The cell suspensions were microwaved on high for 3.5, 2, 1.5, 1 and 0.5 min, between each heating step the cell suspensions were vortexed briefly. The cells were freeze-thawed by freezing at -80 °C for 10 min, followed by heating at 95 °C for 5 min. To pellet the cells debris the suspension was centrifuged at 1 000 *x*g for 15 min using a 96-well plate swing-out centrifuge. The supernatant served as the template in the following PCR reaction.

The *AOX*1 and *CYC*1 primer set were used to amplify the region between the promoter and termination region. KAPA2G Robust HotStart ReadyMix PCR kit (KAPABiosystems) was used to screen for positive clones. This master mix is suited for amplification from crude samples and is convenient when screening large numbers of colonies. The PCR was performed in a final volume of 25 µl containing 1X KAPA2G Robust HotStart ReadyMix (2X), 5% DMSO, 200 µM dNTPs, 1.5 mM $MgCl_2$ (additional), 0.5 µM of each primer and 7.5 µl supernatant. The reactions were run for 30 cycles: denaturation was at 95 °C for 15 sec, annealing at 60 °C for 15 sec and elongation at 72 °C for 30 sec. An initial denaturation at 95 °C for 3 min and a final extension at 72 °C for 1.5 min. The PCR products were resolved on a 1% agarose the detailed protocol is outlined in the appendix.

The PCR products were used in a nested PCR with the Phos-FP and *Fse*I-RP gene specific primers to confirm integration of the *fLA* gene. PCR products were sent to Inqaba Biotech for sequencing with the *AOX1, CYC1* and *fLA* internal primers.

### 3.6 Expression of recombinant PichiaPink$^{Tm}$ strains and optimisation

Recombination may occur in many different ways, each of which affect protein expression; therefore four positive colonies for each strain were selected for expression evaluation. The following controls were used, the parent PichiaPink$^{Tm}$ strains (untransformed), PichiaPink$^{Tm}$ strains transformed with the parent plasmid and recombinant PichiaPink$^{Tm}$ strains uninduced (grown in BMGY media only).

Positive colonies of each strain were selected from fresh plates. A single colony was then used to inoculate 10 ml BMGY medium and placed in 125 ml flasks. The cells were grown with shaking at 300 rpm at 28 °C for 1 day. The cells were pelleted at 1 500 x g for 5 min and the BMGY media was removed. The cell pellet was then resuspended in 1 ml BMMY media to induce expression. Cells were grown in 50 ml conical tubes with shaking at 300 rpm at 28 °C for 4 days. One hundred microliters of each sample was taken at 0, 12, 24, 48, 72 and 96 hr, the volume was replaced with 100 µl of 40 % methanol each time. Samples were stored by pelleting the cells at 1 500 x g for 10 min. For SDS-PAGE analysis the supernatant and pellet were stored separately. The samples were snap frozen in liquid nitrogen and stored at -80 °C.

In an attempt to optimise expression, the above expression protocol was repeated with the following modifications: the initial growth step was increased to 2 days to enhance cell densities; cells were grown in 125 ml baffled flasks in order to promote aeration; induced cells were grown in a final volume of 5 ml and the 1 ml of sample was removed replaced with 1 ml of 50% methanol. Cells were harvested by centrifugation and the supernatant and pellet were stored separately (After 96 hr of induction the total 5 ml sample was centrifuged).

### 3.7 SDS-PAGE analyses of secreted protein expression

Protein expression was analysed using the protocol for tricine-SDS-PAGE gels according to Schägger, (2006). The detailed protocol for gel and buffer preparation is outlined in the appendix (A.1). For analyses of secreted protein expression for each strain, the crude samples were used, 5 µl of the reducing sample buffer was mixed with 15 µl of the supernatant sample (media). The samples were then incubated at 37 °C for 30 min. For analyses of the secreted protein fraction, 16% separating gel with a 4% stacking gel was used to separate proteins. 20 µl of the protein sample was loaded and for the marker 10 µl of Unstained Protein Marker, Broad Range (2-12 kDA) (New England Biolabs) was loaded. The protein marker was boiled at 100 °C for 5 min before loading. Gels were stained using the

coomassie staining method. The gels were photographed with the ChemiDoc$^{TM}$MP Imaging System (Biorad).

### 3.8 Comparison of cell disruption methods for protein extraction

Four disintegration methods were tested in an attempt to find the most suitable method to extract protein for SDS-PAGE analysis and activity assays. Methods were adapted from Klimek-Ochab *et al.,* (2011) with some modifications.

The 96 hr sample from the expression trails had the largest pellet and was selected for disruption, for consistency one sample for strain 4 was used for all four disruption methods. For each sample the pellets were weighed and resuspended in breaking buffer where 1 ml of breaking buffer per 1 g of pellet was used. Before and after disruption 100 µl of the cell suspension was removed to measure breaking efficiency. Following disruption the cell debris were pelleted by centrifugation at 1 500 *xg* for 15 min at 4 °C. The supernatant and pellet were separated and analysed by SDS-PAGE.

#### *3.8.1 Bead Milling*

Bead milling involves glass or zirconium beads circulating in cell suspensions leading to disintegration of cell walls. The cells suspension was mixed with an equal volume of zirconium beads (0.1 mm and 0.5 mm). Cells were vortexed with a $Zx^3$ vortex (Velp Scientifica) at 40 Hz continuously for a total of 30 sec followed by cooling on ice for 30 sec. This was repeated for a total of 8 cycles.

#### *3.8.2 Osmotic Shock*

For cell disruption by osmotic shock the cell pellet was resuspended in a hypertonic solution instead of a breaking buffer. For every 1 g of cell pellet, 1 ml of hypertonic solution (50 mM Tris-HCl buffer, pH 7.5 containing 10 mM EDTA and 30% (w/v) sucrose) was added. The cell suspension was shaken at 4 °C for 20 min. The cells were pelleted by centrifugation at 10 000 *x*g for 15 min at 4 °C using a MiniSpin Plus benchtop centrifuge (Eppendorf). The supernatant was discarded and the pellet was resuspended in an equal volume of 10 mM $MgCl_2$ for 20 min.

#### *3.8.3 Sonication*

Sonication is a mechanical method that shears cell walls with high frequency ultrasonic waves. The cells suspension was sonicated using the microson$^{TM}$ XL2000 ultrasonic cell

disruptor (Misonix, *inc.*). Cells were sonicated for bursts of 30 sec at 20 kHz on ice, followed by resting on ice for 30 sec. This was repeated for 8 cycles.

*3.8.4 Freeze Crushing*

Liquid nitrogen was poured directly into 2 ml tubes containing the cells and micropestles were used to crush the cells. Cells were milled for approximately 30 sec and additional liquid nitrogen was added, this was repeated 8 times. Conical tubes were used at first and it was found that these did not facilitate the adequate crushing of cells, thereafter round bottom tubes were used which allowed sufficient access to the pellet.

The efficiency of each method was determined by measuring two parameters: calculating the breaking efficiency and measuring the concentration of the soluble proteins in the cell-free extract. Intact cells were counted before and after disruption using a Neubauer chamber hemocytometer (Celeromics). The cells were diluted 10X and a small drop was added to the chamber. The grid was examined under the 40X objective using an Olympus CX22LED light microscope (Olympus). The cells in ten 0.04 $mm^2$ squares were counted and the concentration calculated (the formula is given in appendix A.4). The breaking efficiency for each method was calculated according to Wang *et al.,* (2015), which determines the percentage of disrupted cells (the formula is given in appendix A.5). The protein concentration from the cell free extract was then measured. The cell free extract and the cell pellet were analysed on 16% SDS gels (as described in 3.7) with samples diluted 5X.

**3.9 Analysis of intracellular protein expression by SDS-PAGE**

To determine if the recombinant protein was not secreted, the intracellular fraction was analysed using SDS-PAGE gels. The total protein was extracted from the cell pellet for each strain using the freeze crushing method described above (section 3.8.4). After the cells were disrupted the cell debris were pelleted and the supernatant then analysed on SDS-PAGE gels. The crude extract was diluted 5X and 15 µl was mixed with 5 µl of the reducing sample buffer. The protocol outlined in section 3.7 was followed.

**3.10 Fluorinase enzyme activity assay using HPLC**

A fluorinase enzyme activity assay was done to determine if the expressed protein was biologically active. The crude protein extract was tested as the fluorinase was not purified. The assay is based on the reaction of the fluorinase enzyme, which converts SAM and fluoride ion to 5-FDA and methionine. HPLC is used to identify 5-FDA, the product of the reaction. The method was followed according to (Schaffrath *et al.*, 2003).

All assays and controls were separated on an Ascentis® Express HILIC HPLC 2.7µm column. A gradient mobile phase with an initial mobile phase of 95:5 and final mobile phase of 80:20, (50 mM Sodium phosphate buffer (pH7.4): Acetronitrile) was used. The flow rate was set to 1ml/min for a maximum of 10 min.

For a reference 5-FDA the product of the reaction and SAM the reactant were used as standards and subjected to HPLC in order to obtain a reference. The 5-FDA is not commercially available and was synthesised by Dr Allan Prior (School of Chemistry, University of the Witswatersrand). S-adenosyl methionine dichloride was purchased from Sigma Aldrich. In order to identify the retention time of 5-FDA and SAM, a standard solution containing 5 mM 5-FDA and 0.1 mM SAM in 50 mM sodium phosphate buffer was eluted through the column. A second standard solution containing 2mM 5-FDA in 50 mM sodium phosphate buffer was eluted through the column. To determine background peaks for the intracellular fraction, the breaking buffer used in protein extraction was eluted through the column.

Once reference peaks were identified, enzyme activity for the intracellular and extracellular protein fraction was tested. The assay was performed in 200 µl 50 mM sodium phosphate buffer (pH 7.4) containing 20 µl crude extract, 50 mM NaF and 0.1 mM SAM. The reaction was incubated at 30°C for 30 min and then heated at 95 °C for 10 min to precipitate the protein. The protein was pelleted by centrifuging the mixture at 10 000 *x*g for 5 min using MiniSpin Plus benchtop centrifuge (Eppendorf). Twenty microliters of the supernatant was separated on the column.

# 4. Results

## 4.1 Confirmation of the *fLA* gene in pUC-57

The *fLA* gene was maintained in a pUC-57 vector, thus facilitating the use of this construct in the transformation of *E. coli* JM109 cells. To confirm the presence of the *fLA* gene, positive clones were screened using colony PCR with the SC forward and SC reverse gene specific primers.

The SC primers amplify approximately 814 bp of the *fLA* gene and a band corresponding to this size was resolved on a 1.0% agarose gel and can be seen in figure 4.1 lanes 2 and 6. This indicates a positive colony was transformed with the recombinant pUC57-*flA*. Negative colonies are seen in lane 3-5 (figure 4.1). A single positive colony was selected and grown in LB-amp media the plasmid was isolated and this served as the template in the PCR based cloning steps.

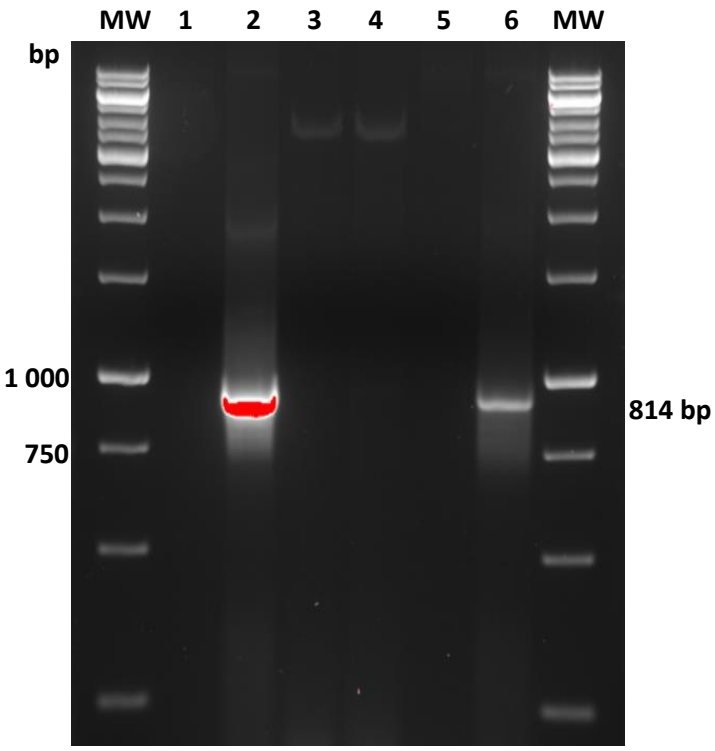



**Figure 4.1**: Colony PCR screening for the pUC-57-*fLA* construct. PCR products separated on a 1% agarose gel. Lane 1- negative control (no template), Lane 2 and 6 Approximately 814 bp of the *fLA* gene was amplified, indicating positive colonies transformed with the construct Lane 3-5 negative colonies, MW- GeneRuler 1kb DNA ladder (Thermo Scientific). Red lines indicate saturated pixels where DNA is concentrated.

## 4.2 Construction of recombinant pPinkα-HC-*fLA* vectors

### *4.2.1 PCR Based Cloning*

PCR based cloning was used to copy the *fLA* gene from the pUC-57-*fLA* construct and to then add restriction sites to the gene of interest, in order to clone this into the pPink α–HC plasmid. A phosphorylated forward primer and a reverse primer with a *Fse* I site were used to create a 5' blunt end and a 3' *Fse* I site, flanking the *fLA* gene.

The FP-Phos and RP-*Fse* I primers were used to amplify the *fLA* gene from the pUC-57-*fLA* construct in a two-step PCR. The PCR parameters had to be optimised as the primers had a high GC content, were long and had high TM values. This caused non-specific binding leading to non-specific products (data not shown). Using Phusion high fidelity DNA polymerase in a two-step PCR, ~900 bp product was amplified from the vector and resolved on a 1.0% agarose gel, figure 4.2A shows the amplicon corresponding to the size of the *fLA* gene. The PCR products were purified and digested with *Fse* I, followed by phosphorylation with T4 polynucleotide kinase.

To create compatible ends on the pPink α–HC plasmid to that of the digested gene, the plasmid was double digested with *Stu* I and *Fse* I. The pPink α–HC plasmid was also digested with each enzyme individually to confirm that the restriction enzymes were working efficiently, particularly as the size of the excised fragment between the two sites is too small to visualise on an agarose gel. Circularised and linearized plasmids differ in conformation and therefore resolve differently on agarose gels. Figure 4.2B shows the 0.8% agarose gel with the resolved circularised and linearized plasmids, the size of the pPinkα-HC plasmid is 7.9 kb. Figure 4.2B shows the plasmid was linearized correctly when compared to the uncut plasmid (lane 1 and 2 respectively). To ensure both enzymes were cutting efficiently the plasmid was digested independently with each enzyme. The plasmid was linearized correctly by each enzyme; lanes 3 and 4 indicate that *Fse* I and *Stu* I were both working efficiently.

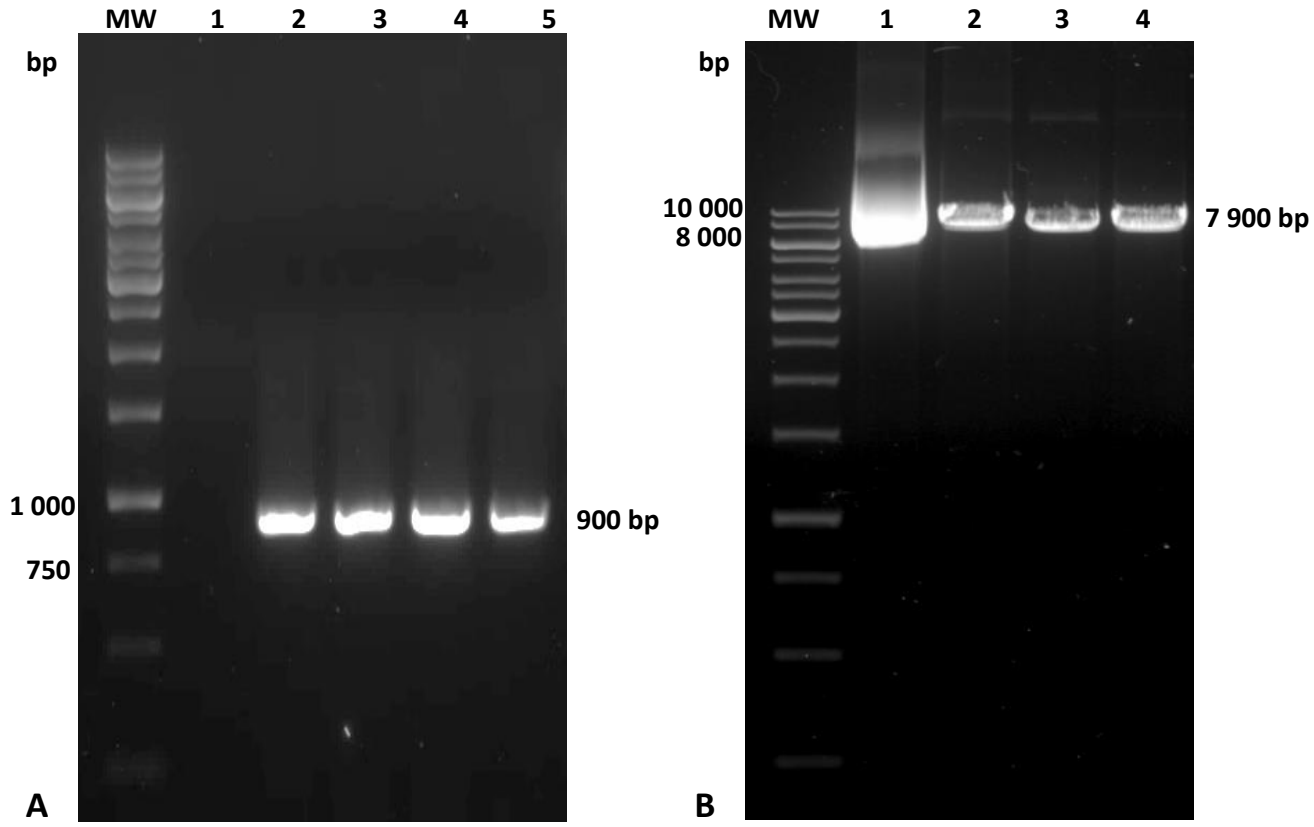



**Figure 4.2:** PCR based cloning for the construction of pPink α–HC-*fLA*. **A**- PCR amplification of the *fLA* gene from the pUC-57-*fLA* with FP-Phos and RP-*Fse* I primers. PCR products were separated on a 1.0% agarose gel. Lane 1- negative control (no template), Lane 2-5 - amplification of the 906 bp *fLA* gene. **B**- Digestion of the pPink α–HC plasmid (7.9 kb), digested fragments resolved on a 0.8% agarose gel. Lane 1- Uncut plasmid, Lane 2- Double digested with *Stu* I and *Fse* I, Lane 3- digested with *Fse* I, Lane 4- digested with *Stu* I, MW- GeneRuler 1kb DNA ladder (Thermo Scientific).

The digested plasmid and gene insert were ligated using T4 DNA ligase. *E. coli* JM109 cells were transformed with the ligation reaction and plated on LB-amp agar for selection. Positive clones were screened for in a colony PCR using KAPA2G Robust HotStart ReadyMix PCR kit (KAPABiosystems) with gene specific primers (FP-Phos and RP-*Fse* I). A ~900 bp product was amplified from eight colonies. Data show that these colonies were successfully transformed with the expression construct (Figure 4.3).

Plasmid DNA from positive colonies were then isolated and sequenced to confirm the DNA sequence and ensure the gene was cloned in frame with the α-MF signal, pPink α–HC-*fLA* constructs were sequenced by Inquaba Biotech. Plasmid specific primers (*AOX1* and *CYC1*) corresponding to the promoter and termination region were used in this procedure. To ensure good coverage an additional primer complementary to the *fLA* gene was used. Sequencing results showed that the gene of interest was successfully cloned into the recipient plasmid and that the gene was cloned in the correct position upstream of the secretion signal sequence. The sequence was confirmed by aligning it with the original sequence using the bioedit sequence alignment editor tool. A single positive clone was selected and plasmid DNA was isolated.

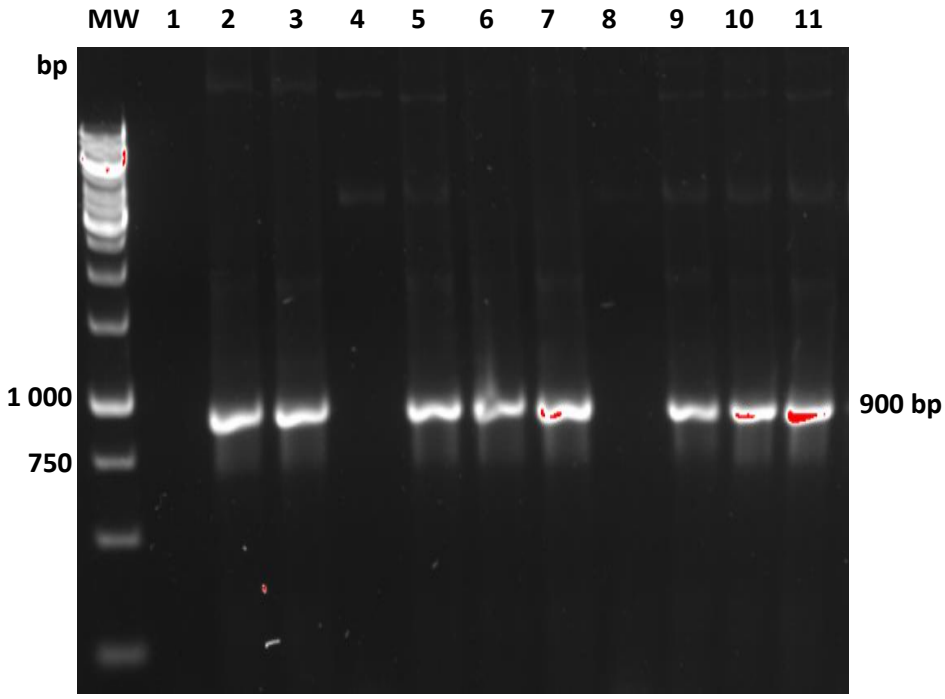



**Figure 4.3:** Colony PCR to screen for the pPink α–HC-*fLA* construct using gene specific primers. PCR products were resolved a 1% agarose gel. Lane 1- negative control (no template), Lane 2-11 a 900 bp product was amplified for positive colonies which corresponds to the size of the *fLA* gene. No products were seen for negative colonies Lane 4 and 8). MW-GeneRuler 1kb DNA ladder (Thermo Scientific). Red lines indicate saturated pixels where DNA is concentrated.

## 4.3 Transformation of PichiaPink™ strains with linearized pPinkα-HC- *fLA*

### *4.3.1 Preparation of transforming DNA*

In order to integrate the foreign gene into the host genome, it is essential to linearize plasmid DNA for homologous recombination to take place. Both the parent plasmid and the expression construct were linearized at the *TRP2* gene with *Spe* I and resolved on a 0.8% agarose gel. Figure 4.4A shows the parent plasmid (pPinkα-HC) resolved at 7.9 kb which corresponds to the size of the parent plasmid. The expression construct (pPinkα-HC- *fLA)* resolved at 8.8 kb which corresponds to the size of the gene and plasmid (Figure 4.4B). Both plasmids were successfully restricted when compared to the uncut plasmid.

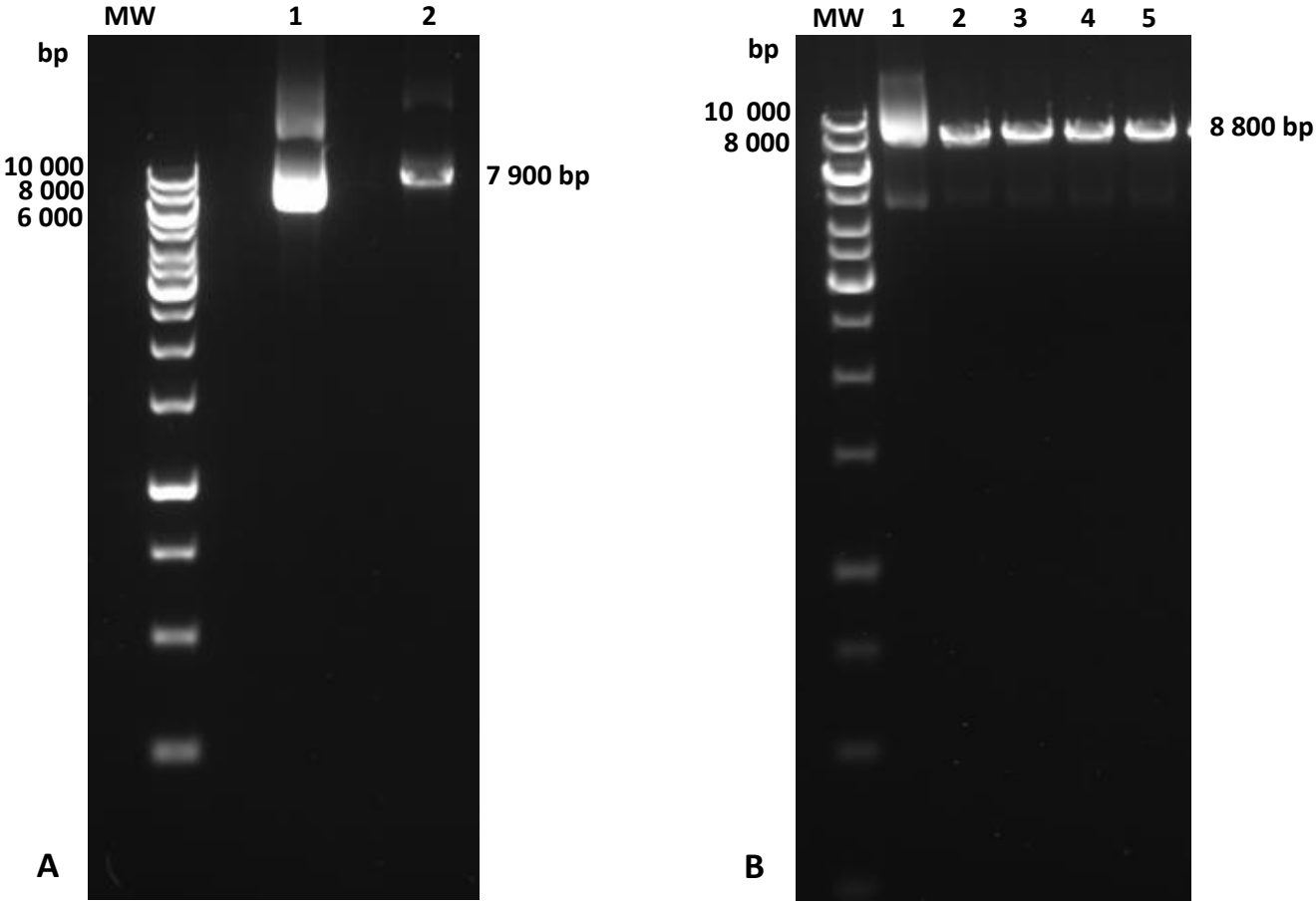



**Figure 4.4:** Linearization of the pPink α–HC parent plasmid (7.9 kb) and pPink α–HC-*fLA* construct (8.8 kb) at the *TRP2* region for integration. Plasmids digested with *Spe* I and resolved on a 0.8% agarose gel. **A**. Lane 1- Uncut parent plasmid, Lane 2- linearized parent plasmid **B.** Lane 1- Uncut expression construct, Lane 2-5- expression linearized construct. MW- GeneRuler 1kb DNA ladder (Thermo Scientific).

### *4.3.2 Transformation of* Pichia*Pink™ strains by electroporation and selection of transformants*

Each of the four PichiaPink™ strains were transformed with the linearized parent plasmid and the expression construct by electroporation. The cells were pulsed with the ECM® 830 electroporation system (BTX Harvard Apparatus) in an attempt to optimise the electroporation step, and three different sets of parameters were used. The voltage, pulse length, pulse number and interval were varied. Following electroporation cells were incubated in YPDS recovery solution for 4 hr at 28 °C (changed to 6 and 8 hr on the 2nd and 3rd attempt) and plated on PAD selection plates (not all data shown). Figure 4.5A shows representative results from the transformation with the ECM® 830 electroporation system. For each attempt at transformation no distinct colonies formed on the selection plates after 10 days. A film of pink cells can be seen in figure 4.5A which is indicative of untransformed cells as these are incapable of growing on agar lacking adenine.

The Gene Pulser Xcell™ electroporation system (Biorad) was used to pulse cells following previous attempts. The cells were pulsed at C= 25 uF; PC= 200 ohm; V= 2 000V with an expected time constant of ~5 millisecond. For each transformation the correct time constant was reached between 4.9 and 5. Distinct colonies formed on PAD plates after incubation at 25 °C for 8-10 days (Figure 4.5B-I).

Figure 4B-I shows transformation plates (not all data shown); white colonies represent cells that were transformed with the pPinkα-HC plasmid that contains the *ADE2* gene and therefore is capable of growing on media lacking adenine. Dark pink colonies represent negative clones (figure 4B-C), these do not grow to larger sizes as the ADE2 gene is not integrated and therefore colonies cannot grow on media lacking adenine. Light pink colonies can be seen in figure 4D, these indicate clones that have a low copy of the integrated plasmid, as the colour of the colony can be related to the copy number of the gene insert.

The transformation reactions that were incubated for longer periods at 12 hr (Figure 4.5E-F) showed a greater number of transformants compared to those incubated for at only 6 hr (figure 4.5G-H). After 14-16 days some colonies that were initially white started showing a light pink phenotype and smaller satellite colonies began to grow on the plate (Figure 4.5I).

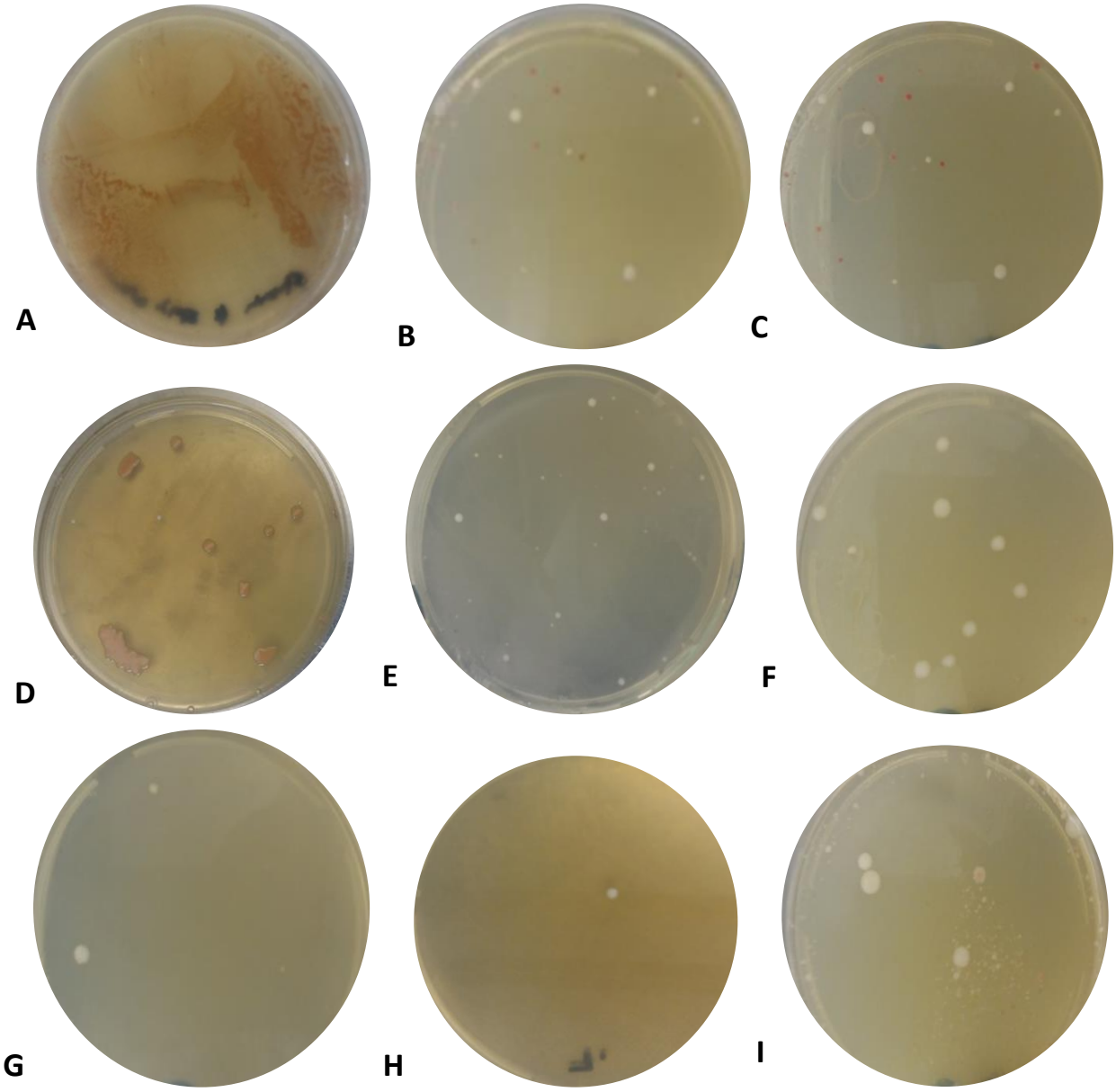



**Figure 4.5:** Following transformation of PichiaPink™ strains with the linearized expression construct (pPink α–HC-*fLA*), cells were plated on PAD selection agar. **A:** Transformation using the ECM® 830 electroporation system. No distinct colonies formed on plates after 10 days, **B-I**: After transformation with Biorad Genepulser Xcell distinct colonies formed on PAD plates after 8-10 days. White colonies represent cells that were transformed with the pPinkα-HC plasmid. **B and C:** Dark pink colonies represent negative clones, **D:** Light pink colonies represent clones with a low copy of the integrated plasmid, **E-F:** transformation reactions recovered for 12 hr (at 28 °C and 37 °C respectively) prior to plating, **G and H:** transformation reactions recovered for 6 hr (at 28 °C and 37 °C respectively) prior to plating, **I:** After 14-16 days smaller satellite colonies appeared on PAD and some colonies that were initially white started to appear light pink.

### 4.4 PCR analysis of PichiaPink$^{Tm}$ integrants

Colony PCR was used to directly test PichiaPink$^{TM}$ clones for insertion of the gene of interest. For each strain, ~twenty white colonies were screened using the *AOX1* and *CYC1* primers. The parent plasmid (pPinkα-HC) was used as a negative control and the expression construct (pPinkα-HC-*fLA*) as the positive control. The PCR products from the colony PCR were resolved on a 1.0 % agarose gels. A prominent band corresponding to the promoter and transcription termination region was amplified from each clone (figure 4.6A). The size of the promoter and termination region of the parent plasmid was 426 bp; an amplicon corresponding to this size can be seen in lane 2 for the negative control (Figure 6A). A band corresponding to the size of the negative control was amplified for each colony which can be seen in lane 4-24 (Figure 4.6A). The size of the gene plus the promoter and transcription terminator region flanking the gene was 1 308 bp and a band corresponding to this size was amplified from the expression construct positive control seen in lane 3 (figure 4.6A). A faint band corresponding to the size of the positive control was also amplified for some of the colonies for each strain. To confirm if the faint band was in fact corresponding to the *fLA* gene, a nested PCR with gene specific primers was performed.

To confirm the presence of the *fLA* gene, the PCR products were used in a nested PCR with gene specific primers for colonies that showed a faint band corresponding to the positive control. A 900 bp product corresponding to the size of the *fLA* gene was amplified from the PCR products that showed a faint band from the first round of PCR (figure 4.6B).
The results show that the gene of interest was integrated into each of the PichiaPink$^{TM}$ strains. However the plasmid without the gene was also integrated into the host genome.
The colonies that showed integration of the *fLA* gene in the nested PCR were selected for expression. The PCR products were sequenced by Inqaba Biotec (Pretoria, South Africa) and the gene was confirmed.

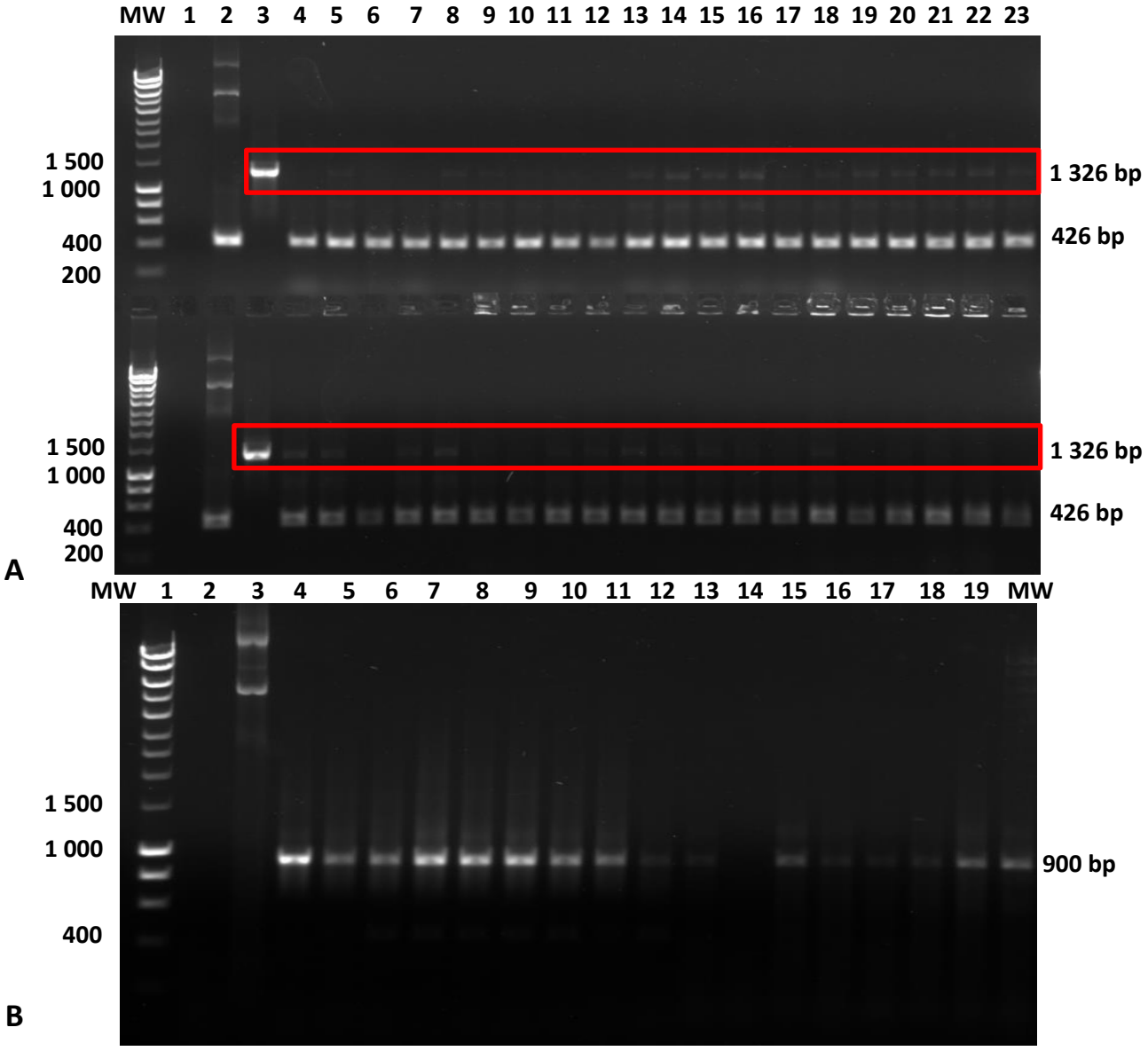



**Figure 4.6: A.** Colony PCR for direct testing of PichiaPink™ clones for insertion of the *fLA* gene with *AOX1* and *CYC1* primers. Lane 1- no template control (negative control), Lane 2- pPink α–HC parent plasmid used as a template (negative control), Lane 3- amplification of the *fLA* gene from the pPink α–HC-*fLA* construct (positive control), Lane 4-24 amplification of a prominent band corresponding to the negative control (426 bp) and a faint band corresponding to the positive control (1 326 bp) is outlined on the gel. **B.** To confirm integration of the *fLA* gene, PCR products that showed a faint band corresponding to the positive control were used in a nested PCR with gene specific primers. Lane 1- no template control (negative control), Lane 2- amplification of the promoter and transcription termination region (461 bp) from the pPink α–HC parent plasmid (negative control), Lane 4-19 a 905 bp amplicon corresponding to the size of the *fLA* gene was amplified from the PCR products indicating integration of the gene of interest. MW – HyperLadder™ 1kb DNA Marker (Bioline)

### 4.5 Expression and analysis of secreted protein expression by SDS-PAGE

Integration can occur in multiple ways which may affect expression of the protein, therefore four positive colonies for each strain were expressed. The following controls were used, the parent PichiaPink$^{Tm}$ strains (untransformed), PichiaPink$^{Tm}$ strains transformed with the parent plasmid and recombinant PichiaPink$^{Tm}$ strains uninduced (grown in BMGY media only).

Positive colonies of each strain were selected from fresh plates and grown in 5 ml BMGY medium for 1 day. The cells were pelleted and resuspended in 1 ml BMMY media to induce expression. Cells were subsequently grown in 50 ml conical tubes for 4 days. One hundred microliters of each sample was then taken at 0, 12, 24, 48, 72 and 96 hr, the volume was replaced with 100 µl of 40 % methanol each time. The supernatants of each sample was analysed on 16% SDS gels but no bands were seen (data not shown).

In an attempt to optimise expression, the above steps were repeated with the following modifications: the initial growth with BMGY media was increased to 2 days to reach higher cell densities in a final volume of 10 µl; cells were grown in 125 ml baffled flasks for all steps to enhance aeration; induced cells were grown in a final volume of 5 ml and; 1 ml of sample was removed and replaced with 1 ml of 50% methanol. To analyse the secreted protein fraction the supernatants of all sixteen samples at each time fraction were analysed on 16% SDS gels (not all data shown). Figure 4.7 shows representative data for the expression trials, no prominent bands were seen at 32 KDa, the expected size of the *fLA* protein. For strains 2-4 no recombinant protein was visible. Near the expected size of the protein (below 34.6 kDa), a few bands can be seen for strain 1 (figure 4.7A). As the bands are faint these results do not unequivocally support the correspondence between these bands and the fluorinase, or perhaps, there was simply limited expression of the enzyme.

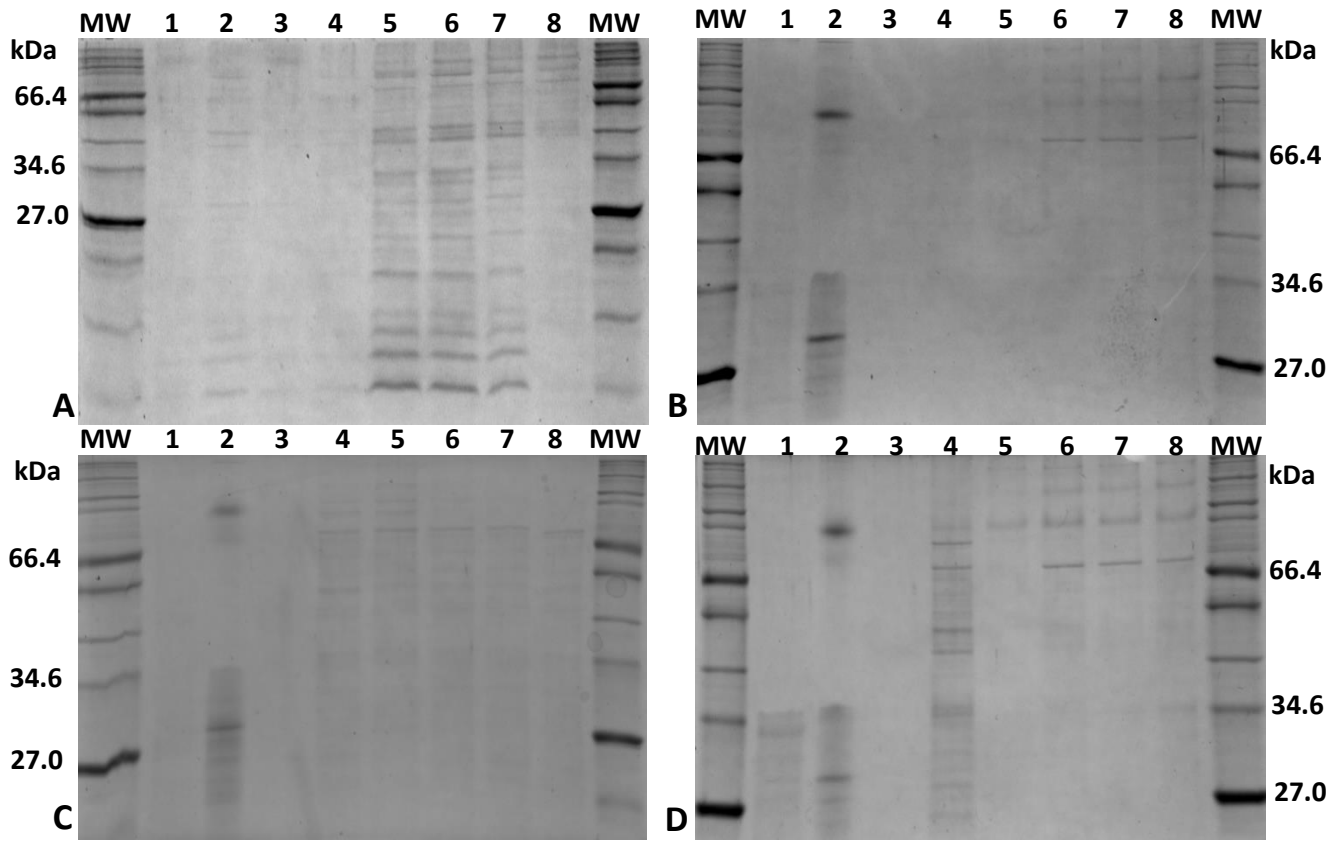



**Figure 4.7:** SDS-PAGE analysis of secreted protein expression. Expression induced with 0.5 % methanol and culture supernatant analysed on 16% SDS-PAGE gels. A- Strain 1, B- Strain 2, C- Strain 3 and D- Strain 4, Lane 1- Untransformed parent strain (strain1-4), Lane 2- Strain 1-4 transformed with parent plamid, Lane 3- Uninduced (Strain 1-4), Lane 4-8 Induced (Strain 1-4) at 0, 24, 48, 72 and 96 hr respectively, no distinct bands at 32 kDa (size of fluorinase). MW- Unstained protein marker broad range (2-212 kDa) (New England Biolabs).

### 4.6 Comparison of cell disruption methods for protein extraction

The most efficient method for protein extraction was established to ensure that the majority of the protein was being extracted for analysis by SDS-PAGE and activity assays. Four disruption methods were tested to determine the most efficient method for cell lysis and optimal protein extraction. Table 4.1 shows the results for each of the cell disruption methods. Mechanical methods yielded the highest breaking efficiency whilst the osmotic shock chemical method was unsuccessful in disintegrating the cell wall. The sonication method had the highest breaking efficiency at 71 %, whilst the freeze crushing method the highest concentration of released protein at 17.7 µg/µl. For each disruption method the supernatant (containing released protein) and the pellet (remaining undisrupted cells) were analysed on 16% SDS-PAGE gels (figure 4.8). For the bead milling (lane 1 and 2) and freeze crushing (lane 7 and 8) method, a higher concentration of the protein was evident in the supernatant compared to the pellet, thus indicating the cells were effectively lysed (table 4.1). However with the sonication method (lane 7 and 8) it was seen that the protein content was higher in the pellet compared to the supernatant, this explains why the measured protein concentration was lower than that measured for the bead milling method, with the lower breaking efficiency. The protein yield for the osmotic shock method was negligible, as this method failed to lyse the cells.

### 4.7 Protein extraction and Analysis of intracellular protein expression by SDS-PAGE

Upon analyses of the secreted protein fraction it was evident that there was either no recombinant protein being secreted into the culture supernatant, or alternatively expression was low and the protein could not be detected by SDS-PAGE. The intracellular fraction was analysed by SDS-PAGE to determine if the protein did not secrete via the α-MF secretion signal. Figure 4.9 shows the SDS-PAGE gels for analysis of the intracellular protein fraction of strain 1-4 induced at 72 and 96 hr. Two prominent bands were seen on the gel for strain 2-3. These bands correspond to the fluorinase protein (32.2 kDA) with the unprocessed α-MF secretion signal.

For strain 3 and 4, at 72 hr and 96 hr induction, bands are prominent (figure 4.9). For strain 2 the band representing expressed intracellular fluorinase protein is less prominent for the 96 hr induction. This lack of prominence, may be related to proteinase activity. For strain 1 the bands are faint for the 72 hr sample and completely absent for the 96 hr sample. Strain 1 is a wild-type strain, therefore proteinases would be present in higher concentrations and lead to the degradation of the foreign protein.

**Table 4.1:** Comparison of four cell disruption methods**,** showing the breaking efficiency and concentration of the soluble protein after disruption

| **Disruption method** | **Breaking efficiency (%)** | **Protein Concentration (**µg/µl) |
|---|---|---|
| Bead Milling | 64.6 % | 15 µg/µl |
| Osmotic Shock | 8 % | 0.2 µg/µl |
| Sonication | 71 % | 13.8 µg/µl |
| Freeze Crushing | 66.8 % | 17.7 µg/µl |

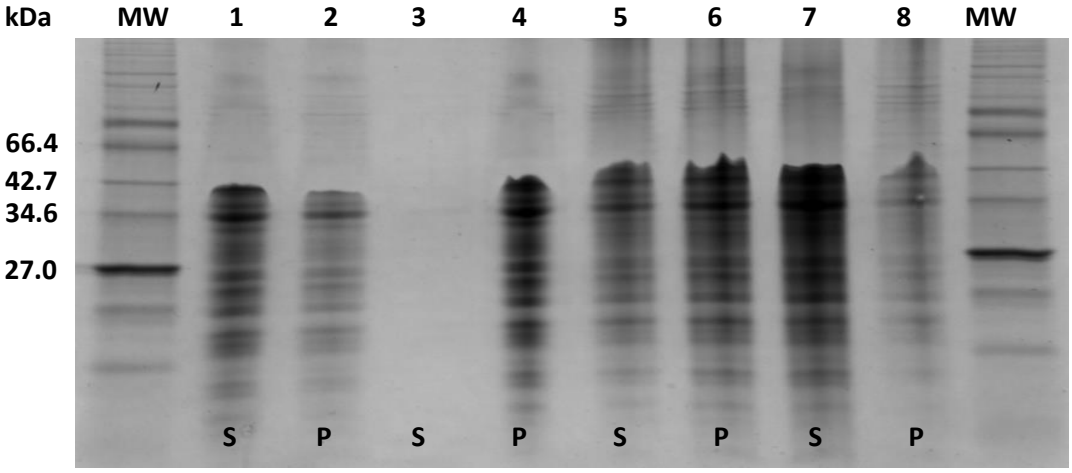



**Figure 4.8:** Comparison of the total protein released after cell disruption- the S- soluble protein and P-pellet separated on 16 % SDS gels, Lane 1 and 2- cells lysed by bead milling, Lane 3 and 4- cells lysed by osmotic shock, Lane 5 and 6- cells lysed by sonication and Lane 7 and 8- cells lysed by freeze crushing.

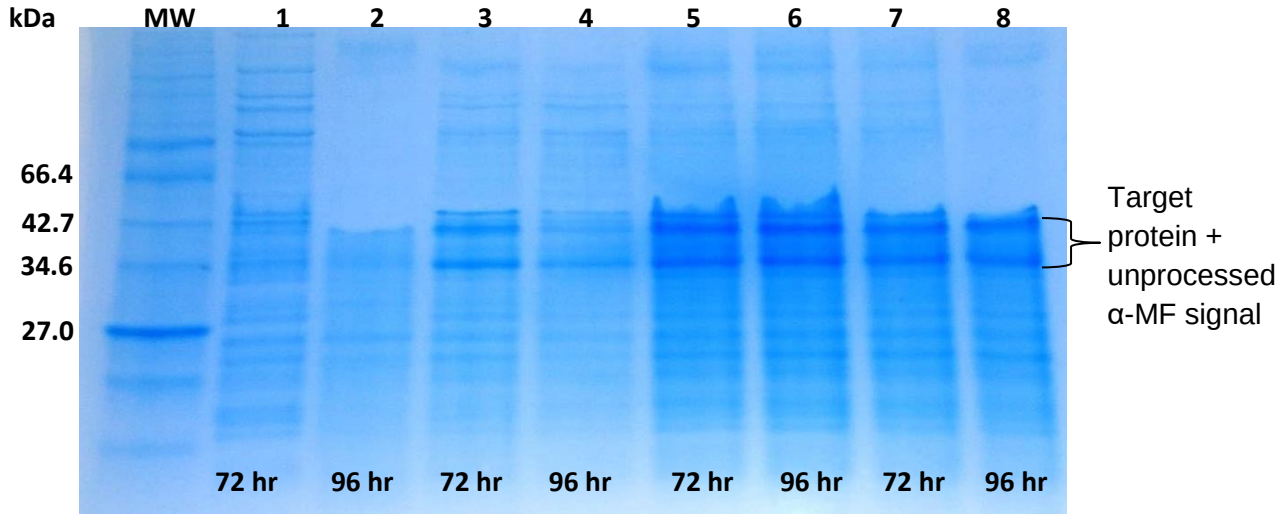



**Figure 4.9:** SDS-PAGE analysis of intracellular fluorinase expression. Expression induced with 0.5 % methanol and soluble protein (crude extract) analysed on 16% SDS-PAGE gels. Lane 1 and 2 - Strain 1, Lane 3 and 4- Strain 2, Lane 5 and 6- Strain 3, Lane 7 and 8- Strain 4, MW- Unstained protein marker broad range (2-212 kDa) (New England Biolabs).

### 4.8 Fluorinase enzyme activity assay using HPLC

In order to detect and determine if the recombinant protein was active, an enzyme activity assay was performed. HPLC was used to analyse 5-FDA the product of the reaction the presence of 5-FDA in the enzyme reaction confirms the presence of the fluorinase protein. In order to identify the retention time of 5-FDA, a standard solution containing 2mM 5-FDA in 50 mM sodium phosphate buffer was eluted through the column (Figure 4.10A) the retention time was 1.5 min. A second standard solution containing 5 mM 5-FDA and 0.1 mM SAM in 50 mM sodium phosphate buffer was eluted through the column (Figure 4.10B). The 5-FDA eluted first at 1.4 min and SAM eluted at 3.4 min.

All enzyme reactions were performed in 50 mM sodium phosphate buffer (pH 7.4) with 20 µl supernatant (media) or crude protein extract, 50 mM NaF and 0.5 mM SAM. The secreted protein fraction for strain 1 was tested for enzyme activity. Figure 4.11 shows the HPLC profile of the enzyme assay for strain 1 induced for 72 hr a peak corresponding to 5-FDA eluted at 1.5 min.

To test the intracellular protein fraction for enzyme activity the crude protein extract was used. The total protein was extracted in breaking buffer. In order to determine background noise due to the buffer, 20 µl of the buffer was eluted through the column. Figure 4.12 shows the HPLC profile for the breaking buffer and the background peaks expected for the test samples.

The SDS-PAGE analysis of the intracellular protein fraction showed a prominent band corresponding to the fluorinase for strain 2 induced at 72 hr and for strain 3-4 at both 72 and 96 hr. Based on these results the samples with evidence of the fluorinase enzyme were tested for enzyme activity.

For strain 2 induced for 72 hr to exhibit fluorinase activity a peak corresponding to 5-FDA eluted at 1.49 min and one corresponding to SAM eluted at 3.2 min (Figure 4.13). For strain 3 induced for 72 hr, 5-FDA eluted at 1.46 min and SAM at 3.82 min (Figure 4.14A) for the 96 hr induced sample, 5-FDA eluted at 1.5 min and SAM at 3.89 min (Figure 4.14B). For strain 4 induced at 72 hr 5-FDA eluted at 1.46 (Figure 4.15A) and for the 96 hr induced sample 5-FDA eluted at 1.47 and SAM at 3.82 (Figure 4.15B). A number of background peaks can be seen which is due to the breaking buffer.

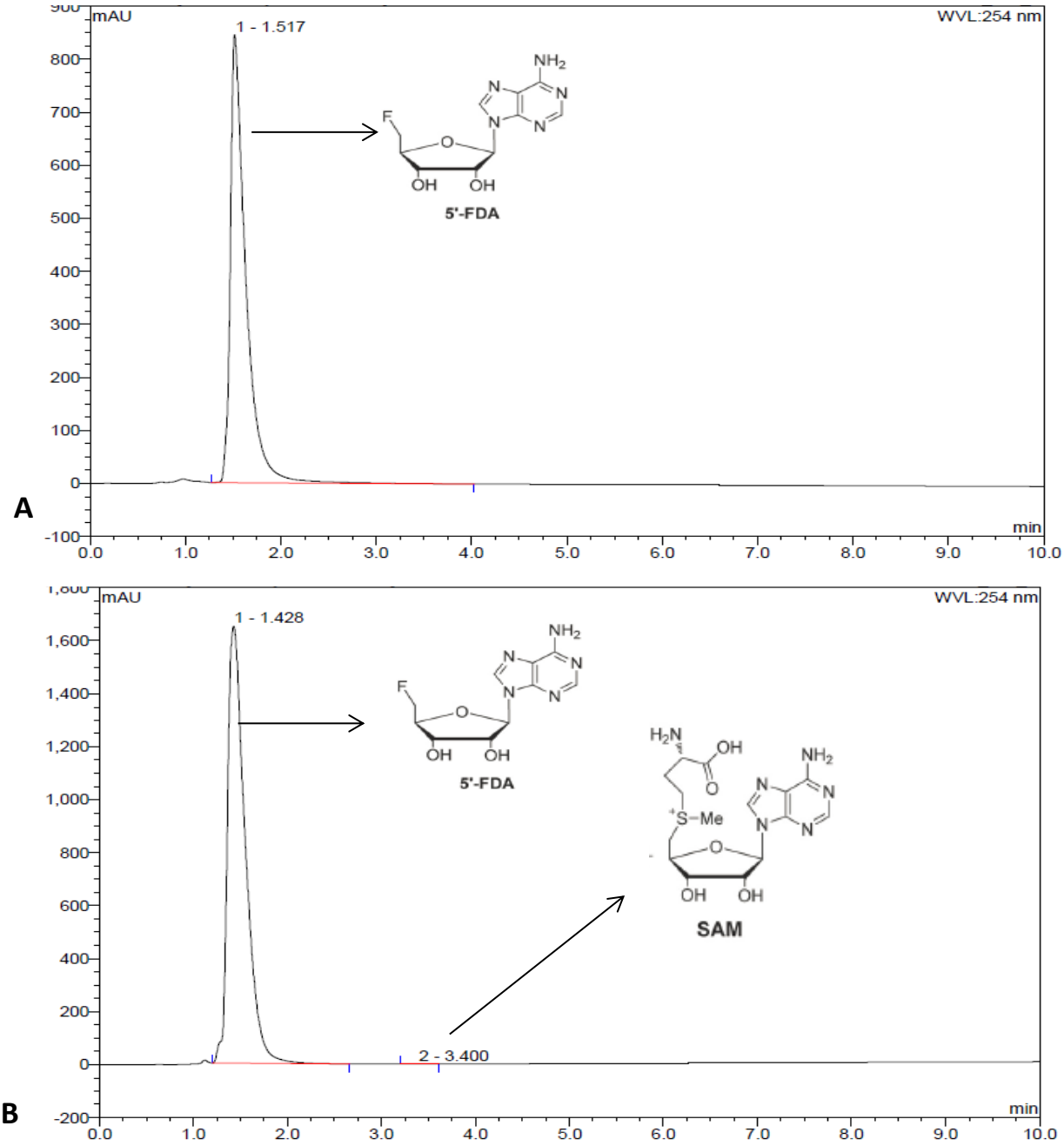



**Figure 4.10:** HPLC profile of 5'-FDA and SAM standards for the fluorinase enzyme activity assay **A.** HPLC profile of 2mM 5-FDA, which eluted at 1.517 min. **B.** HPLC profile of 5mM 5-FDA and 0.1mM SAM separated on the column, 5-FDA eluted first at 1.428 min followed by SAM at 3.4 min. Standards separated on an Ascentis® Express HILIC HPLC 2.7µm column in Sodium phosphate buffer, pH 7.4. A gradient mobile phase with an initial mobile phase of 95:5 and final mobile phase of 80:20, (Sodium phosphate buffer: Acetronitrile) was used. The flow rate was set to 1ml/min for 10 min. The graph shows milliabsorbance units at 254 nm (Y-axis) and retention time in min (x-axis).

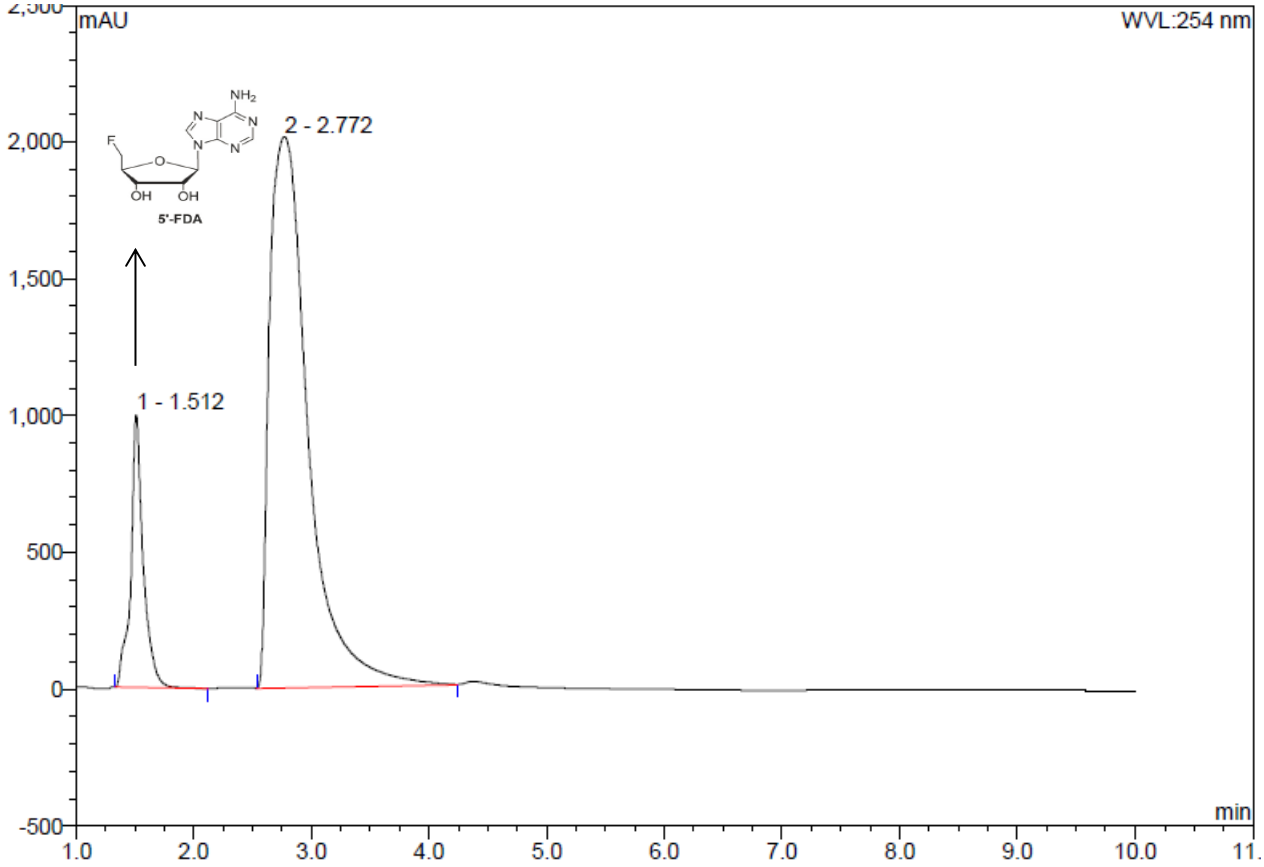



**Figure 4.11:** HPLC profile of the fluorinase enzyme assay testing the secreted protein fraction for activity. Strain 1 induced for 72 hr to express the fluorinase enzyme of was used. The enzyme reaction was performed in 50 mM sodium phosphate buffer pH 7.4, containing 20 µl supernatant (media), 50 mM NaF and 0.5 mM SAM. Twenty microliters of the enzyme reaction was separated on an Ascentis® Express HILIC HPLC 2.7µm column. A gradient mobile phase of 95:5 to 80:20, (Sodium phosphate buffer: Acetronitrile) was used with a flow rate of 1ml/min. 5-FDA eluted at 1.512 min followed by an unknown molecule at 2.772 min. The graph shows milliabsorbance units at 254 nm (Y-axis) and retention time in min (x-axis).

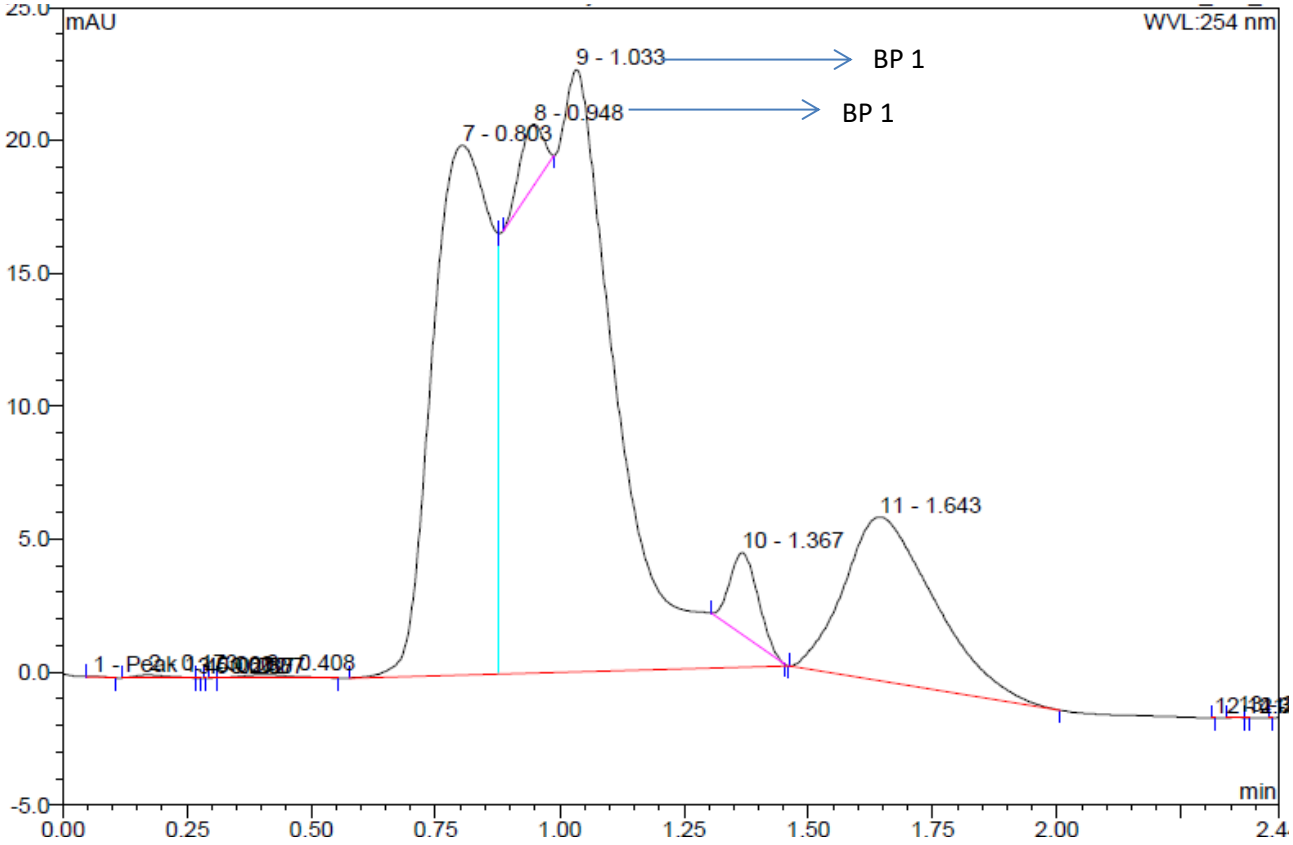



**Figure 4.12:** HPLC profile of the breaking buffer -to determine background peaks for intracellular enzyme activity assays. Breaking buffer c containing 2 mM PMSF; 1 mM EDTA; 5% glycerol in 50 mM Sodium phosphate buffer was separated on an Ascentis® Express HILIC HPLC 2.7µm column. A gradient mobile phase of 95:5 to 80:20, (Sodium phosphate buffer: Acetronitrile) was used with a flow rate of 1ml/min. The graph shows milliabsorbance units at 254 nm (Y-axis) and retention time in min (x-axis).

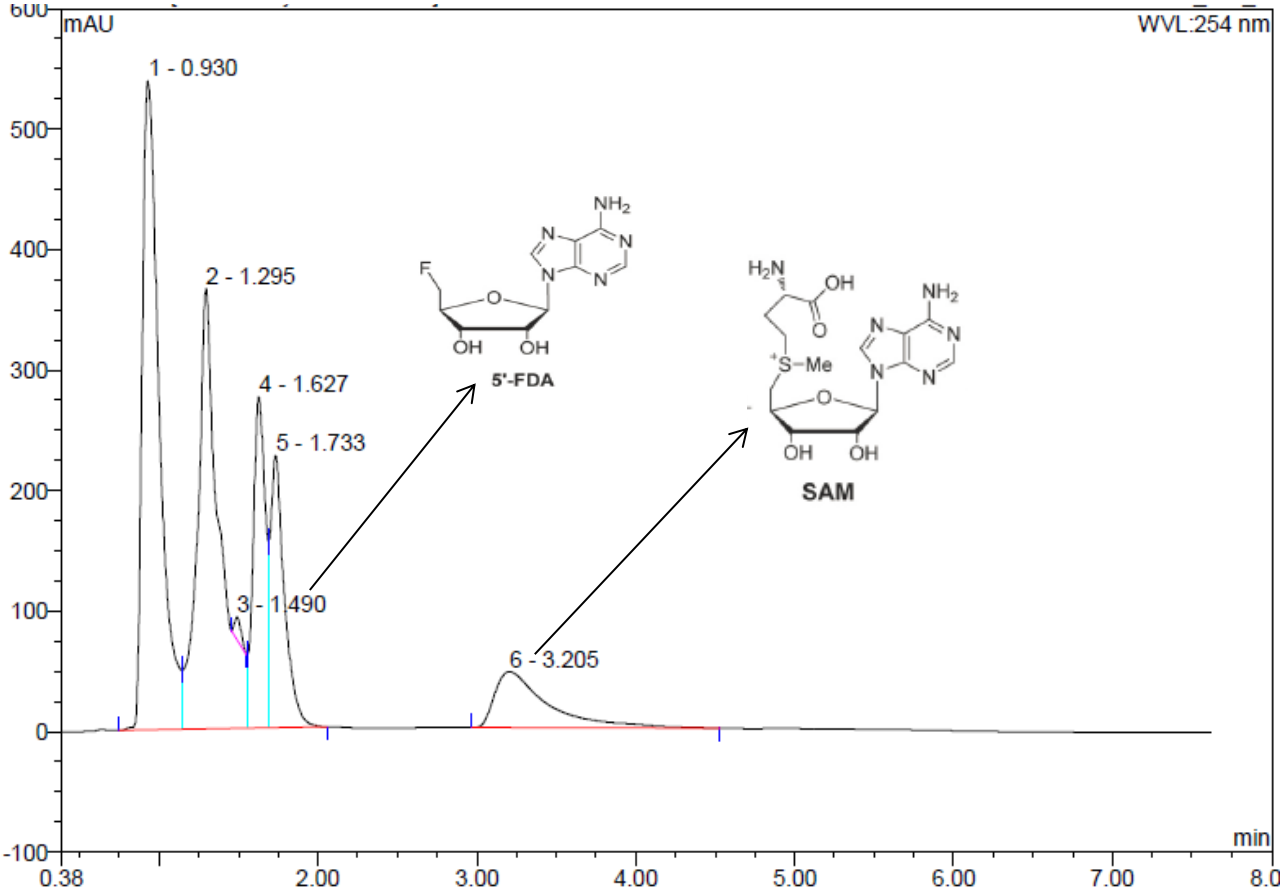



**Figure 4.13:** HPLC profile of the fluorinase enzyme assay testing the intracellular protein fraction for activity. Strain 2 induced for 72 hr to express the fluorinase enzyme of was used. The enzyme reaction was performed in 50 mM sodium phosphate buffer pH 7.4 containing 20 µl crude extract, 50 mM NaF and 0.5 mM SAM. Twenty microliters of the enzyme reaction was separated on an Ascentis® Express HILIC HPLC 2.7µm column. A gradient mobile phase of 95:5 to 80:20, (Sodium phosphate buffer: Acetronitrile) was used with a flow rate of 1ml/min. 5-FDA eluted at 1.490 min followed by SAM at 3.205 min (background peaks due to breaking buffer components). The graph shows milliabsorbance units at 254 nm (Y-axis) and retention time in min (x-axis).

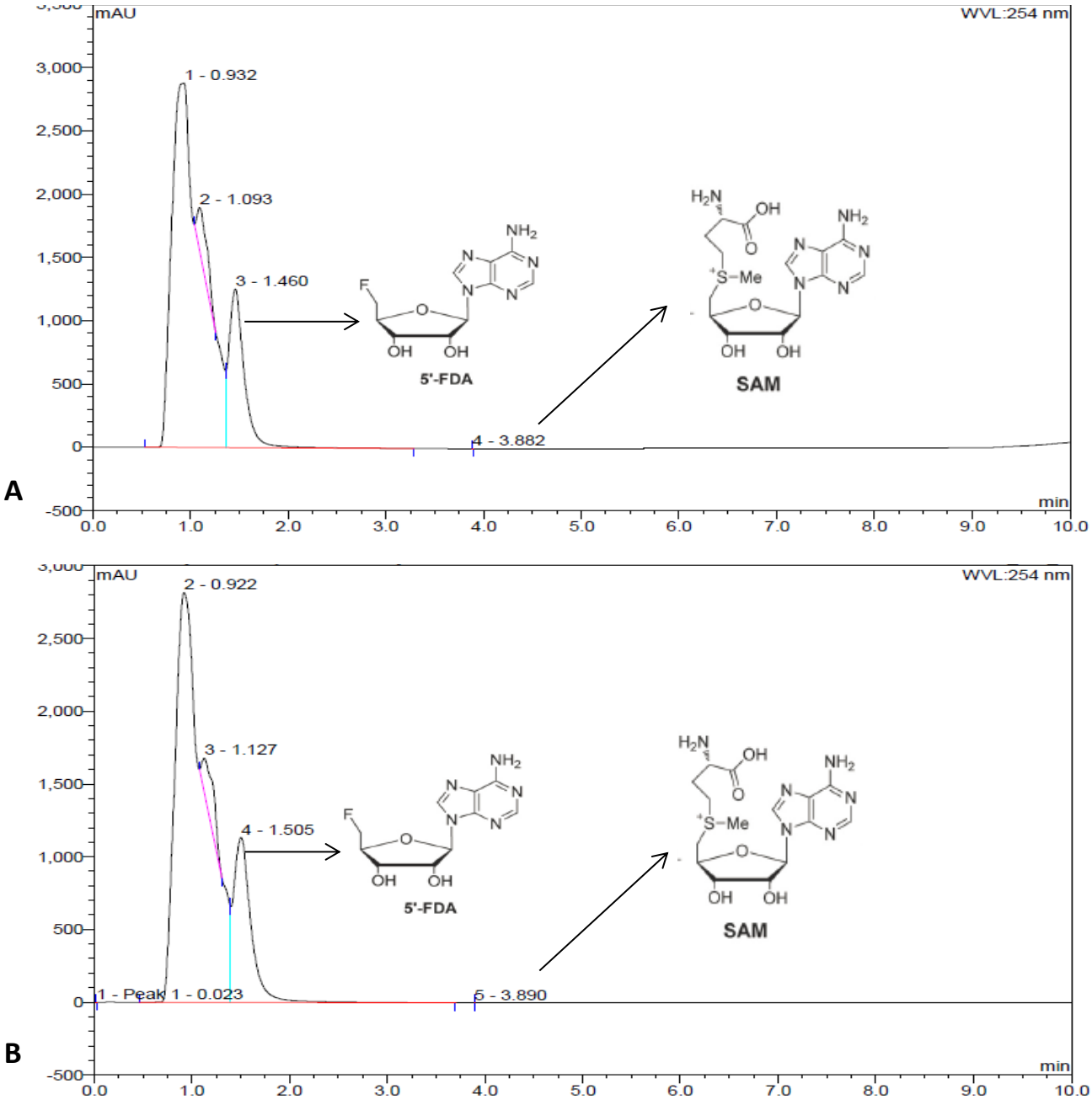



**Figure 4.14:** HPLC profile of the fluorinase enzyme assay testing the intracellular protein fraction of strain 3 for activity. The enzyme reaction was performed in 50 mM sodium phosphate buffer pH 7.4 t containing 20 µl crude extract, 50 mM NaF and 0.5 mM SAM. Twenty microliters of the enzyme reaction was separated on an Ascentis® Express HILIC HPLC 2.7µm column. A gradient mobile phase of 95:5 to 80:20, (Sodium phosphate buffer: Acetronitrile) was used with a flow rate of 1ml/min. **A.** Strain 3 induced for 72 hr to express the fluorinase enzyme was tested, 5-FDA eluted at 1.460 min followed by SAM at 3.882 min. **B.** Strain 3 induced for 96 hr to express the fluorinase enzyme was tested, 5-FDA eluted at 1.505 min followed by SAM at 3.890 min. The graph shows milliabsorbance units at 254 nm (Y-axis) and retention time in min (x-axis).

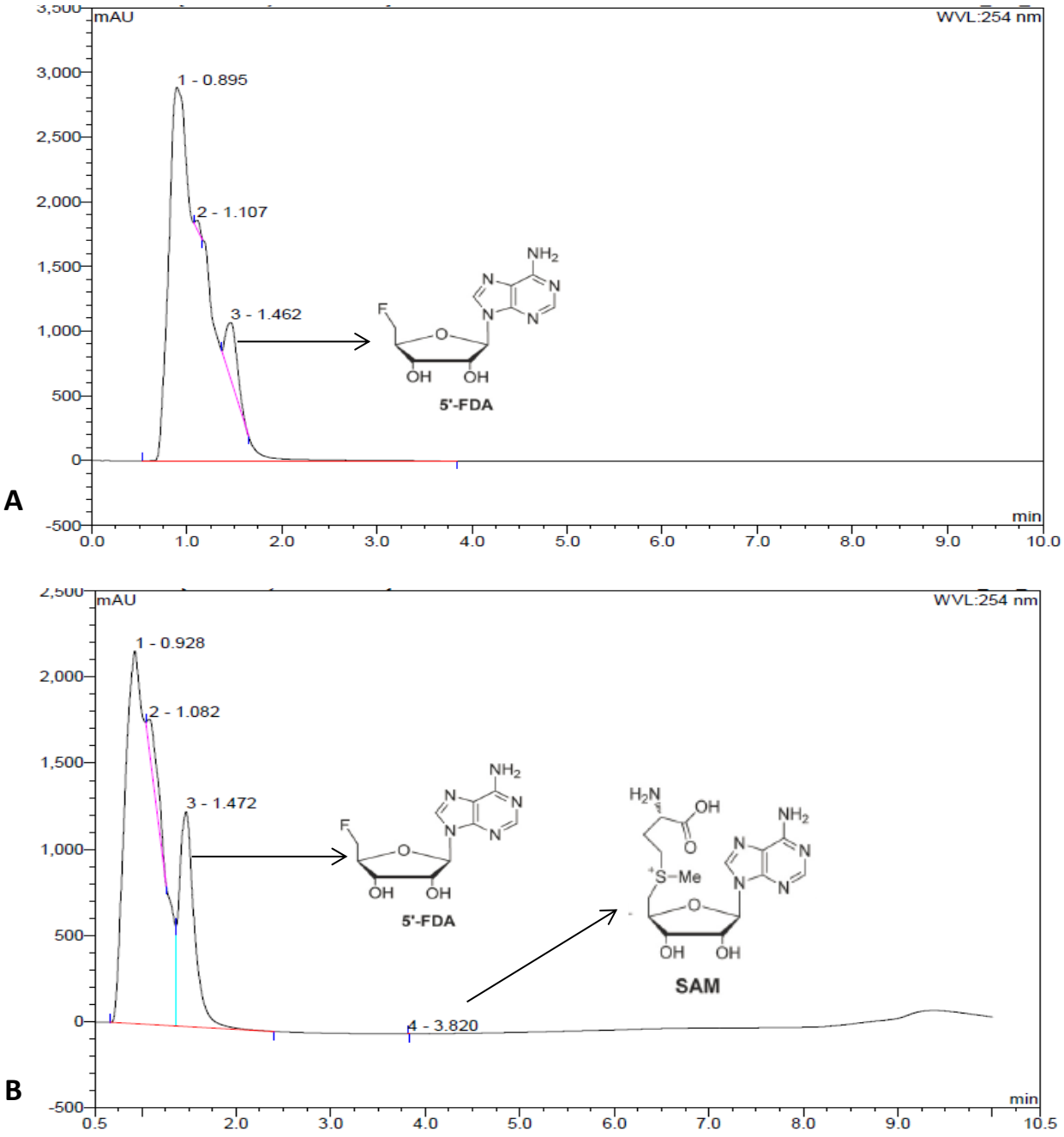



**Figure 4.15:** HPLC profile of the fluorinase enzyme assay testing the intracellular protein fraction of strain 4 for activity. The enzyme reaction was performed in 50 mM sodium phosphate buffer pH 7.4 t containing 20 µl crude extract, 50 mM NaF and 0.5 mM SAM. Twenty microliters of the enzyme reaction was separated on an Ascentis® Express HILIC HPLC 2.7µm column. A gradient mobile phase of 95:5 to 80:20, (Sodium phosphate buffer: Acetronitrile) was used with a flow rate of 1ml/min. **A.** Strain 4 induced for 72 hr to express the fluorinase enzyme was tested, 5-FDA eluted at 1.462 min, no peak for SAM seen **B.** Strain 4 induced for 96 hr to express the fluorinase enzyme was tested, 5-FDA eluted at 1.472 min followed by SAM at 3.820 min. The graph shows milliabsorbance units at 254 nm (Y-axis) and retention time in min (x-axis).

# 5. Discussion

## 5.1 Construction of recombinant PichiaPink™ vectors (pPinkα-HC) with the gene of interest

Construction of the pPinkα-HC-*fLA* plasmid proved successful. Sequencing results confirmed the gene sequence and that the gene was in frame with the secretion signal. Due to the limited restriction sites in the multiple cloning site of the pPinkα-HC (shown in figure 4.1) devising a cloning strategy can be difficult. The only restriction enzyme that could be used was *Fse* I, which is a low efficiency enzyme. Additional steps had to be taken to ensure the enzyme was not affected by PCR components and salts in solution. Further, *Stu* I cuts within the gene of interest and therefore steps were taken to ensure the enzyme was completely removed before setting up a ligation reaction. Cloning in Pichiapink™ is simple and easy to follow as it is in *E. coli.*

## 5.2 Transformation of the four PichiaPink™ strains with the expression construct

### *5.2.1 Factors influencing the transformation efficiency*

The transformation efficiency is calculated as the number of transformants per microgram of DNA. The transformation efficiencies for each strain was low, for example 37 white colonies were obtained for strain 2 (3.7 transformants per microgram of DNA).

The transformation efficiencies are lower for *P. pastoris* compared to other yeast species like *S. cerevisiae.* Transformation is an inefficient process in *P. pastoris* because not only must the foreign DNA enter the cells, it must also integrate into the host genome. Entry of the DNA involves passage through the cytoplasm, followed by the nucleus and finally locating the homologous region in the host genome. The natural and induced competence of the cells is a possible rate limiting step in this process as the cells must be competent to take up foreign DNA from the external environment (Klinner and Schäfer, 2004).

A study done by Wu and Letchworth, (2004) showed that there are several factors that influence the transformation efficiencies in *P. pastoris*. These factors include: the concentration of foreign DNA, the cell density, the cell phase, the choice of integration site as well as the electroporation parameters.

Wu and Letchworth (2004) determined that the cell phase and density as well concentration of linearized DNA influences the transformation efficiency. Their findings were that higher concentrations of DNA (300-1000 ng) yielded lower transformation efficiencies compared to lower concentration of DNA (1-10 ng). Further, it was found that transformation efficiencies

were highest for yeast cells grown to an $OD_{600}$ of 1.5, with the best results achieved between $OD_{600}$ of 1.15-2.6. Values above or below this yielded a drop in the transformation efficiency. Transformations were more efficient at high cell densities, $2.5X10^9$ cells per 80 µl yielded transformation efficiencies of $67.6±44.2X10^4$ and $10X10^9$ cells per 80 µl yielded transformation efficiencies of $334.7±119.3X10^4$.

The transformation efficiency is affected by the site of integration, Wu and Letchworth, (2004) demonstrated that the pPIC9K vector digested at the selection marker (*HIS4*) or at the 3' *AOX1*, region yielded lower transformation efficiencies than those digested in the 5' *AOX1* promoter region. The alternative to integration at the *TRP2* region for pPinkα-HC was to linearize the plasmid in the *AOX1* region. Further, electroporation parameters affect the transformation, higher voltages of 2000 V yielded lower transformations compared to 1 500 V and lower capacitance of 25 µF yielded better results compared to higher capacitance at 50 µF. The Gene Pulser Xcell™ electroporation system (Biorad) was used, employing the built in protocol for *P. pastoris.* In future work, variation in the voltage should be explored, as lowering the voltage could improve transformation efficiencies.

In this research, the protocol for transformation was followed as per the Invitrogen guide, the PichiaPink™ cells were grown to a cell density between 1.3 and 1.5 which according to Wu and Letchworth, (2004) is in the optimal range. According to the Invitrogen guide 1 $A_{600}$= $5X10^7$ cells per ml, Wu and Letchworth used the same formula and strains of *P. pastoris* that was used in this study. For the transformation the PichiaPink™ cells were grown to an $OD_{600}$ of 1.5 in 100 ml of media which is equivalent to $7.5X10^9$ cells and the cells were concentrated in 300 µl of sorbitol. For each transformation 80 µl of cells were used, which is equivalent to $2X10^9$.cells per 80 µl reaction. The Invitrogen guide recommends you use 10 µg of DNA per transformation reaction, whilst Wu and Letchworth showed that 0.3-1 µg of DNA yields the lowest transformation efficiencies. In light of the contradictory findings, the amount of DNA used to transform PichiaPink™ cells was too high, which could account for the low number of transformants produced. For the transformation, the concentration of DNA used was well over the optimal range demonstrated by Wu and Letchworth, 2004, where 1000 ng of DNA yielded the lowest whilst 10 ng yielded the highest. A recommendation for future research is to optimise transformation by varying DNA concentrations as well as cell densities

The method used to make competent cells should also be optimised. According to Klinner and Schäfer, (2004) integration is affected by the method of transformation and the cell competence. Wu and Letchworth, (2004) showed that treating cells with lithium acetate

(LiAc) and dithiothreitol (DTT) prior to sorbitol increased the overall transformation efficiencies.

From the results it was seen that recovery times clearly affected the transformation efficiencies, as reactions incubated for longer time periods (figure 4.5 E-F) showed a higher number of transformants compared to those incubated for shorter times, 6 hr (figure 4.5G-H). Equal amounts of the transformation reaction was plated on each plate and incubated for different time periods. Varying the recovery temperature seemed to have no effect on the transformation efficiency, as the results for both 28 °C and 37 °C were the same (figure 4.5E-F).

The transformation of PichiaPink™ was not optimal yielding low transformation efficiencies; this was the case for all four strains. Optimising the parameters of transformation can be done by exploring the factors mentioned above. With increasing transformation efficiencies a higher number of transformants will be available for screening to find the integrant with the highest expression levels.

## 5.3 Screening of PichiaPink™ clones for integration

### *5.3.1 ADE2 complementation*

When screening for positive integrants a direct PCR approach was taken, where the yeast cells were lysed by several rounds of microwaving, followed by freezing and boiling. The supernatant containing genomic DNA was used as a template in a PCR with the *AOX1* and *CYC1* plasmid specific primers. Attempts to extract genomic DNA failed as the cells may have not been lysed efficiently using the ZR Fungal/Bacterial DNA MiniPrep™ kit (Zymo Research).

The results from the colony PCR with the plasmid specific primers yielded prominent bands corresponding to the parent plasmid control. The size of the amplicon was 426 bp which corresponds to the size of the promoter, α-MF signal sequence and the transcription termination region (Figure 4.6A). A faint band corresponding to the pPinkα-HC-*fLA* construct positive control was seen for some of the PichiaPink™ clones (Figure 4.6A). This was confirmed by a nested PCR with gene specific primers, and a 900 bp product was amplified from the first PCR reaction (Figure 4.6B), which corresponds to size of the gene. The gene was confirmed by PCR. Confirmation that this faint band was not due to contamination e,g contamination of the primers with the positive control, the no template control (water only) showed no bands. Secondly the PCR products were sequenced to ensure the band was corresponding to the gene of interest. The linearized DNA for transformation was selected

from a single positive colony and these were also sequenced to ensure the gene sequence was correct and cloned in correctly. Therefore it is highly unlikely that these were mixed with the parent plasmid lacking the gene of interest.

From these results it was clear the gene was inserted into the host genome, however expression cassettes without the gene were also integrated. Although PCR is not quantitative it is clear this would lead to a lower copy number of the gene. The exact copy number can only be determined by screening using a quantitative method such as real-time quantitative PCR or by western blots with an antibody to the protein.

The colour of the colonies can be related to expression levels of the gene of interest. Pink colonies express very little *ADE2* gene product and white colonies express higher amounts of the *ADE2* gene product, which indirectly relates to the gene copy number. In this situation this feature was not applicable as the results showed the expression cassette without the gene of interest was inserted into the host genome. These negative cassettes include the *ADE2* gene which would influence the levels of *ADE2* gene product being produced thereby contributing to the white phenotype and large size of the PichiaPink™ clones.

As a whole, selection based on *ADE2* complementation is a convenient method for screening for positive integrants. The downside to this screening method is the fact that you cannot identify which white colonies have the higher copy number.

Sunga *et al* (2008) demonstrated a method for screening multiple-copy vectors called posttransformational vector amplification (PVA). This system is based on antibiotic selection markers, to screen for clones with the highest copy number. The original transformants can be plated on a master plate of low antibiotic concentration. Colonies that grow on the master plate are then replica plated on plates containing increasing amounts of antibiotics; in this way, the colonies that are capable of growing on media with the highest concentrations of antibiotic are essentially clones that contain multiples copies of the selection marker (Sunga *et al,* 2008). Unfortunately with the PichiaPink™ system there is no way to identify which of the white clones have a higher copy number.

### *5.3.2 Integration by insertion and genetic instability*

The presence of the expression cassette without the gene of interest can be explained by the genetic instability associated with integration by insertion. Multi-copies are produced by tandem multiple integration events and due to repeated recombination events (Daly *et al.,* 2005), which leads to repeat sequences in the genome. Unlike homologous recombination, recombination can occur between repeated sequences on a chromosome. This leads to

chromosomal rearrangements such as deletion, insertions or inversions (Aguilera *et al.,* 2000). Deletions most commonly occur when there is a long section of repeat sequences, where these deletions can occur by any of the primary mechanisms (figure 5.1).

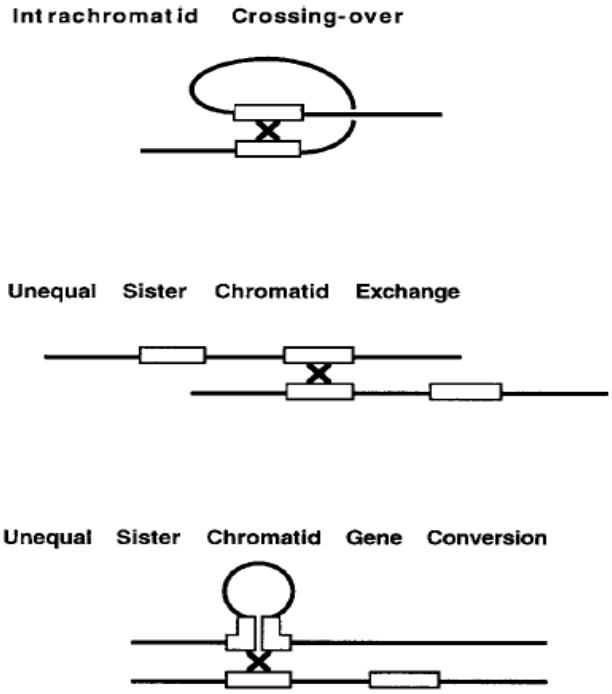



**Figure 5.1:** Recombination between repeat sequences lead to deletions and insertion in chromosomal DNA, this occurs by: Intrachromatid crossing-over, unequal sister chromatid exchange and gene conversion (adapted from Aguilera *et al.,* 2000).

These repeated regions of homology can recombine and whole, or parts ,of the vector can be excised from the host genome by the loop-out effect outlined in figure 5.2 (Aw and Pollizi, 2013). Aw and Polizzi (2013) demonstrated that with an increase in the expression cassette copy number there is an increased number of loop outs that can occur. *P. pastoris* is easily manipulated to take up foreign DNA and is quite easily genetically modified. Aw and Polizzi (2013) stated “it is perhaps this highly recombinogenic nature that results in unstable clones. Theoretically, an organism that so readily accepts DNA can lose it just as fast”.

From the results, it is evident that the integration of the gene into the host genome for each of the strains was unstable. This lead to the production of transformants with a low-copy number of the gene, this is based solely on the presence of the expression cassette without the gene. As the precise gene copy number was not quanitified, due to the fact that an antibody to the gene of interest was not available.

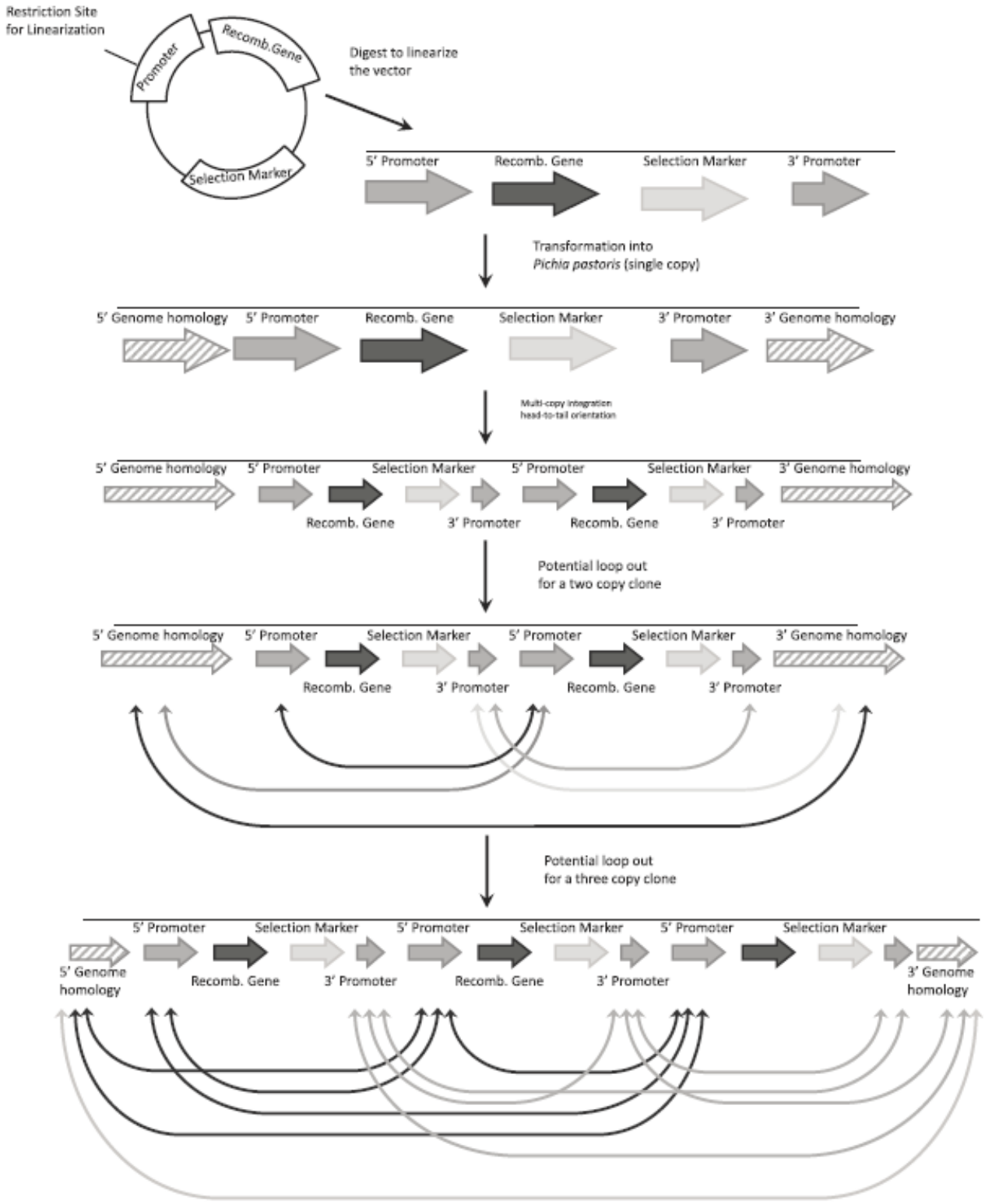



**Figure 5.2:** illustrates the mechanism of the loop-effect that leads to deletion of whole or parts of the expression cassette in *P. pastoris*, the number of possible loop-outs increase with increasing repeated sequences (adapted from Aw and Pilozzi, 2013).

The use of a vector that contains a gene for antibiotic resistance should be considered when there is no antibody available for the gene of interest. This would assist in determining which clones have the highest copy number versus moving onto expressing a large number of clones and then determining the highest expression. In this way, time and cost can be reduced when a group of integrants with high copy numbers have been screened for using the PVA method, then only can you start expressing protein. This method evidently would only apply once stable integration has been established.

According to Klinner and Schäfer, (2004) integration is affected by the method of transformation and the cell competence. Therefore, as discussed in section 5.2.1, transformation protocols should be optimised before continuing onto screening and transformation.

### 5.4 Comparison of mechanical disruption methods for protein extraction

The morphology of yeast cell walls change during growth and cell division, and could also change following genetic modification (Klimek-Ochab *et al.,* 2011). Therefore, cells from the expression trials were used to test disruption methods as these were the cells that the protein would ultimately be extracted from.

In an attempt to determine the most efficient method for lysing *P. pastoris* cells for the extraction of protein, four cell disruption methods were tested. Bead milling, sonication and freeze crushing are mechanical methods and osmotic shock a chemical method. Another chemical method is the use of enzymes. Enzymatic digestion is a recommended method for cell disruption in yeast, however this option was not explored as it is costly for use in large applications, in terms of preparing a large number of samples for protein extraction. Although mechanical methods can be laborious and time consuming, these are non-expensive compared to enzymatic methods. The use of bead mill machines and homogenisers can make these methods more efficient; however this equipment was not sourced for use in this research.

The breaking efficiency was calculated by measuring the percentage of broken cells based on the number of cells before and after disruption. The results in table 4.1 showed that the sonication method yielded the highest breaking efficiency of 71 %. The bead milling and freeze-crushing had similar breaking efficiencies of 64.6% and 66.8% respectively. Overall the percentage of broken cells was greater than fixed cells for the mechanical methods. The osmotic shock method was unsuccessful in disintegrating the cell wall.

The concentration of total protein in the supernatant after lysis was highest for samples disrupted by the freeze crushing method (table 4.1). Although sonication had a higher

percentage of broken cells the protein concentration was lower. Stathopulos *et al.,* (2004) emphasises the negative effect sonication has on biological systems, this includes excessive heating and sheer stress that could lead to degradation of proteins. The study demonstrated that sonication causes aggregation of proteins and with heating theses aggregates have a seeding effect on soluble protein. Therefore aggregates form due to sonication and further aggregation occurs due to the heating associated with sonication. These aggregates are insoluble and are pelleted by centrifugation. The findings of Stathopulos *et al.,* (2004) can explain the high breaking efficiency and lower concentration of soluble protein seen for the sonication method. The SDS-PAGE gel in figure 4.8 shows the supernatant (containing released protein) and the pellet (remaining undisrupted cells), for the sonication method the majority of the protein is seen in the pellet which could be due to the formation of insoluble aggregates.

The method chosen for breaking cells for protein extraction has an effect on the quality of the final product yield. Concerns with sonication and bead milling include excessive heating which could lead to the denaturation of protein. Although precaution was taken by including additional breaks on ice in between treatment, the build-up of excess heat cannot be controlled. Advantages of the freeze crushing method is the lack of heating compared to other mechanical methods. This is advantageous when extracting sensitive proteins such as enzymes minimising chances of denaturation due to excess heat.

The liquid nitrogen freeze crushing method was selected as the method to extract intracellular proteins from the induced cells. This was based on three parameters, the fact that the protein concentration was higher for this method, it showed to have a breaking efficiency close to that of the 'best method' and it provided the ideal environment for extraction of the enzyme in this study as the concern of degradation due to excessive heating was minimal. Inconsistency is a concern with the liquid nitrogen- freeze crushing method, as this is a manual method, the force with which the cells are crushed is inconsistent from sample to sample. This can be a major problem as results may change from one sample to the next. In this case a small amount of cells can be removed from the sample and the degree of lyses can be checked. If the protein concentration is found to be too low the cells can be crushed further to increase protein yields.

## 5.5 Analysis of protein expression by SDS-PAGE

### *5.5.1 Analysis of the secreted protein fraction by SDS-PAGE and the problems associated with the α-MF secretion signal*

To test if the recombinant protein was secreted into the media by the α-MF secretion signal, the secreted protein fraction was analysed by SDS-PAGE. For strain 2-3 (Figure 4.7B-D) no bands were seen corresponding to the size of the fluorinase (32.2 kDa), thus indicating that there was no recombinant protein being secreted into the media. For strain 1 there were multiple bands seen between 27.0 and 34.6 kDA but this may be due to inefficient separation of the pellet from the supernatant.

Common practise is to express a protein intracellularly if it is not naturally secreted by its host organism (Cregg *et al,* 2000), because it has been reported that intracellular proteins have been successfully secreted in the *P. pastoris* system (Daly *et al.,* 2005), secretion of the fluroinase was attempted. Opting for targeted secretion of the fluorinase that is naturally an intracellular protein in *S. cattleya* may have been the cause of the failed secretion. However, failed secretion may be attributed to the problems related to the α-MF secretion signal. A major problem associated with the α-MF secretion signal sequence is the processing of the signal from the recombinant protein. It is not the ability of the secretion signal to direct secretion that is the problem, the problem is the proteolytic processing of the α-MF secretion signal.

The processing of the pro-sequence at the *Kex2* site, which is cleavable by endo-peptidases in the golgi apparatus has proven to be problematic (Cregg, 2007; Daly et al 2005). A significant number of proteins have shown that during heterologous production, the cleavage at the *Kex2* site does not occur. Therefore the foreign protein is produced with whole or parts of the secretion signal still attached to the protein (Cregg, 2007). A solution to this problem was the introduction of the *Ste13* site (Glu-Ala-Glu-Ala), an additional cleavage site upstream of the *Kex2* site (Figure 5.3).

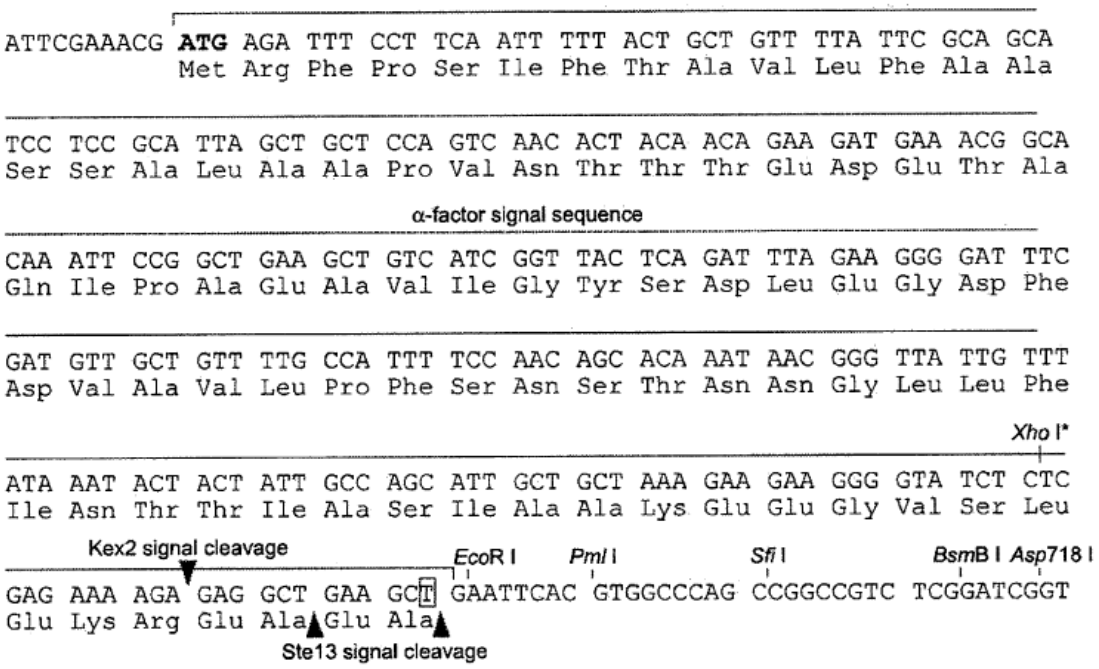



**Figure 5.3:** sequence of the alpha mating factor secretion signal, with the *Kex2* and *Ste* 13 signal cleavage sites (adapted from Cregg, 2007).

The introduction of the *Ste13* allows for complete proteolytic processing of the *Kex2* site and the smaller *Ste13* site remains attached to the recombinant protein. The lack of proteolytic processing of the *Kex2* site in some recombinant proteins occurs due to the structure and folding of the protein. The N-terminus of the protein folds in such a way that the cleavage site becomes unavailable for cleavage by the endo-peptidases. The addition of the *Ste13* site extends the *Kex2* cleavage site making it available for cleavage, however the *Ste*13 site remains which is much smaller (Cregg, 2007). In some cases the *Ste13* site is cleaved successfully by dipeptidyl aminopeptidases (Glick and Fitzgerald, 2014).

The pPinkα-HC vector does not include the Glu-Ala-Glu-Ala *Ste13* cleavage site after the *Kex2* site. The sequence of the alpha mating factor secretion signal is given in the appendix (Figure A.2). Thus the problem associated with the proteolytic processing of the *Kex2* cleavage site could account for the lack of secretion during the expression of the fluorinase gene.

The use of other signal sequences such as the *P. pastoris* acid phosphatase signal should also be explored in future work. A set of secretion signals are available for the PichiaPink™ system, these signals can be cloned into the vector in frame with the gene of interest. Alternatively, when expressing a protein for the first time, one should attempt both intracellular and secreted expression, especially in the case where the foreign gene is not naturally secreted by the host organism. Unfortunately pinpointing the exact problem with

failed secretion is difficult. In some cases *P. pastoris* just cannot secrete certain proteins even with the presence of a secretion signal, and these proteins often remain in the ER or in the golgi apparatus (Lin-Cereghino *et al.,* 2013).

*5.5.2 Analysis of intracellular protein fraction by SDS-PAGE*

Following analysis of the culture supernatant for secreted protein expression, the SDS-PAGE gels showed no evidence of the recombinant protein being secreted into the media. The intracellular protein fraction was analysed by SDS-PAGE to confirm whether expression was low or whether the secretion signal failed to secrete the protein.

Figure 4.9 shows the SDS-PAGE gel for analysis of the intracellular protein fraction of strain 1-4 induced at 72 and 96 hr. Two prominent bands were seen on the gel for strain 2-3. These bands correspond to the fluorinase protein (32.2 kDA) with the unprocessed α-MF secretion signal. Lin-Cereghino *et al.,* 2013 described that *P. pastoris* just cannot secrete certain proteins even with the presence of a secretion signal; these proteins often remain in the ER or in the golgi apparatus.

The size of the pre-pro sequence is 9.35 kDA and the pro-sequence is 7.26 kDA. If the protein remained in the ER with the pre-pro sequence attached the protein and signal would resolve at 41.55 kDA. If the protein remained in the golgi attached to the pro-sequence due to failed proteolytic processing at the *Kex2* site, this would resolve at 39.46 kDA. In figure 4.9 a bright band can be seen below 42.7 kDA which corresponds to the above. The band seen near 34.6 kDA in figure 4.9 corresponds to the target protein (32.2 kDA) and unprocessed sections of the α-MF secretion signal, as described by Cregg, (2007) when cleavage at the *Kex2* site does not occur, the foreign protein is produced with whole or parts of the secretion signal still attached.

As mentioned before both secreted and intracellular protein expression should be attempted. However, only the pPinkα-HC vector was available, if the expression system is to become established in any laboratory, it would be important include the pPink-HC and pPink-LC vector for intracellular protein production.

According to Cregg, (2007) in *E. coli* systems 10% or more of the total protein can be recombinant protein in early phases of development. Due to this the recombinant protein can easily be detected on SDS gels. In the case of *P. pastoris* the recombinant protein represents as little as 1% of the total protein in early stages of development. Finding a commercially available antibody to the foreign protein is not always an option and therefore SDS-PAGE analysis is the method used to screen for strains expressing the protein. With

the low concentration of expressed protein and background proteins of the yeast, this can be difficult to do (Li *et al,* 2011). A sensitive assay such as an enzymatic assay or western blot for detecting recombinant protein is critical (Cregg, 2007).

The SDS-PAGE results for the expression proved difficult to interpret. To test for the recombinant protein and to test if the protein was functional a HPLC based enzyme assay was done. The secreted and intracellular protein fraction for each strain was tested.

**5.6 Fluorinase enzyme activity assay using HPLC**

In order to detect and determine if the recombinant protein was active, an enzyme activity assay was done. The assay is based on the reaction of the fluorinase enzyme, which converts SAM and fluoride ion to 5-FDA and methionine. HPLC is used to identify 5-FDA the product of the reaction, the presence of 5-FDA in the enzyme reaction confirms the presence of the fluorinase protein and confirms activity.

Based on the results for the analyses of protein expression by SDS-PAGE (figure 4.10), only the secreted protein fraction for strain 1 was assayed as the bands corresponding to the fluorinase were not seen in the intracellular. The intracellular protein fraction for strain 2-4 were assayed, however only the 72 hr induced sample for strain 2 was assayed as the 96 hr induced sample showed less prominent bands.

The retention time for 5-FDA (1.4-1.5 min) and SAM (3.4 min) standards are shown in figure 4.10A-B. Figure 4.11 shows the HPLC profile for the enzyme reaction for the 72 hr induced secreted fraction of strain 1. The profile shows a peak at 1.5 min which corresponds to the standard peaks seen for 5-FDA. For strain 2 the 72 hr induced intracellular fraction was assayed. A peak can be seen at 1.49 min and a second peak at 3.2 min, which corresponds to 5-FDA and SAM respectively (figure 4.13). For strain 3 the 72 hr and 96 hr induced sample produced a peak at 1.46 and 1.5 respectively corresponding to 5-FDA, a second peak eluted at 3.88 and 3.89 respectively which corresponds to SAM (figure 4.14A-B). For strain 4 the 72 hr and 96 hr induced sample produced a peak at 1.46 and 1.47 respectively corresponding to 5-FDA, a second peak eluted at 3.82 for the 96 hr induced sample which corresponds to SAM (figure 4.15A-B).

The presence of 5-FDA for all of the reactions indicate that the fluorinase was active in that it was capable of converting SAM to 5-FDA and methionine. The methionine does not absorb at 254 nm and therefore does not show on the HPLC profile. The presence of the SAM in the assay shows that not all of the reactant was converted in the reaction, the remainder of the

SAM was present in the assay for strain 2 (72 hr), strain 3 (72 and 96 hr) as well as strain 4 (96 hr). The breaking buffer was eluted through the column to determine background noise, the two prominent compounds that eluted at 0.948 and 1.033 (figure 4.12) can be seen on the HPLC profiles for strain 2-4.

For each assay the crude protein sample was used and therefore the assay was not used quantitatively. If however, the fluorinase was purified and known concentrations were used in the assay, then a standard curve could have been determined (Kupiec, 2004). From this the amount of 5-FDA produced could have been quantified and the activity of the fluorinase for each strain could have been more adequately compared. In this study the assay was solely used as a method to qualitatively determine if the protein produced was active and thus the use of a standard curve not necessary.

The HPLC results confirm that the protein failed to secrete for strain 2-4 and was seen in the intracellular fraction, also confirming that the fluorinase produced by strain 1 was secreted as it was not seen in the intracellular fraction (figure 4.9).

Protease knock-out strains are used to decrease protease activity in the media during secreted protein expression. In this study the protease knock-out strains 2-4 failed to secrete the fluorinase protein, whereas the wild-type strain 1 successfully secreted the protein. This finding supports the evidence that wild-type strains are generally healthier. Cregg (2007) describes protease knockout strains for *P. pastoris* as less stable, slower growing, showing lower levels of expression and also lower transformation efficiencies. This indicates the benefits of transforming four strains before moving on to large-scale fermentation processes, allowing for the selection of the strain that yields the highest expression levels.

# 6. Conclusion

Construction of the pPinkαHC-*fLA* construct was successfull and all four strains were transformed with the construct. Although the integrants were unstable all four strains were found to produce fluorinase during expression. Although the secretion failed for strain 2-3, all four strains exhibited the ability to produce a biological active protein which produced 5-FDA in the fluorinase enzyme assay. Additionally, the most suitable method for extracting protein from recombinant PichiaPink$^{TM}$ cells was determined, which was freeze crushing with liquid nitrogen. A number of considerations should be taken into account when using *P. pastoris* as an expression host for the production of industrial enzymes the nature of the protein should be the determining factor when selecting a specific strain or vector. Both intracellular and secreted protein expression should be explored to determine the most suitable method for the targeted protein. Secondly, the transformation efficiency should be optimised before expressing the protein. The concentration of DNA and the cell density should be optimised in order to determine the most efficient transformation protocol. Another consideration is to vary the electroporation parameters as well as explore other methods for making competent cells. Increasing the transformation efficiencies would lead to a higher number of clones which in turn allows for screening for the most stable integrants with the highest copy numbers which would inevitably lead to an increase in protein expression.

# 7. References


Aguilera, A., Chávez, S. and Malagón, F. (2000). Mitotic recombination in yeast: elements controlling its incidence. *Yeast,* 16, pp. 731-754.

Ahmad, M., Hirz, M., Pichler, H. and Schwab, H. (2014). Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Applied Microbiology and Biotechnology*, 98, pp. 5301-5317.

Aw., R. and Polizzi, K.H. (2013). Can too many copies spoil the broth? *Microbial Cell factories*, 12, pp.128.

Cereghino, J.L. and Cregg, J.M. (2000). Heterologous protein expression in the methylotrophic yeast *Pichia pastoris. FEMS Microbiology Reviews,* 24, pp. 45-66.

Cregg, J.M., Madden, K.R., Barringer, K.J., Thill, G.P. and Stillman, C.A. (1989). Functional characterization of the two alcohol oxidase genes from the yeast *Pichia pastoris*. *Molecular and Cell Biology,* 9, pp.1316-1323.

Cregg, J.M., Cerenghino, J.L., Shi., J. and Higgins, D.R. (2000). Recombinant protein expression in *Pichia pastoris. Molecular Biotechnology*, 16, pp. 23-52.

Cregg, J.M. (2007). Introduction: Distinctions between *Pichia pastoris* and other expression systems. In: J.M. Cregg ed., *Pichia Protocols*, 2nd ed. Humana Press Inc., pp. 1-7.

Cocolin, L., Bisson, L.F. and Mills, D.A. (2000). Direct profiling of the yeast dynamics in wine fermentations. *FEMS Microbiology Letters*, 189, pp. 81-87.

Couderc, R., and J. Baratti. (1980). Oxidation of methanol by the yeast Pichia pastoris. Purification and properties of alcohol oxidase. *Agricultural and Biological Chemistry*, 44 (10), pp. 2279-2289.

Curran, B.P.G. and Bugeja, V.C. (2009). The biotechnology and molecular biology of yeasts. In: J.M. Walker and R. Raply ed., *Molecular Biology and Biotechnology,* 5th ed. Cambridge: RSC Publishing, pp. 159-162.

Daly, R. and Hearn, T.W.M. (2005). expression of heterologous proteins in *Pichia pastoris*: A useful experimental tool in protein engineering and production. *Journal of Molecular Recognition*, 18, pp. 119-138.

Deng, H., O'Hagan, D. and Schaffrath, C. (2004). Fluorometabolite biosynthesis and the fluorinase from *Streptomyces cattleya. Natural Product Reports.* 21, pp.773-784.

Du, M., Battles, M.B. and Nett, J.H. (2012). A colour-based multi-copy integrant selection system for *Pichia pastoris* using the attenuated *ADE1* and *ADE2* genes as auxotrophic markers. *Bioengineered Bugs*, 3(1), pp. 32-37.

Ellis, S.B., Brust, P.F., Koutz, P.J., Waters, A.F., Harpold, M.M. and Gingeras, T.R. (1985). Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast *Pichia pastoris*. *Molecular and Cell* Biology, 5, pp. 1111-1121.

Furuya, Y., Kamlet, A.S. and Ritter, T. (2011). Catalyses for fluorination and trifluoromethylation. *Nature,* 473(7348), pp. 470-477.

Glick, B.S. and Fitzgerald, I. (2014). Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting. *Microbial Cell Factories*, 13, pp. 125.

Hanahan, D., Jessee, J. and Bloom, F.R. (1991). Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymology*, 20, pp. 63-113.

Harbron, S. (2009). Molecular cloning and protein expression. In: J.M. Walker and R. Raply ed., *Molecular Biology and Biotechnology,* 5th ed. Cambridge: RSC Publishing, pp. 20-23.

Invitrogen Corporation: PichiaPink™ expression system for high-level and largescale expression and secretion of bioactive recombinant proteins in *Pichia pastoris.* (2014). Publication number MAN0000717. [http:/tools.lifetechnologies.com/content/sfs/manuals/PichiaTM_expression_system-man.pdf]

Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. and Boyer, H.W. (1977). Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. *Science*, 198(4321), pp. 1056-1063.

Khow, O. and Suntrarachun, S. (2012). Strategies for the production of active eukaryotic proteins in bacterial expression system. *Asian Pacific Journal of Tropical Biomedicine*, 2(2), pp. 159-162.

Klimek-Ochab, M., Rodak-Brzezińska-Rodak, M., Żymańczyk-Duda, E., Lejezak, B. and Kafarski, P. (2011). Comparative study of fungal cell disruption-scope and limitations of the methods. *Folia Microbiologica*, 56, pp. 469-475.

Klinner, U. and Schäfer, B. (2004). Genetic aspects of targeted insertion mutagenesis in yeasts. *FEMS Microbiology Reviews*, 28, pp. 201-223.

Koch, C., Neumann, P., Valerius, O., Feussner, I. and Ficner, R. (2016). Crystal structure of alcohol oxidase from *Pichia pastoris*. *PLOS ONE,* 11(2), pp. 1-17.

Kupiec, T. (2004). Quality-control analytical methods: High performance liquid chromatography. *International Journal of Pharmaceutical Compounding*, 8(3), pp. 223-227.

Li, B., Cao, Y., Zhou, L., Liang, C. and Sun, F. (2011). A novel protein expression system-PichiaPink$^{TM}$- and a protocol for fast and efficient recombinant protein expression. *African Journal of Biotechnology*, 10(83), pp. 19464-19472.

Li, P., Anumanthan, A., Gao, X-G., Ilangovan, K., Suzara, V.V., Düzgüneş, N and Renugopalakrishnan, V. (2007). Expression of recombinant proteins in *Pichia pastoris*. *Applied Biochemistry and Biotechnology,* 142, pp. 105-124.

Lin-Cereghino, G.P., Godfrey, L., de la Cruz, B., Johnson, S., Khuongsathiene, S., Tolstorukov, I., Yan, M., Lin-Cerenghino, J., Veenhuis, M., Subramani, S. and Cregg, J.M (2006). Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in *Pichia pastoris. Molecular and Cell Biology,* 26(3), pp. 883-897.

Lin-Cereghino, G.P. and Lin-Cereghino, J. (2007). Vectors and strains for expression. In: J.M. Cregg ed., *Pichia Protocols*, 2$^{nd}$ ed. Humana Press Inc., pp. 14-18.

Lin-Cereghino, G.P., Stark, C.M., Kim, D., Chang, J., Shaheen, N., Poerwanto, H., Agari, K., Moua, P., Low, L.K., Tran, N., Huang, A.D., Nattestad, M., Oshiro, K.T., Chang, J.W., Chavan, A., Tsai, J.W. and Lin-Cereghino, J. (2013). The effect of the α-mating factor secretion signal mutations on the recombinant protein expression in *Pichia pastoris. Gene* 519(2), pp. 311-317.

Macauley-Patrick, S., Fazenda, M.L., McNeil, B. and Harvey, L.M. (2005). Heterologous protein production using the *Pichia pastoris* expression system. *Yeast*, 22, pp. 249-270.

O'Hagan, D. and Deng, H. (2015) Enzymatic fluorination and biotechnological developments of the fluorinase. *Chemical Reviews,* 115, pp. 634-649.

Orr-Weaver, T.L., Szostak, J.W. and Rothstein, R.J. (1981). Yeast transformation: A model system for the study of recombination. *Genetics,* 78(10), pp. 6354-6358.

Richey, M.G., Marek, E.T. , Schardl, C.L. and Smith, D.A. (1989). Transformation of filamentous fungi with plasmid DNA by electroporation. *Phytopathology*, 79, pp. 844-847.

Romanos, M.A, Scorer, C.A. and Clare, J.J. (1992). Foreign gene expression in yeast: a review. *Yeast,* 8, pp. 423-488.

Schaffrath, C., Deng, H. and O'Hagan, D. (2003). Isolation and characterisation of 5'fluorodeoxyadenosine synthase, a fluorination enzyme from *Streptomyces cattleya.* FEBS Letters, 547, pp. 111-114.

Schägger, H. (2006). Tricine-SDS-PAGE. *Nature Protocols,* 1(1), pp. 16-22.

Shigekawa, K. and Dower, W.J. (1988). Electroporation of eukaryotes and prokaryotes: A general approach to the introduction of macromolecules into cells. *Biotechniques,* 6, pp. 742-751.

Stathopulos, P.B., Scholz, G.A., Hwang, Y-M., Rumfeldt, J.A.O., Lepock, J.R. and Meiering, E.M. (2004). Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Science*, 13, pp. 3017-3027.

Sunga, A.J., Tolstorukov, I. and Cregg, J.M. (2008). Posttransformational vector amplification in the yeast *Pichia pastoris. FEMS Yeast Research,* 8, pp. 870-876.

Wang, Y.L., Hu, Y.S., Lei, Y., Lv, Y.Y., Wang, L. and Wei, H.W. (2015). Survey of Intracellular Protein Extraction Methods from *Pichia pastoris*. *World Journal of Engineering and Technology*, 3, pp. 1-6.

Wang, Z., Troilo, P.J., Wang, X., Griffiths II, T.G., Pacchione, S.J., Barnum, A.B., Harper, L.B., Pauley, C.J., Niu, Z., Denisova, L., Follmer, T.T., Rizzuto, G., Ciliberto, G., Fattori, E., Monica, N.L., Manam, S. and Ledwith, B.J. (2004). Detection of integration of plasmid DNA into host genomic DNA following intramuscular injection and electroporation. *Gene Therapy*, 11, pp. 711-721.

Wu. S. and Letchworth, G.J. (2004). High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. *Biotechniques,* 36, pp.152-154.

Zhang, X-J., Lai, T-B. and Kong, R. Y-C. (2012). Biology of fluoro-organic compounds. *Topics in Current Chemistry,* 308, pp. 365-404.

## A. Appendices

### A.1 Recipes for media and general reagents

**Ampicillin stocks (100 mg/ml)**

Dissolve in $dH_2O$/100% Ethanol and filter sterilise, store at -20 °C

**Agarose gel (0.8-1.0%)**

Dissolve 0.8-1.0 g agarose in 1X TAE buffer by boiling in a microwave. Once melted and cooled add 5 µl of ethidium bromide (10 µg/ml) per 100 ml of agarose.

Following horizontal gel electrophoresis at 90 V for approximately 1 hr, bands were visualised under UV light and photographed with the ChemiDoc$^{TM}$MP Imaging System (Biorad). Band sizes were determined using a Gene Ruler 1kb DNA Ladder (Thermoscientific) or HyperLadder$^{TM}$ 1kb DNA Marker (Bioline).

**LB- broth and agar**

1% Tryptone
0.5% Yeast Extract
0.5% NaCl
For LB-agar add 1.5 % bacteriological agar
Make the volume up with $dH_2O$ and sterilise by autoclaving for 20 mins at 121°C

**2XYT media**

2% Yeast extract
1.5% Tryptone
0.5% NaCl
Make the volume up with $dH_2O$ and sterilise by autoclaving for 20 mins at 121°C

**SOC Recovery media**

Add 2 g tryptone, 0.5 g yeast extract, 1 ml of 1 M NaCl and 0.25 ml of 1 M KCl and make up to 100 ml with $dH_2O$. Sterilise by autoclaving for 20 mins at 121°C. Add 1 ml of 2M magnesium stock (1 M $MgCl_2$, 1 M $MgSO_{24}$) and 1 ml of 2M glucose stock, each to a final concentration of 20 mM. Filter the media through a 0.2 µm filter to sterilise.

**20% Dextrose (10X)**

Add dextrose to $dH_2O$ and Sterilise by autoclaving for 20 mins at 121°C.

**YPD- broth and agar**

1% Yeast Extract

1% Peptone

For YPD-agar add 1.5 % agar

Dissolve yeast extract and peptone in $dH_2O$ and sterilise by autoclaving for 20 mins at 121°C. Add dextrose (autoclaved separately) once cooled.

**YPDS Recovery media**

1% Yeast Extract

2% Peptone

20% Dextrose

1 M Sorbitol

Dissolve yeast extract and peptone in $dH_2O$ and sterilise by autoclaving for 20 mins at 121°C. Add dextrose and sorbitol (autoclaved separately).

**BMGY/BMMY- Buffered complex media containing glycerol/methanol**

Stocks

**0.02% Biotin (500X)**

Dissolve in $dH_2O$ and filter sterilise, store at 4 °C.

**20% Dextrose (10X)**

Dissolve in $dH_2O$ and sterilise by autoclaving for 20 mins at 121°C, store at room temperature.

**5 % Methanol (10 X)**

Filter sterilise the solution and store at 4 °C

**10% Glycerol (10X)**

Dissolve in $dH_2O$ and sterilise by autoclaving for 20 mins at 121°C, store at room temperature.

**1 M Potassium phosphate buffer, pH 6.0**

Adjust pH with Phosphoric acid or potassium hydroxide and sterilise by autoclaving for 20 mins at 121°C, store at room temperature.

**13.4% Yeast Nitrogen Base with ammonium sulphate, without amino acids (10 X YNB)**

Dissolve in $dH_2O$ with heating and filter sterilise, store at 4 °C.

1% Yeast Extract

2% Peptone

100 mM potassium phosphate buffer pH 6.0

1.34% Yeast Nitrogen Base

0.00004% Biotin

1% Glycerol or 0.5% Methanol

Dissolve the yeast extract and peptone in $dH_2O$ and sterilise by autoclaving for 20 mins at 121°C. Add remaining solutions and sterilise by filtering through a 0.2 µM filter.

For BMMY add methanol instead of glycerol.

**Breaking buffer**

50 mM Sodium Phosphate buffer, pH 7.4

1 mM PMSF

1 mM EDTA

5% glycerol

Add the PMSF fresh just before use, store at 4 °C

**SDS gel stocks**

**Reducing sample buffer**

12% SDS (wt/vol)

6% Mercaptoethanol (vol/vol)

30% Glycerol (wt/vol)

0.05% Coomassie blue G-250 (wt/vol)

150 mM Tris/HCl, pH 7.0

**10 X Anode buffer**

1 M Tris (wt/vol)

0.225 M HCl (vol/vol)

Make up the volume with $dH_2O$ and adjust the pH to 8.9 with NaOH or HCl

**10 X Cathode buffer**

1 M Tris (wt/vol)

1 M Tricine (wt/vol)

1% SDS (wt/vol)

Make up the volume with $dH_2O$ and the pH should be ~8.25, the pH of the cathode buffer should not be adjusted.

**3 X Gel buffer**

3 M Tris (wt/vol)

1 M HCl (vol/vol)

0.3% SDS (wt/vol)

Make up the volume with $dH_2O$ and adjust the pH to 8.45 with NaOH or HCl

**Acrylamide-bisacrylamide (AB) stock solution**

Dissolve 48 g of acrylamide and 1.5 g of bisacrylamide in 100 ml of $dH_2O$. Store the solution at 7-10 °C.

**16% Separating gel**

Add 10 ml AB solution, 10 ml 3X gel buffer, 3 ml of glycerol to a final volume of 30 ml with $dH_2O$. The gel was polymerised by adding freshly prepared 150 µl 10% APS and 15 µl TEMED.

**16% Separating gel**

Add 1 ml AB solution, 3 ml 3X gel buffer to a final volume of 12 ml with $dH_2O$. The gel was polymerised by adding freshly prepared 90 µl 10% APS and 9 µl TEMED.

**Coomassie staining**

**Fixing solution**

50% Methanol (vol/vol), 10% Acetic acid (vol/vol) and 100 mM Ammonium acetate

**Staining solution**

0.025% Coomassie blue G-250 (wt/vol) in 10% Acetic acid (vol/vol)

**Destaining solution**

10% Acetic acid (vol/vol)

**A.2 Preparation of calcium chloride competent cells**

*E. coli* JM109 cells were made competent using the calcium chloride method by Hanahan, e*t al.,* (1991). A single colony was used to inoculate 10 ml of LB broth. The culture was grown overnight at 37°C with shaking at 250 rpm. One millilitre of the starter culture was used to inoculate 100 ml of LB broth and grown to an OD of 1.5, in approximately 3 hr. The cells were cooled on ice for 10 min and pelleted at 4100 rpm for 10 min at 4°C. The pellet was gently resuspended in 30 ml of $MgCl_2$-$CaCl_2$ solution, followed by centrifugation as above. The pellet was resuspended in 2 ml of 0.1 M $CaCl_2$-15 % glycerol solution. Competent cells were stored in aliquots at -80°C.

## A.3 Preparation of stocks

### PichiaPink™ strains

To obtain single colonies streak each strain on YPD agar plates and grow for 3-5 days at 25 °C. Inoculate 10 ml YPD with a single and grow for 16-20 hr at 25 °C with shaking at 300 rpm. Use the starter culture to inoculate 200 ml of YPD media to a final $OD_{600}$ of 0.2. Grow with shaking for 1-2 days at 25 °C for 2 days, until the $OD_{600}$ reaches 2-3. Harvest the cells by centrifugation at 1 500 x g for 5 min. Concentrate the cells by resuspending the pellet in YPD medium supplemented with 25 % glycerol to a final $OD_{600}$ reaches 50-100. Aliquot the cells in cryovials, snap freeze with liquid nitrogen and store at -80 °C.

### PichiaPink™ vectors- pPinkα-HC

Propagate and maintain the vector and expression constructs in an *E. coli* strain that is recombination (*rec*A*)* and endonuclease (*end*A*)* deficient. Transform *E. coli* JM109 cells with the pPinkα-HC vector, select transformants on LB-agar supplemented with 100 µg/ml ampicillin. Select a single positive colony and grow in 2 ml of 2XYT media supplemented with 100 µg/ml ampicillin. Grow the cells until the stationary phase is reached; add glycerol to a final concentration of 15 %. Aliquot cells into cryovials, snap freeze with liquid nitrogen and store at -80 °C.

```
ORIGIN
   1 atggctgcca acagcacacg tcgccccatc atcgcgttca tgagcgacct ggggaccacg
  61 gacgactccg tcgcccagtg caaggggctc atgtacagca tctgcccgga cgtcacggtg
 121 gtggacgtct gccactcgat gacgccctgg gacgtcgagg agggcgcccg ctacatcgtg
 181 gaccttccgc gcttcttccc cgagggaacg gtcttcgcca ccaccaccta tccggcgacc
 241 ggcaccacca cccgctcggt ggcggtgcgc atcaagcagg ccgccaaggg cggtgcccgc
 301 ggccagtggg cgggctcggg ggccggcttc gagcgcgccg agggctcgta catctacatc
 361 gcgcccaaca acgggctgct gaccaccgtg ctggaggagc acggctacct ggaggcgtac
 421 gaggtcacct cgccgaaggt catccccgag cagcccgaac cgaccttcta cagccgggag
 481 atggtggcca tcccctccgc gcacctggcc gccggcttcc cgctgtccga ggtcggccgt
 541 ccgctggagg accacgagat cgtccgcttc aaccgcccgg ccgtcgagca ggacggggag
 601 gcgctggtgg gcgtggtctc cgccatcgac cacccgttcg gcaacgtgtg gaccaacatc
 661 caccgcaccg acctggagaa ggcgggcatc ggctacggcg cccggctgcg gctgacgctg
 721 gacggcgtgc tgccgttcga ggcgccgctg accccgacgt tcgccgacgc cggtgagatc
 781 ggcaacatcg ccatctacct caacagccgc ggttacctgt ccatcgcgcg caacgcggcc
 841 agcctcgcct acccgtacca cctcaaggag ggcatgtccg cccgggtcga ggcccgctga

translation="MAANSTRRPIIAFMSDLGTTDDSVAQCKGLMYSICPDVTVVDVC

HSMTPWDVEEGARYIVDLPRFFPEGTVFATTTYPATGTTTRSVAVRIKQAAKGGARGQ

WAGSGAGFERAEGSYIYIAPNNGLLTTVLEEHGYLEAYEVTSPKVIPEQPEPTFYSRE

MVAIPSAHLAAGFPLSEVGRPLEDHEIVRFNRPAVEQDGEALVGVVSAIDHPFGNVWT

NIHRTDLEKAGIGYGARLRLTLDGVLPFEAPLTPTFADAGEIGNIAIYLNSRGYLSIA

RNAASLAYPYHLKEGMSARVEAR"
```

**Figure A.1** The 900 bp nucleotide sequence of the *fLA* gene and the corresponding 299 amino acid sequence of the fluorinase from *Streptomyces cattleya,* (gene accession number: AJ581748).

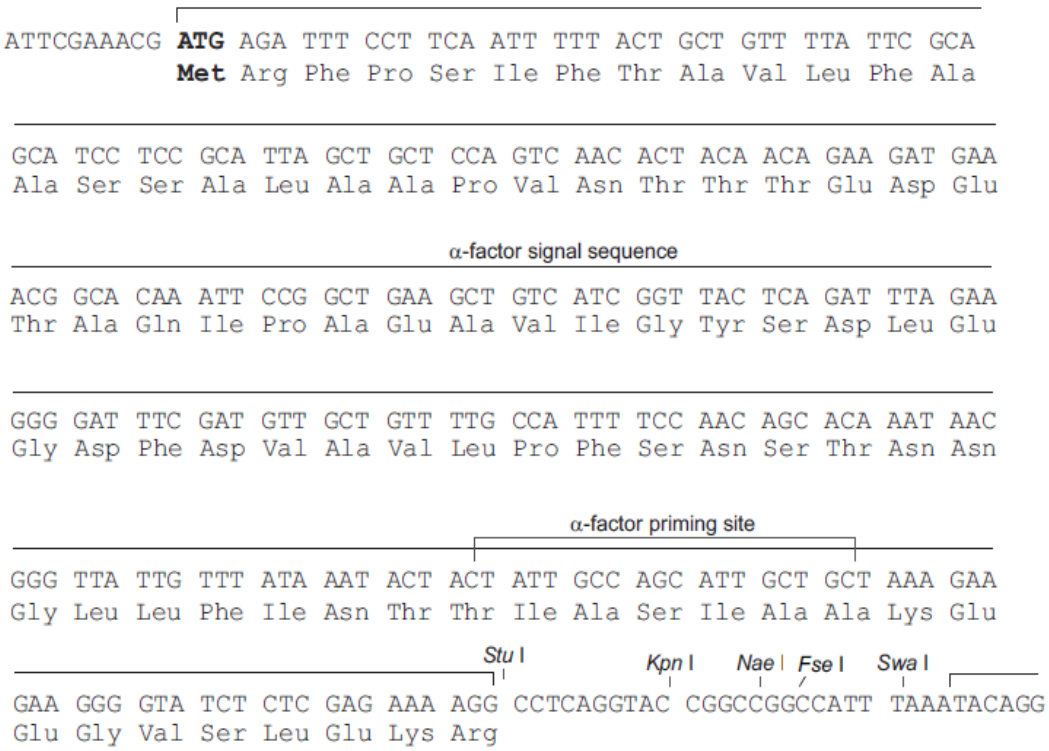



**Figure A.2:** Sequence of the alpha mating factor secretion signal, the *Kex2* cleavage site is located after the final Arg amino acid PichiaPink™ pPinkα-HC vectors. The Glu-Ala-Glu-Ala *Ste13* cleavage site after the *Kex2* site not available (adapted from the Invitrogen)

### A.4 Calculating cell concentration

$$Concentration\ of\ cells = \frac{(average\ number\ of\ cells\ per\ square)X\ df}{volume\ of\ square\ counted(mm^3)}$$

$$cells/ml = cells/mm^3 X 10^3$$

### A.5 Calculation the breaking efficiency

The cell breaking rate was calculated as the percentage of broken cells after disruption

Nb= number off cells/ml before disruption

Na= number of cells/ml after disruption

$$cell\ breaking\ rate(\%) = \frac{Nb - Na}{Nb} X\ 100$$