A combinatorial assembly strategy to optimise biodegradation pathway performance in yeast

Gurdamanjit Singh

A thesis submitted for the degree of
Doctor of Philosophy

School of life Sciences
University of Warwick

Supervisor: Prof. John E.G. McCarthy

March 2021

Table of Contents

List of Figures

List of Tables

Acknowledgements

Firstly, I would like to give a special thanks to Prof. John McCarthy for giving me the opportunity to work in his lab and acquire the skills I have. I appreciate all the advice that John has provided me regarding my PhD as well as career path.

I would also like to say thanks to Dr Maja Firczuk, Dr Estelle Dacheux for their guidance and support in this project. I am grateful to my colleagues Tailise, Sandie, Alan, Mallory and Byron for their friendship and assistance in the lab. I would like to thank Dr. Sarah Bennett, Dr. Corinne Hanlon and Dr. Graham Jones for their assistance in WISB related matters as well as their friendship. I would like to say thank you to Sophie Piquerez for her guidance, inspiration and advice towards my scientific career as well as her friendship. A special thanks to Dr. Fabio Rodrigues for his assistance and friendship throughout the PhD.

Thank you to my collaborators Ingenza, Dr. Ian Fotheringham and Dr. Stephen McColm for their continued support and assistance throughout the PhD as well as the opportunity for a brilliant placement.

I would like to give special thanks to Dr. Chrysi Sergaki for her continued support and advice throughout the PhD. I will always cherish and appreciate the time we spent together throughout this PhD that helped me to shape the person I am today.

I would like to thank my mother Navjit Kaur as she has supported me and continues to support me in my life unconditionally. Every time I fall, she is there to support me and tell me to get back up and for that, I deeply appreciate her. Please know that without you, none of this would have been possible. To my sister Jaz Kaur, I say thank you for your support and belief in me and always being there for me when I needed it, big or small. Thank you to my grandparents Geeta Devi and Lekh Raj for their love and continued support. To my father Charanjit Singh, I thank you for your hard work in life to ensure a better life for your family. Mya, thank you for pushing me when I needed it.

Finally, I would like to thank the University of Warwick, MIBTP, BBSRC, and WISB.

Declarations

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. The work presented (including data generated and data analysis) was carried out by the author unless otherwise stated.

Abstract

Combinatorial assembly methods, enhanced by automation, aim to facilitate the optimisation of a wide range of biosynthetic and biodegradative pathways to obtain a maximal product yield by varying the rate of expression of genes. The aim of this thesis was to develop a combinatorial strategy that allows the variation of expression of metabolic pathways in yeast. Initially, two yeast species, *Pichia pastoris* and *Saccharomyces cerevisiae,* were selected and compared to determine what the better suited host would be for metabolic pathway expression. The secretion of a fluorescent reporter (ymNeonGreen) was used as marker and *P. pastoris* was found to be a more suitable host for extracellular secretion of recombinant proteins. This PhD work was an iCASE collaboration with the biologics company Ingenza, Edinburgh, Scotland, which provided seven novel cellulase sequences that could be optimised for efficient cellulose degradation. The individual heterologous extracellular secretion of these cellulases was carried out in *P. pastoris* and the cellulases that were not expressed were not selected for further optimisation. To be able to co-express these cellulases and create a cellulose degradation pathway in *P. pastoris,* a plasmid utilising marker recycling via *cre*-recombinases was selected. However, unwanted recombination of this plasmid was occurring in *E. coli* prior to *P. pastoris* transformation, preventing high throughput use of this plasmid. Therefore, modifications were made to the plasmid that allowed the development of a high-throughput integration method. A combinatorial strategy was implemented in this method with the use of promoters with different expression strength. This strategy was used to create a cellulose degrading *P. pastoris* strain. The same method was utilised for the integration of ferulic acid degradation pathway in *P. pastoris,* validating this method as an efficient high-throughput method to integrate and express metabolic pathway products in *P. pastoris*.

Abbreviations

5'UTR	5' untranslated region
a.a.	Amino acid
Abs	Absorbance
AKR	Aldo-keto reductase
Amp	Ampicillin
AMP	Adenosine Monophosphate
AOX	Alcohol oxidase
AOX1	Alcohol oxidase 1
ATP	Adenosine Triphosphate
ATP	Adenosine Triphosphate
BCMY	Buffered cellulose Complex Medium
BG	Beta-glucosidase
BMGY	Buffered Glycerol Complex Medium
BMMY	Buffered Methanol Complex Medium
B-ME	β-mercaptoethanol
bp	Base Pairs
CA	Coniferyl aldehyde
CaCl2	Calcium chloride
CAR	Carboxylic acid reductase
CBH	Cellobiohydrolase
CBM	Carbohydrate binding module
CD	Catalytic domain
CE1	Carbohydrate esterase family 1
Cg-AKR-1	*Coptotermes gestroi* aldo-keto reductase
Col	Coniferyl alcohol
CRISPR	clustered regularly interspaced short palindromic repeats
CV	Column Volume
Da	Dalton
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
ddH2O	Double-distilled water

dGTP	Deoxyguanosine triphosphate
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
E	Elution
E. coli	Escherichia coli
EC	Enzyme comission
EMA	European medicine agency
EnG	Endoglucanase
ER	Endoplasmic reticulum
ExG	Exo-glucanase
FA	Ferulic acid
FACS	Fluorescence-Activated Cell Sorting
FDA	American food and drug administration
FP	Fluorescent proteins
g	Gram
GAP	Glyceraldehyde 3-phosphate
GFP	Green Fluorescent Protein
GOI	gene of interest
GRAS	Generally regarded as safe
HIS4	Histidine biosynthesis trifunctional protein
HisCar5	Nocadia iowensis carboxylic acid reductase
HPLC	High Performance Liquid Chromatography
ID	identification card
IMAC	Metal ion affinity chromatography
Kan	Kanamycin
KCl	Potassium chloride
kDa	Kilo dalton
KOH	Potassium hydroxide
l	Litre
lacO	Lac operon operator
LB	Lysogeny Broth
LC-MS	Liquid chromatography–mass spectrometry

lem	Emission wavelength
lex	Excitation wavelength
LiAc	Lithium Acetate
loxp	locus of X-over P1
M	Molar
MAT	Mating alpha factor secretion signal
$MgCl_2$	Magnesium chloride
mg	Milligram
mins	Minutes
mM	Millimolar
mRNA	Messenger Ribonucleic Acid
NaCl	Sodium chloride
NAD	Nicotinamide adenine dinucleotide
NADPH	Nicotinamide adenine dinucleotide phosphate
ng	Nanogram
nm	Nanometre
nt	nucleotide
oC	Degree Celsius
OD595	Optical Density at 595 nm
OD595	Optical Density at 595 nm
OD600	Optical Density at 600 nm
ORF	Open-Reading-Frame
P. pastoris	*Pichia pastoris*
PCP-domain	phosphopantetheine carrier domain
PCR	Polymerase chain reaction
PEG8000	Polyethylene glycol 8000
PGAP	Promoter Glyceraldehyde 3-phosphate
pH	log10[H^+]
PI	isoelectric point
PNP	p-nitrophenyl
PNPG	p-nitrophenyl-β-glucopyranoside
POP	Pre-Ost1-Pro

PPTase	Phosphopantetheinyl transferase
RE	Restriction Enzyme
RFP	Red Fluorescent Protein
RFU	Relative fluoresence units
RNA	Ribonucleic acid
rt	Retention time
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate - Polyacrylamide Gel
SOB	Super Optimal broth
SOC	Super Optimal broth with Catabolite repression
TAE	Tris-acetate-EDTA
TE	Tris-EDTA buffer
TFP	Teal Fluorescent Protein
TLC	Thin-layered chromatography
Tris-HCl	Tris(hydroxymethyl)aminomethane Hydrochloride
Tris.HCl	Tris(hydroxymethyl)aminomethane Hydrochloride
URA3	Orotidine 5'-phosphate decarboxylase
UV	Ultraviolet
V	Volts
v/v	Volume by volume
w/v	Weight by volume
w/w	Weight by weight
WAX	Wheat arabinose-xylans
yEGFP	Yeast-enhanced green fluorescent protein
YEP	Yeast extract Peptone
YEPD	Yeast extract Peptone Dextrose
YEPDA	Yeast extract Peptone dextrose Agar
YEPG	Yeast extract Peptone Glycerol
YFP	Yellow Fluorescent Protein
ymNG	ymNeonGreen
µg	Microgram
µL	Micro litre

µM	Micro molar
µm	Micrometre
λ_{ex}	Excitation wavelength
λ_{em}	Emission wavelength

Chapter 1 - Introduction

1.1 Synthetic Biology

Synthetic Biology is defined as 'the design and construction of new biological parts, devices, and systems, and the re-design of existing, natural biological systems for useful purposes' (Roberts et al., 2013). It is an emerging interdisciplinary field that aims to study the manipulation of biology through the combination of scientific and engineering approaches. This field has experienced rapid growth and innovation addressing challenges such as the global food crisis and the need for renewable energy (Shapira et al., 2017). Synthetic biology can also improve gene-based applications as well as to provide a great toolset for advancing industrial science (Church, 2014). Major advancements in the last 50 years in this field have allowed a way of designing and constructing new or existing biological elements (proteins, metabolic pathways, or gene circuits) to meet specific criteria (Serrano, 2007).

Synthetic biology roots can be traced to a landmark publication by Monod & Jacob, (1961) that led to the identification of regulatory circuits involved in the response of a cell to its environment (Cameron et al., 2014). Great leaps in this field were made in 1970 with the discovery and development of the first type II restriction enzyme which cleave DNA at specific recognition sequence sites (Smith and Welcox, 1970). The utilisation of this recombinant DNA technology allowed the cloning of genes from one organism into another (Morrow et al., 1974). The use of this technology led to the first patent being awarded in 1981 for a genetically engineered *Pseudomonas* bacterial strain that could degrade crude oil and as a result outlining the use of genetically modified organisms for metabolite breakdown/production (Chakrabarty, 2010). Further developments in synthetic DNA synthesis and thus gene synthesis (Caruthers, 1985) aided in the mass scale production of target drugs in the biopharmaceutical industry (Aharonowitz and Cohen, 1981). The most notable early achievement was the creation and commercialism of synthetic insulin (Ladisch and Kohlmann, 1992). In addition, the polymerase chain reaction (PCR) allowed the amplification of millions of copies of target genes using small oligonucleotides. This provided an excellent tool for fast and cost-efficient recombinant gene expression (Mullis et al., 1986).

A rise in synthetic biology research was observed worldwide, going from approximately 170 publications annually between 2000 and 2005 to approximately 1200 annually in 2015 (Shapira et al., 2017). The global value of synthetic biology has been projected to reach 14 billion USD by the year 2026 (Freemont, 2019). The biopharmaceutical sector observed around 1,357 products approved by the American food and drug administration (FDA) and up to 2,016 products approved by the European medicine agency (EMA) (Jozala et al., 2016). It has been projected that in the next 10 years, up to 50% of all drugs developed will be obtained from biopharmaceuticals (Jozala et al., 2016). The use of synthetic biology has been widely adopted by many biologics companies as a result of huge financial investments and incentives (Wang and Zhang, 2019).

One specific way of adopting synthetic biology is by engineering metabolic pathways to achieve the biological manufacturing of a desired chemical or pharmaceutical product (Stephanopoulos, 2012). The engineering of metabolic pathways can be a complex procedure as they require the synergistic action of multiple enzymes to efficiently degrade a substrate into the desired product. Due to the current climate challenges, one substrate of interest is lignocellulosic biomass as it is an abundant (with an estimated 181.5 billion tonnes produced per year [Dahmen et al., 2019]), relatively inexpensive ($24 to $121 per tonne depending on the crop [Mapemba and Epplin, 2004]), renewable resource that can be degraded to yield many high-value products such as glucose and coniferyl alcohol (Arevalo-Gallegos et al., 2017; Tramontina et al., 2020). Scaling up the biosynthesis of these products through the cellulose degradation pathway and the coniferyl alcohol pathway can be challenging. In order to achieve this, suitable enzymes are sought from various microorganisms and are produced in suitable industrial hosts as recombinant proteins using recombinant DNA technology. The production of recombinant proteins is often maximised to ensure high levels of proteins in the pathways. In order to achieve high yields of recombinant protein production for industrial use, the choice of substrate, host organism and combinatorial strategy are crucial factors requiring thorough consideration.

1.1.1 Gene expression and recombinant protein production

Cellular protein synthesis is composed of many processes taking place whereby DNA is copied to mRNA (transcription) followed by the mRNA being decoded to a specific amino acid sequence to form proteins (translation) (Kapp and Lorsch, 2004). RNA polymerases are multi-subunit enzymes carrying out transcription, producing RNA from DNA. There are three types of RNA polymerases. RNA polymerase I involved I the synthesis of most ribosomal NRA, RNA polymerase II which is involved in the synthesis of all protein coding mRNAs as well as miRNAs and RNA polymerase III which is involved in the synthesis of tRNAs. RNA polymerase II is recruited at the core promoter region of genes by the pre-initiation complex (PIC) followed by activation of RNA polymerase II by DNA-binding transcription factors (Hori and Carey, 1994). Some eukrayotic promoters contain a TATA box consisting of 'TATTAA' DNA sequence (upstream of transcription start site) to allow the binding of the tata box binding protein for initiating the transcription to the promoter site. This sequence can affect the rate at which transcription is carried out. They may also contain other regulatory sequence to increase or decrease transcription (Watson et al., 2014). These transcriptional activators result in the initiation of transcription followed by elongation (via elongation factors). Transcription is terminated by the recognition of the poly(A) signal in the transcribed RNA via cleavage and polyadenylation specificity factors and cleavage simulation factors. This results in the recruitment of other proteins that then carry out polyadenylation (adding adenines to the cleaved 3' RNA end) and RNA cleavage. The termination may occur as a result of conformational changes to the elongation complex and/or via the degradation of RNA still attached to RNA polymerase II via an exonuclease (Hirose and Ohkuma, 2007). RNA splicing is carried out on the pre-messenger RNA to produce the mRNA which is transported to the ribosome for protein synthesis (translation). The mRNA is a single-stranded molecule containing a 5' cap, 5' UTR, a coding region, a 3' UTR and a poly(A) tail. The 5' cap is formed of a 7-methylgunylate bound to the 5' end of the mRNA and is involved in nuclear export, stability, and promotion of translation by assisting in 40s ribosomal subunit binding to the mRNA via eukaryotic initiation factors (Araujo et al., 2012). This is followed by the 5' UTR which may contain secondary structures that can impede the 40s ribosomal subunit from scanning,

affecting the rate of translation. However, there are some eukaryotic initiation factors that are involved in identifying and unwinding any secondary structures in this region to allow for ribosomal units to scan for the initiation codon without being impeded. Once the start codon (AUG) has been identified by the ribosomal unit, the codon anti-codon base pairing can begin. The methionine charged initiator tRNA is brought to the 40s ribosomal subunit occupying the P-site leading to the dissociation of translation factors and the association of the 60s ribosomal subunit (Szostak and Gebauer). Elongation is then carried out whereby the A-site can receive the next aminoacyl-tRNA based on the codon sequence forming a peptide bond followed by shifting the mRNA by one codon. The preference of one codon over another to encode for the same amino acid (codon usage bias) by a eukaryotic organism can affect the rate at which translation may be carried out. Therefore, to improve heterologous gene expression, the open reading frame is often optimised to contain the more frequently used codons for that host organism (Hanson and Coller, 2017). Elongation is terminated when the ribosomal unit encounters a stop codon (usually UAG, UAA or UGA). The elongation release factors trigger the hydrolysis of the polypeptidyl-tRNA, resulting in the release of the completed protein product (Jackson et al., 2012). The stop codon is followed by the 3' UTR which can aid in regulation of translation. They may also contain AU-rich regions that regulate stability (Mayr, 2019). The poly(A) tail may also play a role in promoting recycling of the terminating ribosomes for another round of translation (Szostak and Gebauer, 2012).

Obtaining desired protein yields from natural sources can be a challenging task. Recombinant protein production is a well-known strategy to generate high expression yields of gene sequences derived from one species in another well-studied host, for example *Escherichia coli*. This has been of particular value in the biotechnology sector as a means of producing proteins for various purposes (García-Fruitós, 2015). Heterologous protein production requires procedures involving the selection of a suitable host, as well as codon optimisation, gene regulation, promoters, terminators, and gene integration.

Both prokaryotic and eukaryotic organisms are used as hosts, with both having their advantages and disadvantages (Elena et al., 2014). The prokaryotes *Bacillus subtilis*

and *E. coli* have been used for heterologous protein production. *B. subtilis* is a gram-positive bacterium that has been used widely to produce heterologous proteins. It has been favoured as it contains secretory pathway and a single cell membrane allowing for the secretion of heterologous proteins. It's wide gain of use has resulted in genetic tool manipulation with CRISPR-cas9 systems being recently developed. Zhang et al., (2017) achieved up to 2.5 mg/mL extracellular expression of a β-cyclodextrin glycosyltransferase. However, it has been reported that *B. subtilis* may result in reduced protein secretion due to secretory protein degradation and incorrect folding Westers et al., 2004. The prokaryote *E. coli* has been a commonly used expression system for many reasons, including fast growth rate (Sezonov et al., 2007), high growth density (Shiloach and Fass, 2005), inexpensive materials and quick transformation using exogenous DNA (Pope and Kent, 1996). Protein production is a relatively quick and efficient process, with cells ready to be lysed overnight after appropriate induction of cultures. However, protein production depends on the complexity of the protein in question. For example, if post-translational modifications such as protein glycosylation and disulphide bonds are necessary, then a eukaryotic organism may be capable of performing the correct type of protein glycosylation. Furthermore, *E. coli* lacks an efficient secretory pathway, adding to costs downstream, whereas other organisms, including some yeasts, are capable of secreting proteins (Rosano and Ceccarelli, 2014).

Yeasts are widely used as host organisms for biotechnological purposes due to the large database of pre-existing research data available, well known simple methods for gene expression and in some cases, the ability to secrete the protein, eliminating the necessity for cell lysis. This is an attractive option for the biotechnology field, cutting down on costs, and improving protein efficacy and yields (García-Fruitós, 2015). Yeasts are also less expensive to use for recombinant protein production compared to mammalian or insect cells (Bill, 2012) as these cells have slow cell growth, continuous CO_2 supply requirement and expensive transfection reagents (Rosano and Ceccarelli, 2014). Two yeasts in particular, *Saccharomyces cerevisiae* (5-10 µm in diameter) and *Pichia pastoris* (also known as *Komagataella phaffi*) (4-6 µm in diameter), have been commonly used in the biotechnology sector for recombinant protein production (Laluce et al., 2012; Zahrl et al., 2017). In the 1990s, *P. pastoris* was recognised as a

possible protein expression system when it was first used in industry for the production of hydroxynitrile lyase (Hasslacher et al., 1997). By 2010 it is known to have expressed more than 500 different recombinant proteins, and has been successfully used for the production of more than 100 million units of Hepatitis B vaccine and human insulin (Sreekrishna, 2010). As for *S. cerevisiae*, its use has been dated back to ancient times for wine production (Mortimer, 2000). *S. cerevisiae* has been widely used in the biopharmaceutical industry for recombinant protein production. It is associated with producing up to 20% of the global market of recombinant pharmaceuticals that are protein-based (Martínez et al., 2012) such as vaccines and insulin (Demain and Vaishnav, 2009). Both yeasts contain a complete secretory pathway that allows the recombinant protein to be secreted into the medium.

1.1.2 Secretory pathway

The secretory pathway in yeasts is a complex process involving hundreds of proteins (Delic et al., 2013). A secretion signal sequence consisting of a pre (16-amino acids) and pro (66-amino acids) region is attached to protein sequences that are required to be secreted. The intracellular secretory pathway carries out secretion of proteins out of the cell. The secretion signal peptide is recognised and bound by a signal recognition particle and is co- or post-translationally translocated into the endoplasmic reticulum lumen through the *sec*61 translocon. The ER-resident chaperone binding protein (BiP) binds the nascent polypeptides and the glycoproteins are bound by ER chaperone calnexin (Idris et al., 2010). Proteins undergo strict folding including covalent modifications of signal sequence processing, N-glycosylation, di-sulphide bond formations as well as degradation and sorting (Sheng et al., 2017). The pre-region of the signal peptide is cleaved and proteins that are correctly folded are exported from the endoplasmic reticulum lumen to the Golgi apparatus via COPII-coated vesicles (Jensen and Schekman, 2011). Here, the protein undergoes further glycosylation. However, misfolded, or aggregated proteins are bound by the BiP binding protein complex for redirection to the cytosol. As a result of the prolonged binding of BiP complex, the unfolded protein sequence response (UPR) is initiated, leading to the activation of the ER-associated protein degradation where the misfolded BiP complex bound protein undergoes degradation (Yoshida, 2007). Robinson et al., (1996) reported the importance of BiP complex by observing reduced secretion of

proteins upon its knockout in *S. cerevisiae*. The correctly folded proteins that are translocated into the trans-Golgi network have the pro region of the signal peptide cleaved by the protein Kex2p. The protein is then packed into a secretory vesicle, which fuses to the cell membrane to complete secretion out of the cell (Hou, K. Tyo, et al., 2012).

The signal sequence most commonly used in heterologous recombinant protein secretion in yeast is the *S. cerevisiae* pre-pro-α-factor mating pheromone signal peptide. The α-factor mating pheromone is a peptide secreted by α-mating cells to conjugate with a-mating cells to allow the formation of α/a diploid cells. The pre-pro region ensures correct translocation of the attached protein and its subsequent secretion outside the cell (Lin-Cereghino et al., 2013a). Chaudhuri et al. (1992) highlighted the importance of the pre-pro region in heterologous protein secretion by attaching the pre-pro region to human insulin-like growth factor 1 and observing the protein being secreted into the extracellular medium.

Improvements have been made in the secretion signal peptide to increase its performance. A study by G. P. Lin-Cereghino et al. (2013) implemented site-directed mutagenesis on the pre-pro region of α-mating factor secretion signal to improve the secretion of heterologous proteins. The pro region is thought to be important in allowing proper folding of proteins, however, this can occur by slowing the rate of transport. They targeted these amino acids in the pro region and carried out various deletions. The deletions in the α-factor mating sequence at amino acids 30-43 (*MAT*Δ30-43) and at amino acids 57-70 (*MAT*Δ57-70) resulted in a 123% and 159% increase in the secretion of horseradish peroxidase and lipase respectively, compared to the wild type α-mating factor secretion signal. Another study by Barrero et al. (2018) observed that if certain proteins can fold in the yeast cytosol, they may not cross the ER membrane and so may not be secreted, or be inefficiently secreted. To overcome this, they added an Ost1 signal sequence in between the pre and pro region of α-factor secretion signal. The Ost1 signal would assist in directing co-translational translocation across the ER membrane. They attached this newly modified secretion signal to a fluorescent protein and a BTL2 lipase and reported an increase in total secretion of them of 20-fold and 10-fold respectively. Secretion of recombinant

proteins is highly desirable in the biotechnology sector. As *P. pastoris* is known to be a good secretor of proteins, it is considered to be a suitable chassis organism for future bioproduction (Gasser and Mattanovich, 2018).

1.1.3 *Pichia pastoris* - A brief history

P. pastoris comes under the genus '*Komagataella*', belonging to the yeast family of Saccharomycetaceae, known for their elliptical, oblong acuminate or spherical cells. *P. pastoris* is a methylotrophic yeast, gaining recognition as a stable, easy to manipulate and reliable cell factory for a wide variety of recombinant protein production. Emil Hansen initially established the *Pichia* genus in 1904, with Alexandre Guilliermond identifying a *Pichia* sample from a french chestnut tree, classing it as *Zygosaccharomyces pastori* in 1920 (Ellis, 1921). Then, in the 1950s, Herman Phaff discovered many yeast isolates in the Yosemite region (California USA), resulting in the discovery of a modern-day known variant of *Pichia pastoris*. Later, they were both recognised as a single species occurring from the evolution from two distinct environments and they were since known as *Pichia pastoris* (Barnett et al., 2000).

In the late 1960s, Koichi Ogata discovered certain yeasts were capable of utilizing methanol as a carbon source (Ogata et al., 1969). Phillips Petroleum adopted this approach and decided to use *P. pastoris* for the bioconversion of methane into animal feedstock as single-cell protein. However, this animal feedstock was not economically viable due to the lower prices of soy protein production as animal feed, so they began using *P. pastoris* as a host for recombinant protein expression. With the help of Salk Institute Biotechnology/Industrial Associate Inc. (SIBIA) and from the discovery of the *AOX1* gene, the strong, tightly regulated methanol-induced promoter P_{AOX1} was identified. Furthermore, the genetic tools that were available for *S. cerevisiae* and *E. coli*, such as vectors, transformation protocols and integration knowledge, were ready to be used with modifications in order to manipulate *P. pastoris*. Phillips Petroleum decided to license out the use of *P. pastoris* as a protein expression system, and thus *P. pastoris* gained recognition as a powerhouse in the recombinant protein production system (Sreekrishna, 2010).

1.1.4 *P. pastoris* as a protein expression system

P. pastoris gained the status of a 'GRAS' (Generally Regarded As Safe) organism by the food and drug administration in 2006 and it was recognised as a good choice for recombinant protein production (Werten et al., 2019). As a result, a mass of research followed soon thereafter (Sreekrishna, 2010). The success of *P. pastoris* in recombinant protein production is a result of the combination of many aspects from tightly regulated strong promoters to high cell density. *Pichia* is initially grown under glucose/glycerol carbon source to a high density (for a period of 24 hours up to 2 days) and then the media is switched to one that contains methanol as a carbon source and expression inducer (Lin-Cereghino et al., 2005). In one example, l-glutamate oxidase was expressed in *P. pastoris*, and using exponential methanol feeding, a maximum wet cell weight of 420g/L was reached after 84 hours of cultivation in fermentation (YaPing et al., 2017) highlighting the potential of high recombinant protein production.

As *P. pastoris* is a eukaryotic organism, it can carry out post-translational modifications utilising an extensive set of translation initiation factors of recombinant proteins, compared to prokaryotes. Furthermore, compared to *S. cerevisiae,* hyper-glycosylation in *P. pastoris* occurs less often and the glycosylation can be less immunogenic (Tran et al., 2017; Hirose and Ohkuma, 2007). Hyper-glycosylation may be undesired as it can vary the size of the recombinant proteins obtained due to the addition of excessive glycans, possibly affecting functionality and immunogenicity (Tran et al., 2017). Another advantage of using *P. pastoris* is its ability to grow in a range of pH (3-7) without a significant loss in cell density. It offers a unique scenario, as a low pH reduces the growth of contaminant micro-organisms as well as general proteolytic activity. Furthermore, this ability allows for a custom pH to be set depending on where the protein stability is (in extracellular conditions) (Cereghino and Cregg, 2000). However, protease deficient *P. pastoris* strains have now been developed to improve recombinant protein levels to reduce extracellular recombinant protein degradation (Ahmad et al., 2019).

1.1.5 *P. pastoris* methanol utilisation

Methanol metabolism is a complex process that requires the actions of multiple enzymes (Figure 1.1). *P. pastoris* is well known for its methanol utilisation, therefore, strains were developed that could be selected using the rate at which methanol is metabolised (Romanos et al., 1992). Under normal conditions, catabolism of methanol is carried out and methanol is oxidised to formaldehyde and hydrogen peroxide (toxic to cell) by the enzyme alcohol oxidase (AOX) in the peroxisome. The peroxisomes contain the enzymes AOX and catalase, the latter degrading the hydrogen peroxide to oxygen and water. Some of the formaldehyde produced remains in the peroxisome, where via anabolic process it is catalysed to glyceraldehyde3-phosphate and dihydroxyacetone by the enzyme dihydroxyacetone synthase (DHAS). The glyceraldehyde 3-phosphate and dihydroxyacetone products exit the peroxisome and enter a cyclic pathway resulting in their transformation to raw materials for becoming constituents of cellular components. Glyceraldehyde 3-phosphate enters the catabolic glycolysis pathway and is converted into D-glycerate 1,3-bisphosphate. This is done by the enzyme glyceraldehyde 3-phosphate dehydrogenase under the constitutive control of its promoter *GAP* (Krainer et al., 2012). Due to their strength, promoters from the methanol metabolism and resulting pathways were targeted for the expression of heterologous proteins (Wu et al., 2003).

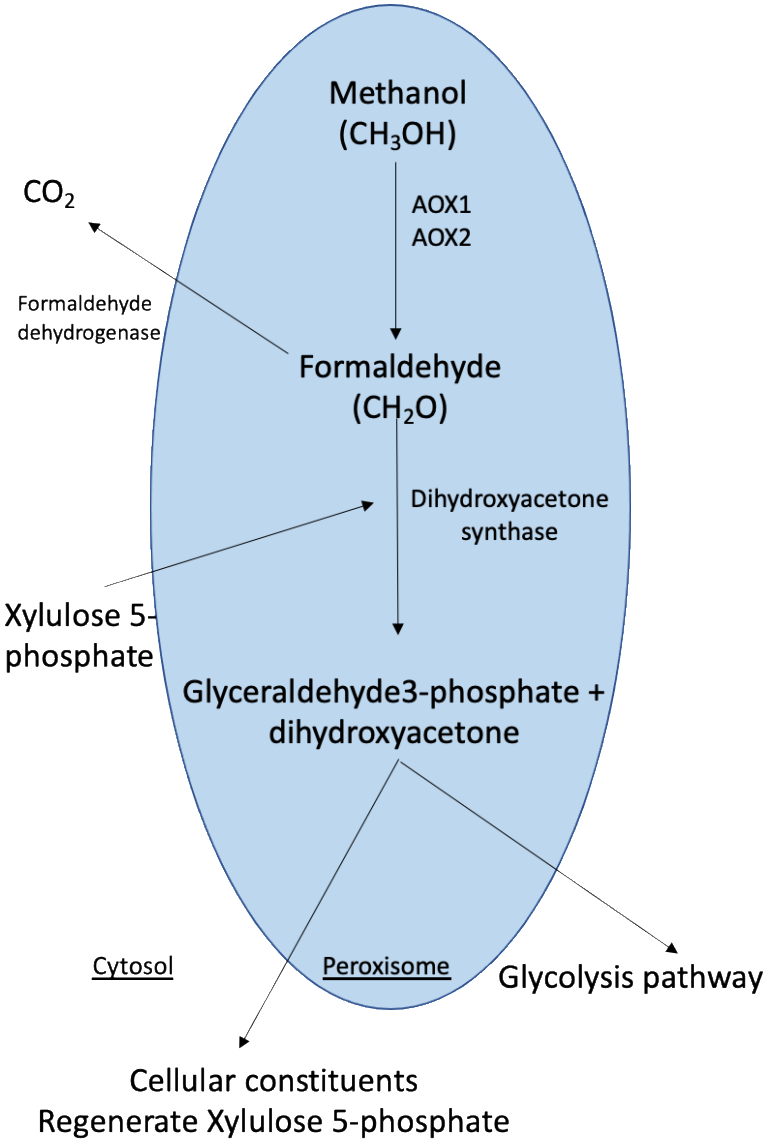

Figure 1.1: Methanol metabolism pathway in Pichia pastoris (adapted from Ali et al., 2019).

1.1.6 *P. pastoris* promoters, selection markers and homologous recombination

Promoters play a key role in the heterologous protein expression in *P. pastoris* as they can determine the rate of transcription, they can be regulated and can be inserted upstream of genes for expression (Peng et al., 2015). *P. pastoris* is well known for its methanol utilisation and so the discovery of the *AOX1* and *AOX2* genes led to the *AOX1* promoter (P_{AOX1}) being utilised heavily (Yang et al., 2018). P_{AOX1} is a very strong expression tightly regulated promoter, induced by methanol and repressed by glucose and glycerol. The transcription factors Mxr1 (methanol expression regulator

1) and glycerol transporter 1 (Gt1) regulate the expression of AOX1 promoter. Mxr1 is activated during carbon derepression and binds to a consensus sequence on P_{AOX1} promoter resulting in activation of the promoter (Zhan et al., 2017). Wang et al., (2016) reported that methanol induced transcription factor 1 (Mit1) also positively regulate P_{AOX1} expression under methanol induction and represses it under glycerol medium. As methanol is a single carbon molecule, its metabolism may result in lower ATP yield compared to the preferred glucose which contains six carbons. Therefore, it may be necessary that P_{AOX1} needs to be highly expressed to ensure the complete metabolism of methanol is favourable in ATP yield.

As a result, it is commonly used in *Pichia* plasmids to drive the expression of heterologous proteins (Türkanoğlu Özçelik et al., 2019). Another advantage it offers is that due to its size (0.9 kb) and location (*AOX* locus), it serves as a homologous recombination site, allowing for genomic integration and strong heterologous protein expression (Daly and Hearn, 2005). However, there are some disadvantages to using the P_{AOX1}. Due to its strength, the production of some heterologous proteins may not be correctly expressed if downstream events such as protein folding, translation or translocation become limiting factors (Prielhofer et al., 2015). In addition, the P_{AOX1} requiring methanol for induction, there may be health and safety-related issues at large volumes (Moon, 2017a).

Another widely used promoter is the *GAP* promoter (P_{GAP}), equal in expression to P_{AOX1}, providing the benefit of being constitutively expressed. It observes the strongest expression under glucose (2 % final concentration) and retains 2/3 and 1/3 expression under glycerol (2 % final concentration) and methanol (0.5 % final concentration) respectively compared to glucose (Cereghino and Cregg, 2000). However, due to its smaller size (0.4 kb), it has a reduced transformation efficiency compared to P_{AOX1} (Wu and Letchworth, 2004). Kung et al., (2013) reported an exponential increase in transformation efficiency when increasing the homologous region from 96 bp to 1000 bp. The work by Qin et al. (2011) resulted in the creation of a library of constitutive promoters through random mutagenesis of the wild type P_{GAP}. The authors created a library of 33 mutant promoters that were capable of protein expression ranging from 0.6-fold up to 19.6-fold when compared to the wildtype P_{GAP}. However, due to the

mutagenesis, the promoter variants could not be used as a homologous recombination site.

There are some plasmids available for *P. pastoris* that could utilise various recombinant sites for genomic integration depending on the requirements. *P. pastoris* utilises homologous recombination and non-homologous end-joining recombination, however, the process is not well understood (Steinle et al., 2010). In *P. pastoris* the longer the homologous region, the higher the chance of recombination occurring. However, even 1kb or longer regions may result in 1-30% specific integration (Vogl et al., 2018). Several integration sites exist including, auxotrophic genes, promoters and terminators. Various mutant strains were developed to utilise auxotrophic markers, with the most commonly used being the strain *P. pastoris* GS115. It contains a mutation in the *HIS4* gene, which, under normal functioning version, produces an enzyme that is responsible for phosphoribosyl-ATP pyrophosphatase, phosphoribosyl-AMP cyclohydrolase and in this case, histidinol dehydrogenase activities (Karbalaei et al., 2020). Histidinol dehydrogenase is involved in the catalysis of the 2nd, 3rd, 6th, 9th and 10th step of histidine biosynthesis, resulting in the production of histidine - an essential amino acid for *Pichia*. Therefore, it is not able to grow on, or, in media lacking histidine. When this *P. pastoris* GS115 strain is transformed with the appropriate vector containing the correctly functioning *HIS4* gene, the *HIS4* gene will recombine with the inactive variant in the *Pichia pastoris* genome, restoring the full function of the gene and it will be able to grow and the transformant would be selectable (Crane and Gould, 1994:4). The *HIS4* gene can be used as a selection marker or a homologous recombinant site. The same process is applied to other auxotrophic markers.

P_{AOX1}, P_{GAP} and the auxotrophic marker *HIS4* are the most commonly used sequences and sites of homologous recombination (Näätsaari et al., 2012). These sites have been implemented in various plasmid designs, with varying amounts of success (Ahmad et al., 2014).

1.1.7 Metabolic pathway engineering in *P. pastoris*

Metabolic pathway expression in *P. pastoris* requires the integration and/or expression of multiple genes into the *P. pastoris* genome. The common design consists of a promoter for both regulating the expression and acting as a homologous recombination site and an antibiotic marker or an auxotrophic marker for successful transformant selection. They can be used in combination, however, as the number of genes goes beyond two, the size of integrated DNA increases and the efficiency reduces (Vogl, Gebbie, et al., 2018). CRISPR/cas9 plasmids have been developed for *P. pastoris* and can be used for successive or multiple integrations (Weninger et al., 2016a, 2016b; Liu et al., 2019). However, to achieve reliable high efficiency, the *ku70* gene involved in the non-homologous end-joining repair mechanism had to be deleted, resulting in a shorter lifespan of *P. pastoris* cells (Weninger et al., 2018). A study by Nakamura et al. (2018) investigated the implementation of an autonomously replicating sequence (*Cen2*) into *P. pastoris* plasmids. Although the plasmid was stable and had high transformation efficiency, the *Cen2* sequence itself was approximately 6.6 kb, rendering it difficult to be used for large genes or multiple copies of genes. C. Li et al. (2017) identified the integration of unnecessary DNA in *P. pastoris* and investigated genome integration via cre-recombinase.

Cre-recombinase is a protein that recognises and catalyses the recombination between two *lox*p (locus of X-over of P1) sites, independent of co-factors. *lox*p sites are 34 bp, containing two 13 bp cre-recombinase binding regions and a spacer region providing orientation to the *lox*p sites. By placing the *lox*p sites in the same direction, cre-recombinase excises any unwanted DNA in between the two *lox*p sites by recombining the two sites, leaving behind the wanted DNA sequence and a new mutant *lox*p site. However, some of these recombined lox sequences can interact or recombine with any new or re-introduced *lox* sites and may result in unwanted recombination. However, two *lox*p sites (*lox*66 and *lox*71) have been identified that once recombined via cre-recombinase form the mutant *lox*72 site. This mutant *lox*p site is not subsequently targeted for recombination, allowing for the reuse of the *lox*61 and *lox*72 sites for multiple integration events (Oberdoerffer, 2003; Pan et al., 2011; C. Li et al., 2017).

C. Li et al. (2017) utilised these *lox*p sites in *P. pastoris* plasmids and created two segments of the plasmid. The first segment was the required DNA expression cassette (consisting of the promoter, gene to be expressed, terminator) and a *HIS4* integration site. The second segment was the unwanted DNA, containing the cre-recombinase gene (under P_{AOX1} expression) and an antibiotic selection marker gene. The two segments were separated by the *lox66* and *lox71* sites. The plasmid would be transformed into *P. pastoris,* using the *HIS4* as the recombination site. After selecting the correct transformant through antibiotic selection pressure, they cultivated colonies in methanol, allowing the expression of cre-recombinase. This resulted in the excision of the unwanted DNA (cre-recombinase and antibiotic selection marker) from the *P. pastoris* genome, leaving behind the wanted DNA (expression cassette) and *HIS4* gene. The authors utilised this selection marker recycling to integrate multiple copies of a single gene into the *P. pastoris* genome.

This plasmid has the potential to allow metabolic pathway integration via marker recycling and allow a combinatorial strategy to be tested. Combinatorial can be defined as the combination of one or more elements. Combinatorial optimisation of recombinant metabolic pathways requires the simultaneous testing of various elements to achieve an optimum flux. These elements can be categorised as transcriptional (promoters, terminators), translational (5'UTRs, ribosome binding sites) or enzymatic combinations (Coussement et al., 2014). A reduced cost of DNA synthesis as well as improvements in recombinant DNA technologies have made it feasible for extensive high-throughput strain screening.

These advances can be utilised to combinatorially optimise metabolic pathways for the degradation of substrates. In particular, degradation of substrates from renewable sources is an appealing process resulting in the production of compounds such as bioethanol (via glucose) and coniferyl alcohol from lignocellulose biomass.

1.2 Lignocellulose Biomass

Lignocellulose is the most abundant biomass on the surface of the earth (Foyle et al., 2007). It is a sustainable alternative to fossil fuels, as it is a renewable resource allowing to produce second-generation biofuels without compromising global food security (Fargione et al., 2008). Lignocellulose has a complex structure, primarily composed of three polymers: cellulose, hemicellulose, and lignin, and it is the main constituent in plant cell walls (Figure 1.2) (Sanderson, 2011). Cellulose accounts for 30-50% of the plant cell wall, forming microfibrils surrounded by monolayers of hemicellulose. Hemicellulose accounts for 15-30% of the cell wall content, forming inter-links via hydrogen bonds with cellulose, binding the cellulose microfibrils together and strengthening the cell wall. Lignin constitutes 10-25% of the cell wall content forming cross-links within it to provide strength and rigidity to the cell. The percentage of each constituent depends on many factors, such as the climate conditions, the age and the type of the plant (Binod and Pandey, 2015) and nutrient availability (Vassilev et al., 2010).

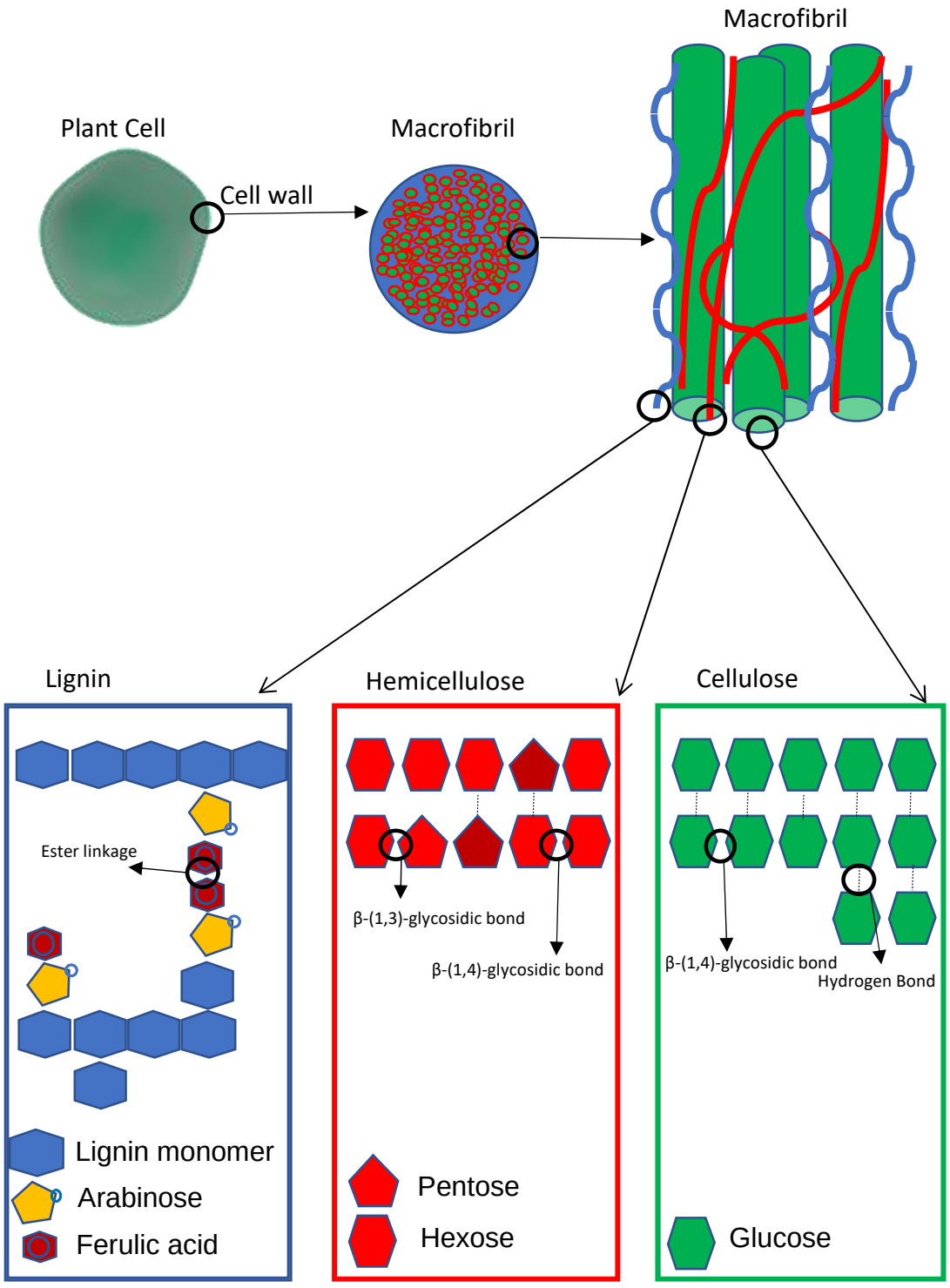

Figure 1.2: Lignocellulose biomass structure (adapted from Baruah et al. 2018; Marriott et al. 2016). Lignocellulose biomass is comprised of lignin, hemicellulose, and cellulose. Ferulic acid forms ester linkages between lignin and hemicellulose to provide strength to the plant cell wall. Hemicellulose consists of pentose and hexose sugars linked by β-(1,3) and β-(1,4)-glycosidic bonds. Cellulose contains glucose units linked by β-(1,4)-glycosidic bonds.

1.2.1 Lignocellulose degradation

The degradation of primary lignocellulosic biomass constituents (cellulose, hemicellulose, and lignin) can lead to the production of various useful compounds (Guerriero et al., 2016). Cellulose is a linear homopolymer consisting of glucose linked units. Its degradation results in the release of single glucose units which are a viable option in the biotech industry as a feedstock for the production of bioethanol (Nanda et al., 2015). Hemicelluloses are a heteropolymer, amorphous structure, composed of units of five-carbon sugars (xylose and arabinose), six-carbon sugars (galactose, glucose and mannose) (Bajpai, 2018), sugar acids (glucuronic acid), acetic acid and phenolic acids (ferulic acid, *p*-coumaric acid) (Broeker et al., 2018). Phenolic acids link with hemicellulose (covalently) and lignin (ester or ether bonds) forming a link between the two, providing rigidity and structure to the cell (Xu et al., 2005). They are identified as "Top value-added chemicals from biomass" to the biotech industry as they can be converted to other more useful and valuable compounds. In particular, ferulic acid can be converted to coniferyl alcohol which is a key compound in synthesizing high-value products such as fragrance compounds, antiviral products and anti-cancer agents (Tramontina et al., 2020). Degradation of lignocellulosic biomass to coniferyl alcohol and bio-ethanol (via glucose) is of high interest to the biotech industry due to its high-value (Rosales-Calderon and Arantes, 2019).

1.2.2 Cellulose degradation

Cellulose is a renewable abundant biopolymer that is utilised in the production of second-generation bioethanol (Wackett, 2008). It can be degraded to single glucose units that are fermented to produce ethanol that can be used as a source of liquid fuel. There are 3 main types of cellulose sources and they can be classified as agricultural waste (eg. corn stover, sugarcane bagasse and wheat straw), wood residue (eg. pinewood and chippings) and dedicated energy crops (eg. switchgrass, timothy grass and miscanthus) (Nanda et al., 2015). Dedicated energy crops are grown for the sole purpose of producing high cellulose content containing lignocellulosic biomass. They produce a higher yield of glucose, however, as they require land for growth they compete with global food security (Gent et al., 2017). In contrast, targeting agricultural residue, i.e. crop components that are inedible to humans and deemed too low in

quality for animal feed, is highly beneficial as it is otherwise considered wasted cellulose. Moreover, the agricultural residue is relatively inexpensive and does not compete with global food security (Fargione et al., 2008).

The degradation of cellulose from agricultural waste is challenging due to its structure and its accessibility as a result of the presence of hemicellulose and lignin (Rosales-Calderon and Arantes, 2019). Cellulose is a polysaccharide consisting of β(1→4) glycosidic linked glucose units (Figure 1.3). This occurs as a result of the inverting of every alternating glucose unit. Therefore, the smallest repeating unit in cellulose consists of two glucose molecules (cellobiose) (Harris and Stone, 2008). The β(1→4) glycosidic bond creates a reducing and non-reducing end and allows the formation of a linear chain up to thousands of glucose units long (500-15,000 nm) (Pérez and Mazeau, 2004). These chains line up in parallel to one another, allowing the formation of inter/intra-molecular hydrogen bonds and van der Waals interactions forming microfibrils. This creates a crystalline structure, which, along with hemicellulose and lignin, make the plant cell wall recalcitrant to enzymatic attack, whilst providing tensile strength (Glazer and Nikaido, 2007). To allow enzymatic degradation of cellulose to produce glucose, ideally, cellulose must be isolated from lignocellulosic biomass.

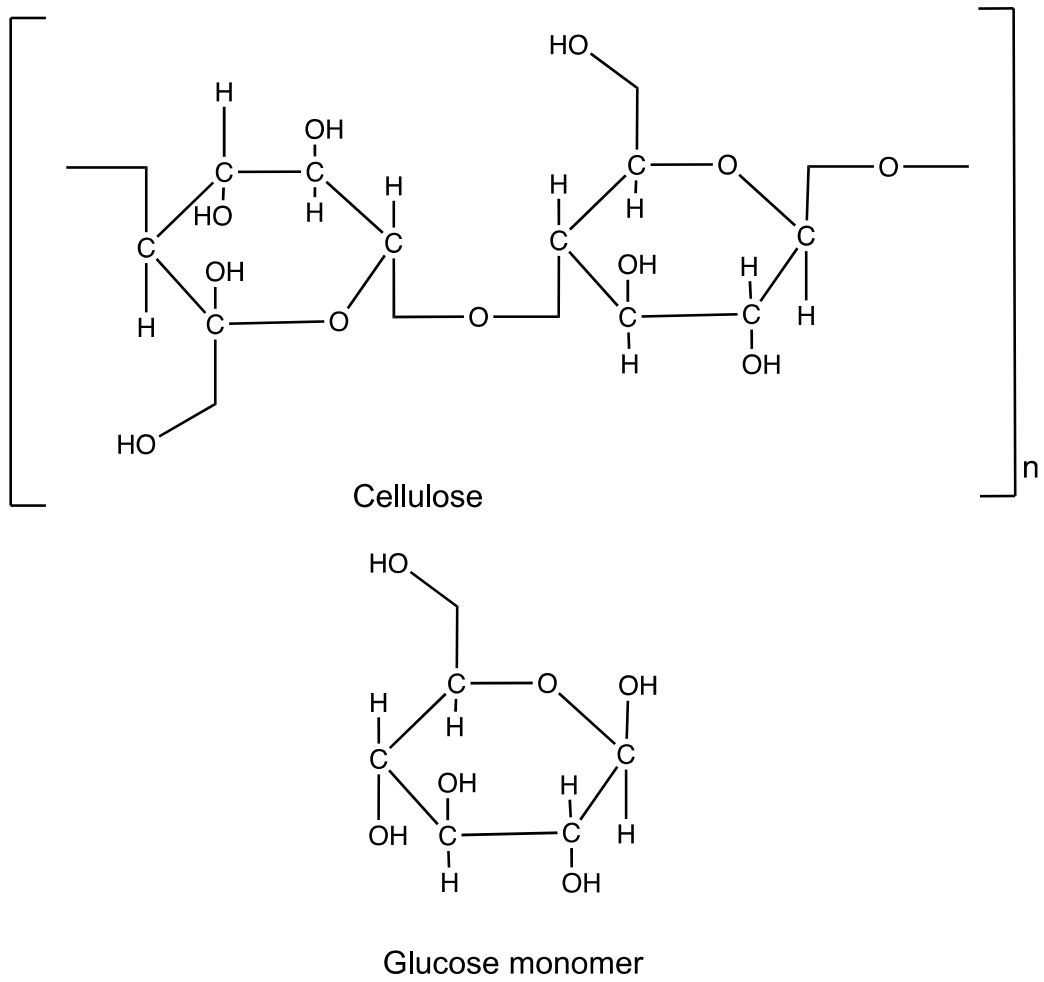

Figure 1.3: Chemical structure of cellulose (adapted from Richards et al. 2012). Cellulose is a polysaccharide of thousands of glucose monomers bonded by β-(1,4)-glycosidic bonds.

1.2.2.1 Cellulases

The knowledge on enzymatic cellulose degradation was gained by observing the microbial degradation of lignocellulosic biomass (cellulose in specific). Cellulases are mainly produced by bacteria (e.g., *Clostridium, Bacillus*) and fungi (eg., *Aspergillus, Penicillium, Trichoderma*) to degrade cellulose content for energy consumption (Bischof et al., 2016). Their occupation of diverse niches is dependent on their ability to obtain carbon and energy sources efficiently from lignocellulosic substrates (Boer et al., 2005). Fungi in particular have adapted well to these conditions by carrying out enzyme (cellulase) secretion and pH modulation. They can produce cellulases that can maintain activity in acidic pH making them ideal for pairing up with acidic pre-treatment conditions (Berlemont and Martiny, 2015). *Trichoderma reesei* was identified as a cellulase producer and is now one of the most popular organisms for

the choice of cellulase production due to its ability to thrive under different climate conditions and its high reproductive capacity (Schuster and Schmoll, 2010). It is known to produce a range of cellulases including endoglucanases (eg., EGI, EGII, EGIII), exoglucanases (eg., CBHI, CBHII) and beta-glucosidases (Miyauchi et al., 2013). Cellulase secretion yields of up to 100g/l have been reported from *T. reesei* at the industrial scale (Martinez et al., 2008).

The degradation of cellulose involves a concerted effort of multiple cellulases. Cellulases are classified (Enzyme Commission) as belonging to the hydrolase group (EC 3) which are responsible for the hydrolysis of β(1→4) glycosidic bonds (EC 3.2) of O- and S-glycosyl compounds (EC 3.2.1) (Lynd et al., 2002). The structure of cellulases consists of two domains: a catalytic domain (CD) and a carbohydrate-binding module (CBM). The two units are linked by a flexible linker. Carbohydrate-binding modules are classified into different families based on their amino acid sequence, substrate specificity and structure (Shoseyov et al., 2006). The carbohydrate-binding module (CBM) is found at the C-terminal of cellulases and is involved in the recognition and binding of polysaccharides (such as crystalline cellulose) (Urbanowicz et al., 2007). The CBMs can be classed into three types depending on its function (Type A, B or C).

Type A CBMs function by interacting with the planar surfaces of crystalline polysaccharides (cellulose). They use their hydrophobic surface which is comprised of aromatic side chains such as tryptophan, tyrosine and phenylalanine (Walker et al., 2015) and have no affinity for soluble polysaccharides. Type B has several subsites with a specific structure arrangement that forms a cleft or groove, allowing the interactions with individual sugar units. These interactions are determined by the aromatic side chains, their orientation and degree of polymerisation (Oliveira et al., 2015) where a lower degree of polymerisation (three or less) results in lower affinity (Boraston et al., 2004). Type C functions by binding to mono-, di- or tri-saccharides, at the non-reducing end as a result of steric restrictions (Pilar Rauter, 2018). The different types of carbohydrate-binding modules allow the recognition and binding to the substrate (crystalline cellulose) in order for the catalytic domain to come into proximity for hydrolysis (Shoseyov et al., 2006). The catalytic domain contains the active site of cellulases involved in the hydrolysis of glycosidic bonds. It contains the

catalytic amino acid residues, proton donors and a nucleophile that lead to hydrolysis via acid/base catalysis, resulting in glycosidic bond cleavage (Li et al., 2010).

The enzymatic degradation of cellulose to release monomers of glucose units is largely dependent on the activity of three types of cellulases. These cellulases are: endo-glucanases, exo-glucanases (Cellobiohydrolases) and Beta-glucosidases. They act together to systematically degrade the rigid crystalline structure of cellulosic material (Figure 1.4) (Shoseyov et al., 2006). After pre-treatment, cellulose is exposed to enzymatic hydrolysis for the production of glucose monomers. Endo-glucanases are responsible for the cleavage of β(1→4) glycosidic bonds in the internal amorphous regions of cellulose. This mostly results in the production of oligosaccharides between two to six saccharides long. This generates new chain ends that are susceptible to hydrolysis by exo-glucanases and beta-glucosidases. The exo-glucanases then act on these new ends (reducing or non-reducing) and result in the production of cellobiose units and may also produce glucose monomers. Exo-glucanases are known to be processive as they remain attached to cellulose chains, performing multiple catalysis reactions. Beta-glucosidases then hydrolyse the cellobiose units, resulting in the production of single glucose monomers (Lakhundi et al., 2015).

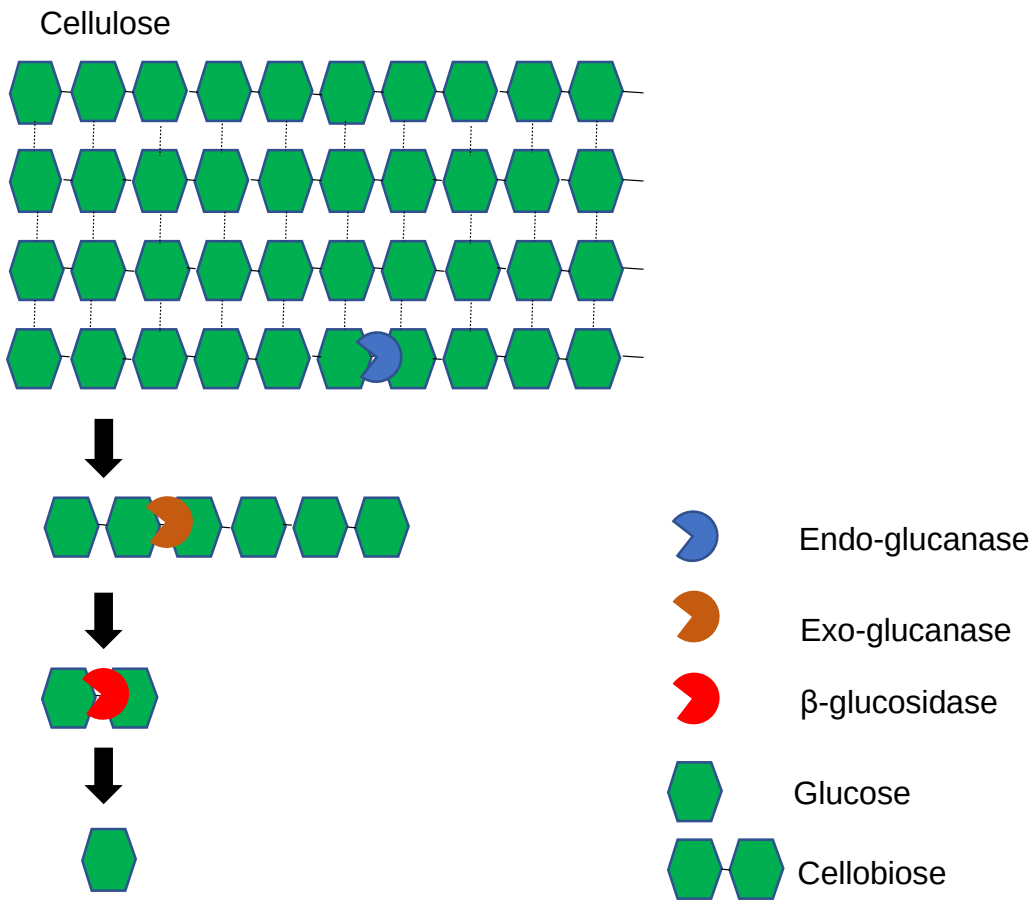

Figure 1.4: Enzymatic degradation of cellulose (adapted from Souza. (2013)). Endoglucanase bind to amorphous cellulose cleaving the internal glycosidic bonds and releasing oligosaccharides between two and six chains long. Exo-glucanases hydrolyse these chains to produce cellobiose. B-glucosidases then hydrolyse the cellobiose units to produce single glucose units.

Cellulases are known to exhibit end-product inhibition where both endo-glucanases and exo-glucanases are inhibited by cellobiose and beta-glucosidases are inhibited by glucose. This makes glucose the rate limiting factor and so an efficient β-glucosidase is a very important factor in improving the conversion rate of cellulose hydrolysis. A study by (Y. Li et al., 2017) identified a highly efficient β-glucosidase (PaBG1b) from the cockroach *Panesthia angustipennis spadica*. PaBG1b was expressed in *P. pastoris* and displayed high specific activity and catalytic efficiency towards cellobiose as well as not exhibiting inhibition for cellobiose up to 100 mM. The degradation of cellulose enzymatically can be achieved in many ways. The efficient degradation of cellulose through these enzymes is aided by synergism where the combined action of these cellulases yields greater hydrolysis than the sum of the individual enzymes (Van Dyk and Pletschke, 2012).

1.2.2.2 Synergism of cellulases

Synergism of cellulases was found to be the highest on the semi-crystalline cellulose-containing high degree of polymerisation and lowest on amorphous cellulose, and almost non-existent on soluble cellulose (Hou and Shaw, 2008). There are four types of synergism observed in enzymatic cellulose degradation. The first is the synergism between endo- and exo-glucanases (endo-exo synergy) where the addition of endoglucanases can increase the amount of substrate available for exoglucanases (Lynd et al., 2002). A study by Jalak et al. (2012) observed the stalling of processive exo-glucanase (TrCel7A) as a result of obstacles encountered in the substrate. The addition of endoglucanase (TrCel5A) resulted in an increased hydrolysis rate. The work by Murashima et al. (2002) showed the addition of cellulases from different families resulting in 1.5- to 3-fold increased activity against soluble cellulose, with no activity increase against insoluble cellulose.

The second form of synergy is exo-exo synergism which results from exo-glucanases processing cellulose substrate from both reducing and non-reducing ends. A study by Igarashi et al. (2011) observed that the movement of an exoglucanase (TrCel7A) on crystalline cellulose would eventually halt, reducing hydrolysis. The addition of a second exoglucanase (TrCel6A) acting at a different end increased the proportion of mobile enzymes on the substrate surface as well as increasing hydrolysis rate.

The third form of synergy is between exoglucanase and beta-glucosidase where the cellobiose is removed. A study by Zhang et al. (2017) highlighted that the addition of beta-glucosidase (TN0602) to an endoglucanase and exoglucanase mixture led to an increased conversion rate of released glucose. The fourth form of synergy is between the carbohydrate-binding domain and catalytic domain. A study by Arai et al. (2003) investigated the truncation of endoglucanase (CelJ) by changing the catalytic domain and carbohydrate-binding domain and found the ratio in which they are expressed had an impact on the production of glucose. By varying the ratio of the truncated CelJ and another endoglucanase (CelC), they managed to observe an increase in glucose production. The work of Kont et al. (2016) analysed the carbohydrate-binding module linker and carried out substitution of amino acids (Trp38 to Ala) in the catalytic domain sequence. The linker resulted in greater affinity towards the substrate and the

amino acid substitution decreased K_{off} rates and increased processivity. Both of these results indicated the action and importance of synergy between the two domains for cellulose degradation.

To improve efficiency in the degradation of cellulose, synergism can be targeted by varying the cellulase mixture by their type and ratio. This is commonly referred to as a "cellulase cocktail". Cellulase cocktails have received considerable attention over the last few years, mainly due to the cost of enzyme production (Aden and Foust, 2009). The cost of enzyme production has reduced over time, however, there is still a need for further improvements to make bioethanol a viable option through enzymatic production (Stephen et al., 2012). Up to 40% of the total cost of bioethanol was due to the cost of cellulase production (Gray et al., 2006). Cellulase cocktails can improve the efficiency of bioethanol production, overcoming the cost of enzyme production as a limiting factor.

Biotechnology companies such as Novozymes and Genencor international have utilised cellulase cocktail mixtures and are now some of the largest providers of these mixtures (Aden and Foust, 2009) offering cellulase cocktails such as accellerase and Cellic CTec2 respectively (Keshavarz and Khalesi, 2016). The creation of a cocktail mixture is very specific to the substrate used, but it is often tailored to certain cellulosic substrates (Gusakov et al., 2007). However, the creation of an optimal mixture of cellulases for specific substrate degradation may not correlate to having an equal outcome when used on complex substrates such as lignocellulosic biomass (Gao et al., 2013).

1.2.3 Ferulic acid degradation

Coniferyl alcohol is a valuable product that can be obtained from the degradation of lignocellulosic biomass. Coniferyl alcohol is derived from the aromatic compound ferulic acid which is a key component of the plant cell wall. Jung & Phillips (2010) investigated a maize mutant that lacked ferulic acid as a feedstock for cows. The lack of ferulic acid resulted in more efficient digestion of the mutant maize indicated by reduced feeding yet increased milk production, highlighting its important role in the plant cell wall. Lignocellulosic biomass is an excellent source of aromatic compounds.

It consists of 10-30% of the total lignocellulose biomass and a valuable source of ferulic acid, which can be more than 2% yield (wt/wt) of the plant biomass (Niño-Medina et al., 2010). Sugarcane bagasse and wheat straw are of interest to the biotech industry as they contain high hemicellulose (xylan) content (Arevalo-Gallegos et al., 2017) incorporating ferulic acid into their cell walls (Harris and Hartley, 1980).

1.2.3.1 Ferulic acid

Ferulic acid is a key component in plant lignocellulosic structure, providing recalcitrance to hydrolysis by forming ester linkages between hemicellulose and lignin (Figure 1.5). In particular, it is found in higher yields in hemicellulose consisting mainly of xylose and arabinose (xylans). The proportion of xylan in lignocellulosic biomass varies depending on the source, but generally, a higher content is found in grasses (20-30 % of the cell wall) than non-grasses. Grass xylan consists mainly of β(1,4) xylose links with various degrees of arabinofuranose substitutions with α-(1,2) or α-(1,3) links (Rennie and Scheller, 2014). Xylan backbone also consists of side chains which are commonly arabinose (forming arabinoxylans) and glucuronic acid (forming glucoronoarabinoxylans). Glucoronoarabinoxylans consist of arabinosyl and glucuronosyl at α-(1,2) links. The molar ratio of arabinosyl to xylose can vary from 1:2 up to 1:30, largely dependent on the type and maturity of the grasses. The side-chain arrangement can dictate the strength of hydrogen bonding to cellulose, influencing the strength of the plant cell wall as a result (Ebringerová and Heinze, 2000).

Arabinoxylans and glucoronoarabinoxylans can form oxidative links between themselves and with lignin through hydroxycinnamic acid residues such as ferulic acid. Fully elongated cells are found to contain higher ferulated arabinosyl units than still elongating cells. Thus, ferulic acid is considered to be important in providing plant cell wall with stability and recalcitrance to hydrolysis (de O. Buanafina, 2009).

Ferulic acid consists of trans-cinnamic acid ($C_6H_5CH=CHCOOH$) containing methoxy ($R\text{-}OCH_3$) and hydroxy (R-OH) substituents at C3 and C4 respectively. The carboxylic acid group of ferulic acid forms an ester bond with the C5 hydroxyl group of the arabinosyl side chain in xylan. When ferulic acid forms ester bonds with

polysaccharides, it can form a dimer through C5-C5 ester bonds, cross-linking the arabinoxylans. However, it is now understood that several types of dimers, trimers and tetramers can also form (Hatfield et al., 2017). In lignin, ferulic acid forms an ether bond with lignin monomers by covalent bonding through its hydroxyl group (Sun et al., 2002). The cross-linking between the hemicellulose and lignin has not been fully understood, however, it is thought that dimers are formed similarly as between arabinoxylans (Hatfield et al., 2017).

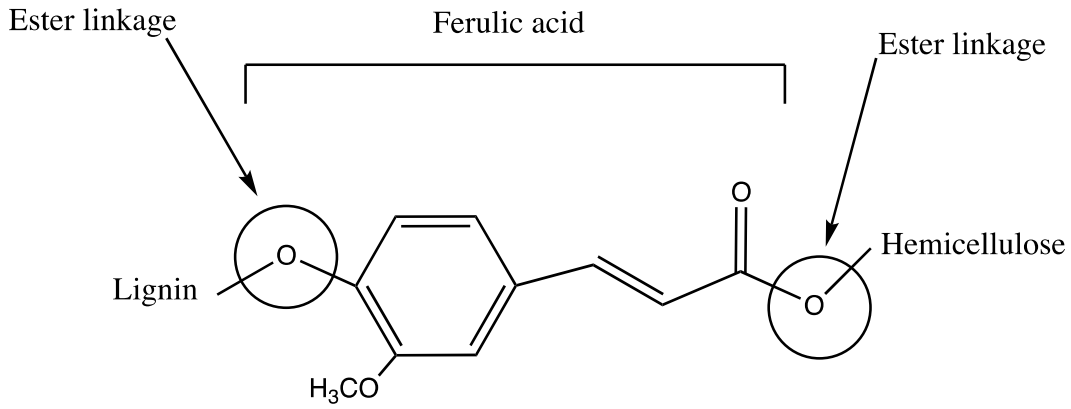

Figure 1.5: Structure of ferulic acid (adapted from Moore and Jung, 2001). Ferulic acid forms ester linkages between hemicellulose and lignin.

1.2.3.2 Ferulic acid derivatives

Lignin is currently seen as a waste product with the paper and pulp industry producing up to 50 million tonnes annually (Bugg and Rahmanpour, 2015). As the need for biofuels increases, so does the waste generated in lignocellulosic degradation (Schoenherr et al., 2018). The extraction of ferulic acid from lignocellulosic biomass and its subsequent conversion to coniferyl alcohol is largely dependent on the lignocellulose biomasses recalcitrance to degradation as well as the challenges in separating and isolating valuable compounds (de Gonzalo et al., 2016). Ferulic acid derivatives can be synthesised through chemical methods such as reduction of the carboxylic acid groups. However, these methods require expensive metals, complex catalyst setups or hazardous substances such as sodium borohydride (Napora-Wijata et al., 2014). Alternatively, ferulic acid derivatives can be produced through plant pathways (Finnigan et al., 2017). Plants use the phenylpropanoid synthesis pathway to synthesize coniferyl alcohol as a building block for lignin. Although it may seem

an applicable pathway, it does require the concert efforts of eight enzymes, which may prove difficult in industrial applications (Wang et al., 2013). Tramontina et al. (2020) exploited the potential applications of metabolic pathway engineering in *E. coli* to release ferulic acid from lignocellulosic biomass degradation and its subsequent conversion into coniferyl alcohol. The authors utilised a three-step cascade consisting of three enzymes, (i) chimeric xylanase (XynZ), (ii) carboxylic acid reductase (niCAR) and (iii) aldo-keto reductase (CgAKR-1) (Figure 1.6).

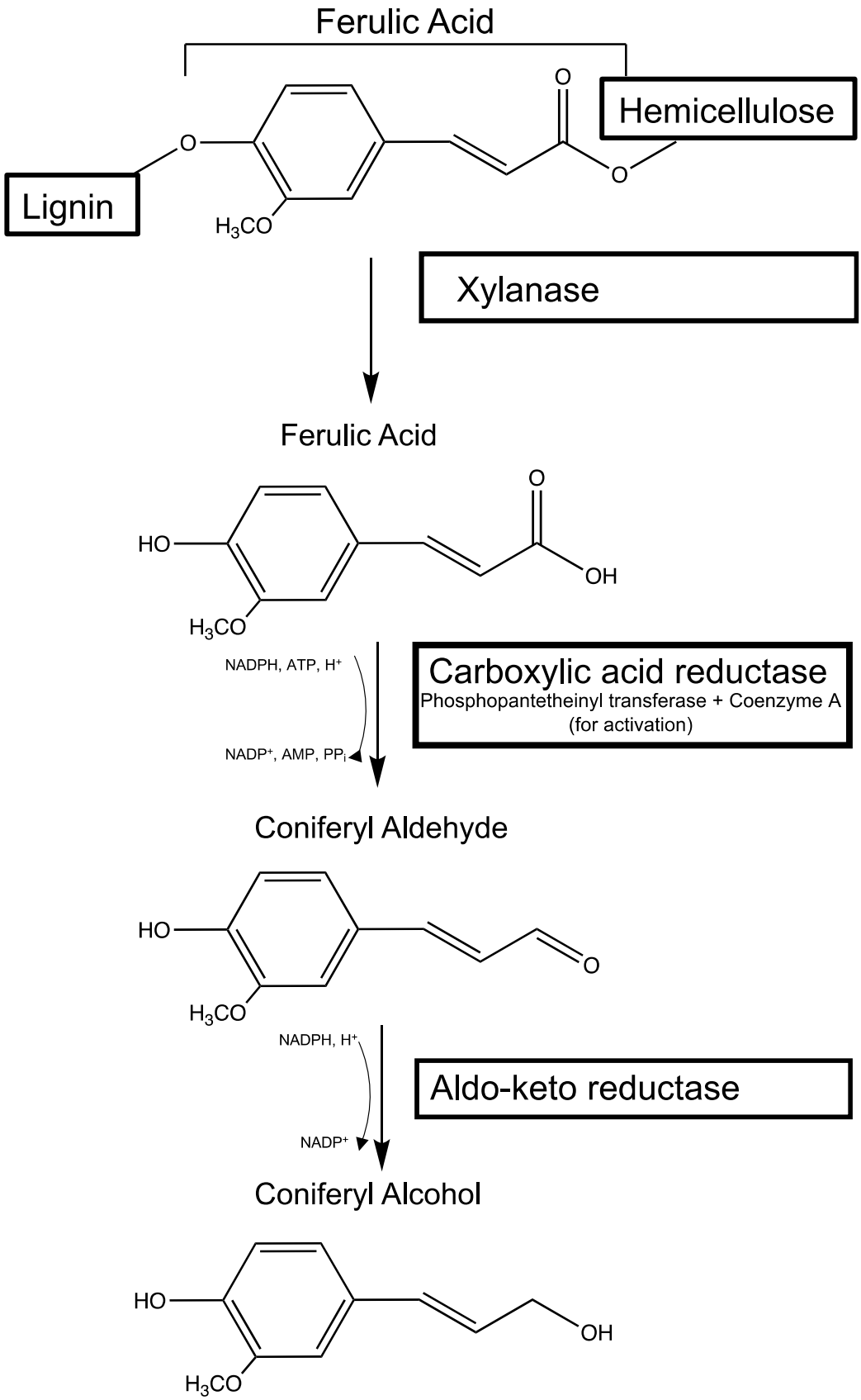

Figure 1.6: Enzymatic conversion of ferulic acid (adapted from (Tramontina et al. 2020). Ferulic acid is released from lignocellulose biomass via the enzyme xylanase. The ferulic acid is then enzymatically converted to coniferyl aldehyde via the carboxylic acid reductase (which is activated by the phosphopantetheinyl transferase and coenzyme A). This is then converted to coniferyl alcohol via the aldo-keto reductase.

1.2.3.3 Enzymes metabolising ferulic acid

The chimeric xylanase XynZ is a xylanase enzyme identified from *Clostridium thermocellum* ATCC 27405 (Blum, Schubot, et al., 2000). Xylanases are a group of enzymes produced by microbial organisms such as bacteria, algae, protozoa and fungi. They are involved in the production of xylose as a carbon sugar source for cell metabolism, from xylan containing sources (Prade, 1996). The xylanase XynZ is unique from other xylanases as it acts as a bifunctional enzyme hydrolyzing the arabinoxylan as well as the ester linkage between the ferulic acid and arabinose (Blum, Kataeva, et al., 2000). XynZ contains a modular structure, consisting of two major domains, glycol hydrolase 10 (GH10) and feruloyl acid esterase. The glycol hydrolase 10 family is known to hydrolyse glycosidic bonds using acid/base catalysis. The GH10 domain has a common mode of action to endo-cellulases, containing a 'cleft' which can accommodate the substrate (Pell et al., 2004).

The feruloyl acid esterase domain belongs to the carbohydrate esterase family 1 (CE1). The feruloyl acid esterases from this family are involved in the hydrolysis of esterified phenolic acids from lignin-carbohydrate complexes (Mandelli et al., 2014). In particular, they can release ferulic acid from lignin polysaccharides (Faulds et al., 1998). They do this by cleaving the ester link between the carboxylic acid group of ferulic acid and the C5 hydroxyl group of arabinose residues in hemicellulose (Hunt et al., 2017). Synergism has been detected between the xylanases and feruloyl esterases where the ferulic acid release is detected in the presence of the xylanase domain. Varying the type of xylanase and feruloyl esterase is also reported to increase ferulic acid production (Wong, 2006). The enzyme XynZ has been reported to produce antioxidant compounds and xylooligosaccharides from sugarcane bagasse (Mandelli et al., 2014). The next step in the cascade is the conversion of ferulic acid to coniferyl aldehyde. This is carried out by a carboxylic acid reductase (CAR).

Carboxylic acid reductases are multi-domain enzymes that reduce carboxylic acid groups to aldehydes using Adenosine Triphosphate (ATP) and Nicotinamide adenine dinucleotide phosphate (NADPH) (Yang et al., 2007). They contain an A-domain which is the substrate-activating adenylation domain. The C-terminus of the A-domain is fused to a phosphopantetheine carrier domain (PCP-domain). The serine in the

phosphopantetheine carrier domain is covalently attached to a phosphopantetheine group and essential in carboxylic acid reductase activity. The C-terminus of the PCP-domain is linked to the reductase domain. The reduction of carboxylic acids is carried out in a multi-step process that involves adenylation, thiolation and reduction. The first step is the ATP-dependent adenylation of carboxylic acid, carried out at the A-domain, resulting in the production of an acyl intermediate. The second step is the transfer of that acyl intermediate group on to the phosphopantetheine linker followed by the third step at the R-domain of the reduction of the acyl-thioester group, resulting in the production of coniferyl aldehyde (Gahloth et al., 2020). The final step involves the conversion of coniferyl aldehyde to coniferyl alcohol through an aldo-keto reductase (cgAKR-1).

Aldo-keto reductases belonging to family 1 are mainly monomeric proteins that act as aldehyde reductases, reducing aldehydes (coniferyl aldehyde) to alcohols (coniferyl alcohol) through an oxidation-reduction reaction. They share a conserved pyridine nucleotide-binding site and a common (B/a)8 structure with the active site in the C-terminus (Chen and Zhang, 2012). They function by reducing the energetic demand of substrates by forming tight binding to NADPH, providing a thermodynamic advantage to reach the transition state. NADPH is used as a redox cofactor and may provide a thermodynamic pathway to favour the reaction, therefore aldo-keto reductases can be very efficient and convert substrates even if the binding is loose (Barski et al., 2008). The aldo-keto reductase used by Tramontina et al. (2020) is the recently reported aldo-keto reductase from the termite *Coptotermes gestroi* (cgAKR-1). It was found to be active against aldehydes commonly found in lignocellulosic biomass (Tramontina et al., 2017). The combined efforts of these enzymes in this unique pathway leads to the degradation of lignocellulosic biomass to release ferulic acid and its conversion to coniferyl alcohol and it was reported to be the first heterologous enzyme cascade of its kind (Tramontina et al., 2020). Although enzymatic degradation of lignin to release ferulic acid can be carried out, it has yet to be replicated at an industrial scale (Silva et al., 2009).

In theory, flux optimisation of ferulic acid degradation or cellulose degradation pathway can be carried out using metabolic control analysis, however, in practicality, there are many challenges to overcome. The enzymes in these metabolic pathways are

commonly recombinantly expressed in ideal host organisms selected for suitability in biotechnology settings. The native and heterologous expression and activity of cellulases can vary, with many factors associated. These factors include things such as the choice of a host organism, codon optimisation, stability intra- and extra-cellularly, stability in growth media, temperature and pH. A fine balance between host organism growth and high product titre is required. Therefore, metabolic engineering requires *in vivo* optimisation, utilising synthetic biology. Most importantly, given that predictable engineering of pathway activities is difficult to achieve, combinatorial strategies are of great interest.

1.3 Objectives and aims

As outlined in the introduction, the pathways of cellulose degradation to produce glucose and the pathways of lignocellulose degradation to coniferyl alcohol are of great interest to the biotech industry as they represent high-value products from a renewable source. Enzymes can be used to degrade the substrates to desired products, whereby secretion from the organisms producing these enzymes is a useful approach. Using a suitable host that can secrete enzyme can lead to reduced benchwork, reduced reagent costs and elimination of the need for cell lysis. The use of enzymatic degradation is an eco-friendly approach that requires recombinant protein production in a suitable host. Yeasts, in particular, *S. cerevisiae* and *P. pastoris* have been awarded 'GRAS' status and have been used in the biotech industry for the production of various biopharmaceuticals. Determining their ability to secrete high levels of proteins can assess their suitability as an industrial organism to express the cellulose degradation and ferulic acid conversion pathway. The utilisation of an effective combinatorial strategy can ensure metabolic pathways are optimised for the production of target compounds.

The work carried out in this thesis was a result of a collaboration between the University of Warwick and Ingenza, Edinburgh, Scotland (iCASE Award PhD). Therefore, the research conducted during this PhD had an industrial focus and aimed at advancing the work of Ingenza.

Chapter 3 describes work intended to compare *S. cerevisiae* and *P. pastoris* in their ability to secrete a green fluorescent protein (ymNeonGreen) by detecting the fluorescent levels using a plate reader. The yeast that displayed the best ability to secrete was then transformed with genes encoding various cellulases provided by the industrial collaborator Ingenza Ltd. The aim was to assess production of these cellulases and to test their activity on simple substrates.

Chapter 4 describes work intended to utilise a cre-recombinase plasmid from C. Li et al. (2017) to determine the feasibility of genomic integration of genes that encode metabolic pathway enzymes. I aimed to modify and improve the performance, efficiency and reliability of this approach as the basis for a new combinatorial strategy for metabolic pathway integration and optimisation in a high-throughput approach.

Three fluorescent proteins (RFP, YFP, iLOV) were used in chapter 5 to assess the integration ability of the newly developed integration method. The new method developed was then utilised to improve the activities of the cellulose degradation pathway and of the ferulic acid metabolic pathway. The more general aim was to understand the effectiveness of the developed combinatorial strategy as a widely applicable tool to enable integration and optimisation of metabolic pathways.

Chapter 2 - Materials and methods

2.1 Media

2.1.1 Bacterial Media

2.1.1.1 LB Broth (Lennox)

LB Broth was used to grow cultures of *E. coli*. To create the media, 10 g/L Tryptone, 5 g/L yeast extract and 5 g/L sodium chloride were added in appropriate volumes of distilled water and autoclaved for 15 minutes at 15 psi, 121˚C.

2.1.1.2 Super optimal broth with catabolite repression (SOC) media

SOC media was used to recover *E. coli* after heat shocks. The media consisted of 2% w/v tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5mM KCl, 10 mM $MgCl_2$ 10 mM $MgSO_4$. 20 mM sterile glucose was added after SOC was autoclaved for 15 minutes at 121 psi, 140˚C.

2.1.1.3 Super optimal broth (SOB) media

SOB media was used to prepare electrically competent *E. coli* cells. The media consisted of 2% w/v tryptone, 0.5% w/v yeast extract, 10 mM NaCl, 2.5mM KCl, 10 mM $MgCl_2$ 10 mM $MgSO_4$ autoclaved for 15 minutes at 121 psi, 140˚C.

2.1.2 Yeast Media

2.1.2.1 Yeast Extract Peptone (YEP)

1% yeast extract (Formedium), 2% Bacto peptone (Sigma-Aldrich) were mixed in appropriate volumes of distilled water and autoclaved for 15 minutes at 121 psi, 140˚C. 4% v/v of 50% sterile dextrose (Fisher scientific) was added to YEP to create

YEPD (Yeast Extract Peptone Dextrose). 4% v/v of 50% sterile glycerol (Fisher scientific) was added to YEP to create YEPG (Yeast Extract Peptone Glycerol) (Table 2.1).

Table 2.1: S. cerevisiae growth media composition

	YEP	**YEPD**	**YEPG**	**YEPDA**
Yeast Extract	1% w/v	1% w/v	1% w/v	1% w/v
Peptone	2% w/v	2% w/v	2% w/v	2% w/v
Agar	-	-	-	2% w/v
Autoclave for 15 minutes at 121 psi, 140˚C				
Dextrose	-	2% w/v	-	2% w/v
Glycerol	-	-	2% w/v	-

2.1.2.2 Buffered Glycerol-complex medium (BMGY) and Buffered Methanol-complex Medium (BMMY)

1% yeast extract (Formedium), 2% bacto peptone (Sigma-Aldrich) were mixed in appropriate volumes of distilled water and autoclaved for 20mins at 121 psi, 140˚C. Then, 1 M potassium phosphate buffer (pH range 5.8 – 8.0) or 1 M sodium acetate buffer (pH range = 3.6 – 5.6) was added for a final concentration of 0.1 M and the pH of the solution was adjusted as required. For potassium phosphate buffer pH adjustment, 1 M phosphoric acid or 1 M KOH was added until required pH was obtained. For sodium acetate buffer, pH was adjusted using 0.1 M HCl or NaOH. Then, filter sterilised Yeast Nitrogen Base with ammonium sulphate (YNB) (Formedium) was added for a final concentration of 1.34% w/v. Filter sterilised Biotin was then added for a final concentration of 0.02% w/v. For BMGY, 2% w/v sterile glycerol was added and for BMMY, 0.5% v/v filter sterilised methanol was added. Where applicable, G418 antibiotic was added for a final concentration of 100 µg/mL and zeocin was added for a final concentration of 50 µg/mL.

Table 2.2: P. pastoris growth media composition

	BMMY	**BMGY**	**BCMY**
Yeast Extract	1% w/v	1% w/v	1% w/v
Peptone	2% w/v	2% w/v	2% w/v
Autoclave for 15 minutes at 121 psi, 140˚C			
Glycerol	-	2% w/v	-
Methanol	0.5% w/v	-	-
Buffer 1M	Potassium phosphate or Sodium acetate 10%v/v	Potassium phosphate or Sodium acetate 10%v/v	Potassium phosphate or Sodium acetate 10%v/v
Yeast nitrogen base (13.4%)	10% v/v	10% v/v	10% v/v
Cellulose (6mm disc)	-	-	1 disc per well

2.2 Cell protocols and assays

2.2.1 Bacterial cells

2.2.1.1 Preparation of electrically competent E. coli cells

E. coli cells were made competent as follows; a single colony from *E. coli* streaked on agar medium was cultivated overnight at 37 ºC and 250 rpm in 10 mL SOB media. 1 mL from this overnight culture was added to a 1L flask containing 250 mL of SOB media pre warmed at 37 ºC. The cells were grown to an OD_{600} of between 0.5 and 0.7 and the cultures were then kept on ice for 15 minutes. The cells were transferred to a 500 mL centrifuge bottle and centrifuged (5000 rpm, 4 ºC) for 10 minutes. The supernatant was discarded and the cells were resuspended in 250 mL of ice-cold

sterilised 10% glycerol. The cells were centrifuged (5000 rpm, 4 °C) for 10 minutes. The supernatant was discarded and the cells were resuspended in residual glycerol and 50 µL aliquots were made into 1 mL sterile Eppendorf tubes. The samples were stored at -80 °C.

2.2.1.2 Transformation of E. coli cells via electroporation

Competent *E. coli* cells were thawed on ice for 30 mins. 3 µL of plasmid DNA (10-100 ng) was added to 50 µL of competent *E. coli* cells and gently mixed and incubated on ice for a further 10 mins. This mixture was transferred to a 2 mm electroporation cuvette (Gene Pulser, 1652086). Air bubbles were removed by tapping the cuvette twice on the countertop and excess water from the cuvette was wiped off. The cuvette was placed in the electroporator (Eppendorf eporator, Catalog No. 4309000035) and pulsed at 1,700 V, τ~5.6ms. Immediately, 950 µL of 37 °C SOC media was added and the mixture transferred to a 15 mL falcon tube. The cells were incubated in a shaking incubator at 250 rpm and 37 °C for 1 hour. The cells were centrifuged at 3250 rpm for 5 minutes. 100 µL of supernatant remained in the tube whilst the rest was discarded. The cells were resuspended gently and plated on to selective medium and incubated overnight at 37 °C.

2.2.2 Yeast cells

2.2.2.1 Preparation of chemically competent S. cerevisiae cells

A single *S. cerevisiae* colony was grown in YEPD overnight at 30 °C. The cells were resuspended to 0.1 OD_{600} in the morning and left to grow between OD_{600} 0.5 and 0.8. The cells were then centrifuged (3250 rpm, 4 °C). Supernatant was discarded and cells were washed once in sterile water and twice in TE/LiAc (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1 M LiAc pH 7.5) where centrifugation followed between the wash steps (3250 rpm, 4 °C) and supernatant was discarded each time. After washing, the cells were resuspended in 200 µL of TE/LiAc and ready to be used immediately.

2.2.2.2 Transformation of chemically competent S. cerevisiae cells

Linearised plasmid DNA (5 µL) and 10 uL (10 mg/mL) denatured salmon sperm (Sigmaaldrich) were added to 50 uL of competent cells followed by 300 µL of 40 % PEG/LiAc (40 % PEG, 10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1 M LiAc pH 7.5). The cells were incubated at 30 ^{o}C for 1 hour followed by a heat shock at 42 ^{o}C for 40 minutes. The cells were washed in sterile water and resuspended in 1 mL of YEPD and incubated at 30 ^{o}C for 3 hours. The cells were then harvested by centrifugation and resuspended in 100 µL of sterile water and plated onto selective YEPDA plates and incubated at 30 ^{o}C for 3 days.

2.2.2.3 Preparation of electrically competent P. pastoris cells

A single colony of *P. pastoris* was grown overnight in a shaking incubator (250 rpm, 30 ^{o}C) in 5 mL YEPD medium. In the morning, the cells were resuspended to OD_{600} of 0.1 and grown in 50mL YEPD medium in the shaking incubator (250 rpm, 30 ^{o}C) until cells reached OD_{600} between 0.5 and 0.8. The cells were harvested by centrifugation (3250 rpm, 4 ^{o}C, 5 minutes). The supernatant was discarded and cells were resuspended in 9mL of ice-cold BEDS solution (10mM bicine-NaOH, pH 8.3, 3 % (v/v) ethylene glycol, 5 % (v/v) (dimethyl sulfoxide) DMSO, 1 M sorbitol) and 1 mL dithiothreitol (DTT). The mixture was incubated (100 rpm, 30 ^{o}C, 5 minutes) followed by centrifugation (3250 rpm, 24 ^{o}C, 5 minutes). The supernatant was discarded and samples were resuspended in 1 mL of ice-cold BEDS solution. 50 µL aliquots were made into 1 mL sterile Eppendorf tubes and stored at -80 ^{o}C.

2.2.2.4 Transformation of electrically competent P. pastoris cells

Competent 50 µL of *P. pastoris* were thawed on ice for 30 minutes. 5 µL of linearised plasmid DNA was added to the cells, gently mixed and incubated on ice for 10 minutes. The mixture was then transferred to a 2 mM electroporation cuvette (Bio-rad) and tapped twice on the countertop to remove any air bubbles. Excess water was wiped, the cuvette was placed into the electroporator (Gene Pulser, 1652086) and pulsed at 1,500 V, τ~5.6ms . Immediately after pulsing, the cells were resuspended in

950 µL YEPD. The mixture was transferred to a 15 mL falcon tube and incubated (250 rpm, 30 °C, 3 hours). The cells were pelleted by centrifugation (5,000 rpm, 5 minutes), the supernatant was discarded and cells were resuspended in 100 µL of sterile water. The mixture was plated onto selective YEPDA medium and incubated for 2-3 days.

2.3 DNA manipulation and cloning

2.3.1 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to separate DNA of differing sizes. A 1 % agarose gel was made by melting 1 g agarose in 100 mL Tris-acetate-EDTA (TAE) buffer (40 mM Tris-HCl, 20mM acetic acid, 1 mM EDTA, pH 8.0). The mixture was allowed to cool to 60˚C and 10 µL SYBR safe DNA stain (Invitrogen) was added to the mixture and dissolved. The mixture was poured into gel cast trays with well selective sized well combs and allowed to cool for 30 mins. The well combs were removed and the gel was placed in a Biorad tank and filled with TAE buffer. DNA loading dye (NEB #B7025, 6x concentration) was added to samples at a final concentration of 1x and the mixture was loaded (12 µL for cPCR and 20 µL for PCR purification). The ladder used was PCR bio ladder II (PCRBio, PB40.12-01 [6 µL per lane]). The gels were electrophoresed at 100 V until sufficient separation of the require DNA fragments had occurred. Upon completion of the electrophoretic run, the gel was imaged in a Biorad Gel Doc XR+ gel documentation system. Where applicable, the DNA bands were sliced out and purified using the Thermo scientific GeneJET Gel Extraction kit.

2.3.2 Isolation and purification of plasmid DNA from *E. coli*

A single colony from the *E. coli* strain containing the plasmid of interest was grown overnight in 5 mL of LB media at 37 °C. In the morning, the cultures were centrifuged (3250 rpm, 10 minutes) and the supernatant was discarded. GeneJET plasmid purification miniprep kit was used to obtain purified plasmid DNA and stored in sterile de-ionised wated and kept at -20 °C.

2.3.3 Plasmid linearisation for *S. cerevisiae* and *P. pastoris* integration

The plasmid of interest was linearised using a selected restriction site. 2 µg of plasmid DNA was added to a 0.2 mL PCR tube. The specific restriction enzymes (2 µL of 20,000 units/mL [40 units per reaction]) were added to the PCR tube alongside 2 µL of appropriate buffer (cutsmart 20x buffer [1x final cutsmart buffer concentration]) as according to the manufacturer's instructions (NEB) and enough sterile water to make up to 20 µL of reaction mixture. The reaction was incubated according to the manufacturer's instructions (dependent on the enzyme being used). 5 µL of 6x purple loading dye (Sigmaaldrich [to increase sample density compared to TAE and follow sample movement through electrophoresis]) was added to the sample (1x purple loading dye final concentration) and analysed via agarose gel electrophoresis. The linearised plasmid was excised and purified using the Thermo scientific GeneJET Gel extraction kit.

2.3.4 Polymerase chain reaction (PCR)

The PCR for product amplification was carried out by mixing components in Table 3.

Table 2.3: PCR components of Q5 enzyme

Component	**50µL reaction volume**	**Final concentration**
10 µM Forward Primer	1.25 µL	0.5 µM
10 µM Reverse Primer	1.25 µL	0.5 µM
Template DNA	Experiment dependent	
dNTPs (Sigma)	0.5 µL	0.2 mM
10x Q5 buffer (NEB)	5 µL	1x
Q5 enzyme (NEB)	0.25 µL	0.02 U/µL
Sterile dH_2O	To complete mix to 50 µL volume	

The PCR was carried out in 0.2 mL PCR tubes in a Bio-Rad thermocycler (T100). The steps in the reaction were carried out as described in Table 2.4.

Table 2.4: Thermocycler PCR steps of Q5 enzyme

Step	**Temperature**	**Time**	**Cycles**
Initial denaturation	98 °C	30 seconds	-
Denaturation	98 °C	10 seconds	25-35
Annealing	Lower Tm +3 °C	30 seconds	
Extension	72 °C	30 seconds per kb	
Final extension	72 °C	5 minutes	-

The PCR products were analysed via agarose gel electrophoresis and purified using Genejet Gel extraction kit or Thermo PCR purification kit where applicable.

2.3.5 Colony Polymerase chain reaction (cPCR)

Colony PCR was carried out by mixing the components in Table 2.5. The cPCR was carried out in 0.2 mL PCR tubes in a Bio-Rad thermocycler (T100). The steps in the reaction were carried out as described in Table 6. The final products were analysed via agarose gel electrophoresis, as described above.

Table 2.5: PCR components of 2x PCRBIO Taq mix red enzyme

Component	**20µL reaction volume**	**Final concentration**
10 µM Forward Primer	0.8 µL	0.4 µM
10 µM Reverse Primer	0.8 µL	0.4 µM
Template DNA	Experiment dependent	
2x PCRBIO Taq Mix Red	10 µL	1x
Sterile dH_2O	To complete mix to 20 µL volume	

Table 2.6: Thermocycler PCR steps of 2x PCRBIO taq red mix

Step	**Temperature**	**Time**	**Cycles**
Initial denaturation	95 ^{o}C	60 seconds	-
Denaturation	95 ^{o}C	15 seconds	40
Annealing	55 – 65 ^{o}C	15 seconds	
Extension	72 ^{o}C	15 seconds per kb	
Final extension	72 ^{o}C	5 minutes	-

2.3.6 Plasmid restriction digest

Plasmids to be ligated were restriction site digested as follows:
Between 1-2 µg of plasmid DNA or PCR product was added to a 0.2 mL PCR tube. The specific restriction enzymes (2 µL of 20,000 units/mL [40 units per reaction]) were added to the PCR tube alongside 2 µL of appropriate buffer (cutsmart 20x buffer [1x cutsmart buffer per reaction]) as according to the manufacturer's instructions (NEB) and enough sterile water to make up to 20 µL of reaction mixture. The reaction was incubated (25 or 37 ^{o}C for between 15 – 60 minutes) according to the restriction enzyme used and the manufacturer's instructions. After restriction digestion, 1 µL (1 unit/µL) of FastAP Thermosensitive phosphatase (Thermofisher) was then added to the reaction to perform vector dephosphorylation and incubated at 37 ^{o}C for 15 minutes. The enzyme was then inactivated by incubating at 75 ^{o}C for 5 minutes. 5 µL of 6x purple loading dye (1x final concentration) was added to the mixture and the digested plasmid was purified via agarose gel electrophoresis and the excised band purified via GeneJET Gel extraction kit.

2.3.7 Plasmid Ligation

Digested plasmid and insert were added together in a 1:3 pmols ratio respectively. 2 µL T4 DNA Ligase buffer and 1 µL T4 DNA ligase were added. Sterile water was added to complete reaction volume to 20 µL. The reaction was mixed and incubated at room temperature for 30 minutes. The mixture was heat inactivated at 65 ^{o}C for 10 minutes followed by cooling step on ice. The reaction was then transformed into *E. coli*.

2.3.8 Gibson Assembly insert preparation

The insert to be cloned into the desired plasmid was PCR amplified containing 5' and 3' overhangs homologous to the plasmid digested sites and place of insert. The overhangs were created containing a minimum of 25 DNA base pairs of 60 ^{o}C Tm. Geneblocks (IDT) to be used as insert were prepared as according to the manufacturer's instructions. 50 µL of sterile ddH_2O was added to the tube, the DNA was dissolved by pipetting and vortexed briefly and incubated to 50 ^{o}C for 20 minutes followed by vertexing briefly again. The tube was briefly centrifuged and stored in -20 ^{o}C.

2.3.9 Gibson Assembly

The Gibson master mix was produced as outlined in Table 2.7 and Table 2.8. The reaction mixture was aliquoted in 15 µL volume into 0.2 mL PCR tubes and stored at -20 ^{o}C. For the reactions, each tube was thawed on ice for 10 minutes. Up to 5 µL of digested plasmid and insert (digested, amplified or geneblock, 50 – 100 ng) was added at a pmols ratio of 1:3 respectively and necessary volume of ddH_2O was added to make the mixture to 5 µL. In the negative controls, equivalent volume of the digested vector only was added (ddH_2O added to complete mixture to 5 µL). The reaction mixture was incubated at 50 ^{o}C for 30 minutes (2 – 3 inserts) or 60 minutes (4 – 6 inserts). Where low Gibson assembly of plasmids efficiency was observed (post-transformation), the incubation step was increased to 120 minutes at 50 ^{o}C. The samples were then transformed into *E. coli*.

Table 2.7: Components of 5x Isothermal reaction mix

Component	**Volume / weight**	**Final concentration**
1 M Tris-Hcl (pH 7.5)	3 mL	0.5 M
1 M MgCl2	300 µL	100 mM
dGTP	60 µL	1 mM
dATP	60 µL	1 mM
dTTP	60 µL	1 mM
dCTP	60 µL	1 mM
1 M DTT	300 µL	100 mM
0.1 M NAD	300 µL	5 mM
PEG8000	1.5 g	30 mM
Sterile ddH_2O	1860 µL	-

Table 2.8: Components of Gibson assembly master mix

Component	**Volume**	**Final concentration**
5x Isothermal Master mix	500 µL	0.266x
10 U/µL T5 exonuclease	1 µL	5.33 U/mL
2 U/µL Phusion DNA polymerase	31.25 µL	33 U/mL
40 U/µL Taq DNA ligase	250 µL	5.33 U/µL
Sterile ddH_2O	1092.75 µL	-

2.3.10 Sanger sequencing

5 µL of plasmid or PCR product (30 – 100 ng/µL or 10 – 50 ng/µL respectively) was mixed with 5 µL 10 µM primer and sent to Genewiz or Eurofins-GATC sequencing. The sequencing data was aligned using SnapGene software (v4.2, GSL Biotech LLC).

2.4 Flow cytometry and Microscopy

2.4.1 Flow Cytometry

A single *S. cerevisiae* colony was selected and grown per tube overnight in 5 mL YEPD at 30 ^{o}C. The culture was refreshed in the morning to OD_{600} of 0.1 and grown until OD_{600} was between 0.5 and 0.7. 1 mL of the culture was transferred to a 15 mL falcon tube and the cells were harvested by centrifugation (4000g, 10 mins). The cells were washed twice in PBS and resuspended in 1 mL PBS. They were then transferred to a flow cytometry tube and sonicated for 15 seconds. The sample was analysed via the flow cytometry (BD Biosciences Fortessa) and measurements were made by selection of emission and excitation wavelengths for the applicable fluorescent protein (GFP B488-530/30A). Flow cytometry analysis was carried out using FACSDiva software (v 642213 Rev A).

2.4.2 Microscopy

Samples were grown to OD_{600} 0.5 – 0.7 and 5 µL was spotted on microscope slides. The cover was attached on top and the images were acquired using a Nikon TI-E Eclipse motorized spinning disk confocal microscope using either 60x or 100x oil immersion objective. The lens' used were (CFI Plan Apochromat DM 60x Lambda Oil Ph3, NA 1.4, WD 0.13mm, Ph3 and CFI Plan Apochromat DM 100x Lambda Oil Ph3, NA 1.45, WD 0.13mm, Ph3). The use of brightfield or specific wavelength emission excitation allowed fluorescent colony images to be obtained (GFP filter λ_{ex}/ λ_{em} 525/50 nm for iLOV, ETCFP filter λ_{ex}/ λ_{em} 436/480 nm for mTFP1 and ETEYFP λ_{ex}/ λ_{em} 500/25 nm for mCitrine and ETmCherry λ_{ex}/ λ_{em} 560/630 nm for mCherry. The images were taken at 10-30% light power and 200 milliseconds exposure. The software ImageJ (v2.0.0-rc-66/1.52b) was used to analyse the images.

2.5 Robotics and Plate reader

2.5.1 Plate reader

The strains that were to be analysed were diluted to OD_{600} 0.1 (unless specified) in fresh YEPD medium. 1 mL of the culture was added to each well of a 24 well plate (Eppendorf, 0030741005) including appropriate controls. Black plates with clear bottoms (Eppendorf, 0030741005) were used for all OD_{600} and fluorescent measurements. Tecan F200 Pro was used as the plate reader for all measurements. The programmes were created to allow a measurement every designated period of time. The conditions in the plate reader were kept at 30 ^{o}C and 250 rpm. Growth rate was measured as the OD_{600} of the samples and fluorescence was measured using specific excitation and emission wavelengths (mNeongreen λ_{ex}/ λ_{em} 506/517 nm respectively, iLOV λ_{ex}/ λ_{em} 447/497 nm respectively, mCherry λ_{ex}/ λ_{em} 587/610 nm respectively, mTFP1 λ_{ex}/ λ_{em} 462/492 nm respectively and mCitrine λ_{ex}/ λ_{em} 516/529 nm respectively). The output was the value of each parameter in a Microsoft Excel spreadsheet.

2.5.2 Robotics - Tecan Freedom Evoware

Tecan freedom Evoware liquid handling robotics was used to measure growth of multiple plates over time. Cultures (samples and controls) were transferred to a 24 well plate (Eppendorf, 0030741005) and the plates were added to the storage compartment. A script was initialised that allowed the addition of those plates in to a shaking incubator. The plates underwent orbital shaking (250 rpm) at 30 ^{o}C to grow cultures. At each designated time point, the plate was removed by the robotic arm and transferred to the plate reader that allowed O.D. measurement or fluorescence measurement to be made. The protocol is highlighted in the appendix (Table S2.2).

2.6 Protein Purification and identification

2.6.1 Buffers

Table 2.9: Composition of protein analysis buffers

Buffer A	**Buffer B**	**Buffer C**
250 mM Tris-HCl pH 6.8	50 mM Potassium phosphate buffer pH 7.5	50 mM Potassium phosphate buffer pH 7.5
10 % SDS	100 mM NaCl	100 mM NaCl
0.008 % Bromophenol Blue	5% glycerol	5% glycerol
10 % Glycerol		300 mM Imidazole
2.854 M β-Mercaptoethanol		

2.6.2 Protein production

Cells were grown overnight in YEPD at 30 ^{o}C and 250 rpm and diluted in the morning to OD_{600} 0.1 in fresh YEPD. The cultures were then grown in 50 or 250 mL YEPD in a 1 L flask at 250 rpm, 30 ^{o}C. Cultures were grown for specific time points. Upon completion of specific time point, cell cultures were centrifuged (3250 rpm, 4 ^{o}C, 10 minutes). Supernatant was separated from cell pellet, centrifuged again and stored on ice. The cell pellet was lysed using a French press. The cell pellet was resuspended in 20 mL buffer (20 mM HEPES pH 7.4, 750 mM KCl, 2 mM MgCl2, 1 mM PMSF, 20 mM imidazole, 0.5 mM B-ME). The cells were disrupted twice at approx. 1500 psi, keeping on ice in-between. The cell lysate was centrifuged (3250 rpm, 15 minutes, 4 ^{o}C) and supernatant was kept on ice for further purification.

2.6.3 Nickel affinity column protein purification

The cell culture supernatant and cell lysate supernatant from 2.6.2 to be analysed were kept on ice. The supernatant was then pushed through a Histrap FF column (Scientificlabs) at a flowrate of 1 mL/min using a peristatic pump. The column was washed and equilibrated with 10mL buffer B (Table 2.9) and several 500 µL elutions were made with Buffer C (Table 2.9). The samples were either analysed via SDS-PAGE or stored at -80 ^{o}C.

2.6.4 SDS-PAGE

30 µL of purified protein samples were resuspended in 10 µL buffer A (4x conc) (Table 2.9) and denaturated at 95 ^{o}C for 15 minutes. The samples were cooled, vortexed briefly and centrifuged briefly. The sample was loaded into a precast gradient gel (4 – 20 %) (Expedeon) and placed into a Bio-Rad mini PROTEAN Tetracell system containing Expedeon RunBlue running buffer (0.8 M Tricine, 1.2 M Tris, 2% SDS). The electrophoretic conditions were 100 V for 80 minutes (constant voltage mode) and stained with Expedeon Instant Blue dye. The resulting gel was destained via multiple washes using distilled water and imaged using a GelDoc XR+ imaging system. The protein bands of interest were sliced for mass spectrometry identification.

2.6.5 Mass spectrometry

Protein bands were excised from SDS-PAGE and cut into small cubes (~1 mm^3). The pieces were dehydrated in acetonitrile for 10 mins twice and the supernatant was discarded. The samples were dried using SpeedVac. 25 mM ammonium bicarbonate was added to the cubes until they were covered (50 – 100 µL). The reaction was incubated at 56 ^{o}C for 1 hour. Supernatant was discarded and tube allowed to cool at room temperature. 100 mM alkylate with iodacetomide in 25 mM ammonium bicarbonate was added to the tube and left at room temperature for 30 minutes. Supernatant was discarded and samples washed in 25 mM ammonium bicarbonate. The samples were dehydrated again in acetonitrile for 10 minutes and dried using SpeedVac. The gel pieces were rehydrated in buffer containing 50 mM bicarbonate and 12.5 ng/µL fresh trypsin. The gel pieces were then fully covered using 50 mM ammonium bicarbonate. The reaction was digested overnight at 37 ^{o}C. In the following morning, samples were briefly centrifuged and supernatant was collected in a new tube. 3 cycles of 5 % formic acid in 50 % acetonitrile were added to the gel pieces and incubated for 20 minutes. The supernatant was added to the tube containing the previous supernatant. Sample volume was reduced in a SpeedVac to volume 5 µL and samples were submitted at -20 ^{o}C to the proteomics facility where reverse phase chromatography was used to separate the peptides and electrospray ionisation method

was used to analyse the peptides via Thermo Orbitrap fusion system (Q-OT-qIT, ThermoScientific).

The software Scaffold (ver. Scaffold_4.3.2, Proteome Software inc.) was used to visualise the files provided by the proteomics facility. Peptide fragments were matched against provided sequences and accepted if they contained a 95 % or greater probability of belonging to the sequence. A protein match was accepted if more than two peptides matched the provided sequence.

2.7 HPLC analysis

2.7.1 Metabolite extraction

Compounds of interest (ferulic acid, coniferyl aldehyde and coniferyl alcohol) were extracted using acetonitrile. 1 mL of cell culture was centrifuged (13,300 rpm, rt, 10 minutes). The supernatant was collected into a new tube and acetonitrile (100 %) was added in equal volume. This was done to obtain compounds in the extracellular medium. 0.2 mL acetonitrile was added to the cell pellet, vortexed briefly and centrifuged (13,300 rpm, rt, 5 minutes) and the supernatant was added to a new tube. This was done to obtain compounds from intracellular medium. This process was repeated 4 times in total and the respective extracellular and intracellular samples were collated. Both samples (extracellular and intracellular supernatants containing acetonitrile) were then centrifuged using a nylon-based spin column (Sigma, CLS8169) ready for further analysis or stored at -20 ^{o}C. Samples to be processed were placed in 2 mL brown HPLC vials.

2.7.2 HPLC

HPLC was carried out on the Agilent 1260 infinity II LC system. The mobile phase A was 100 % acetonitrile (HPLC grade) and the mobile phase B was 0.1 % formic acid in ddH_2O (HPLC grade). The column was C18 (Ascentis® Express C18, 2.7 µm HPLC Column, 2.7 µm particle size, L × I.D. 15 cm × 2.1 mm). Sample injection volume was set to 20 µL. The method consisted of 2 minutes at 8 % A, increased to 20 % A in 20 minutes, further increased to 100 % A in 10 minutes, maintained at 100

% A for 5 minutes followed by a decrease of A to 8% in 5 minutes and a final maintain of A at 8 % for 5 minutes to give a 55 minutes method. The data was recorded using the diode array detector at specified wavelengths (280 and 340 nm). Data was exported as signal peak at respective times to Microsoft Excel spreadsheet.

2.7.3 Thin-layered chromatography

TLC Silica gel 60 F_{254} (Sigma) were used as the plates for their ability to absorb UV at 254 nm and fluoresce in the visible spectrum. The compounds ferulic acid, coniferyl alcohol and coniferyl aldehyde also absorbed at this wavelength, allowing spots to be observed on the plate. The TLC plates were marked with the starting position of the compound and 5 µL spots of the compound of interest was added. The samples were allowed to dry before placing the TLC plate in a beaker containing the mobile phase (Chloroform:Methanol:Formic acid, 85:14:1 respectively). The beaker was covered with a glass piece and the mobile phase was allowed to elute until 1-2 cm from the top. The TLC plate was then removed and air dried. To check for compounds, the TLC plate was exposed to UV light emission (λ=254 nm) and pictures were taken. To check for ferulic acid, the plate was submerged in 1 % ferric chloride solution and imaged.

2.8 Cell culture assays

2.8.1 pNPG assay

pNPG assay was used to determine the activity of the beta-glucosidases. 0.5 mL of cell culture was centrifuged (13,300 rpm, rt [room temperature], 10 minutes) and 10 µL of the supernatant (or blank medium) was added to a 1 mL Eppendorf tube. 100 µL of substrate solution containing 10 mM pNPG (p-nitrophenyl B-D-glucopyranoside) (Sigma) in 50 mM sodium acetate buffer pH 5.5 was added to the 1 mL tube to start the reaction. The mixture was incubated for 30 minutes at 45 ^{o}C. 1 mL of 0.6 M sodium carbonate was added to stop the reaction. The activity was determined by measuring the absorbance at 410 nm via a spectrophotometer indicated by the presence of p-nitrophenyl from pNPG as a result of the activity of the beta-glucosidases on the pNPG substrate.

2.8.2 Cellulose degradation and glucose concentration

Whatman filter paper no.1 was used as the cellulose substrate and 6mm disc shapes were cut out using a hole puncher. The discs were added to wells in a 24 well black, clear bottom plate (Eppendorf) and 1 mL of BGXY without the glycerol was added to create the media BGCY. Colonies were added at a starting OD_{600} of 0.6 to ensure exponential phase allowed sufficient production of proteins for growth. The OD_{600} of the plates were continually monitored for specified time lengths. Where applicable, glucose measurements were made using a HK glucose assay kit according to the manufacturer's instructions (Sigma, GAHK20).

2.8.3 Statistical tests

The significance of the differences between samples was calculated using a student's t-test (Kalpić et al., 2011), throughout the whole thesis. A difference was considered significant between the two samples when the p-value in the test was less than 0.05. A p-value greater than 0.05 resulted in no significance. The relative standard deviation (RSD) was obtained using the formula RSD=Standard deviation of sample/average of the sample. The growth constant (k) was obtained via the log of exponential growth using the formula Y=logY0 + k*X in the software package Graphpad (v8.4.3).

2.9 Strains and plasmids

2.9.1 Strains used in this study

Table 2.10: Escherichia coli strains

Strain name	**Collection number**	**Genotype**	**Remarks**	**Source**

Top 10F’	TBD	F′ [*lac*I^{q}, *Tn*10(Tet^{R})] *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K *rps*L(Str^{R}) *end*A1 *nup*G	Cre-recombinase plasmid cloning	INVITROGEN
Top 10	PTC46	F$^{–}$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ$^{–}$ *rps*L(Str^{R}) *end*A1 *nup*G	plasmid cloning	INVITROGEN

Table 2.11: S. cerevisiae strains

Strain name	**Collection number**	**Genotype**	**Source**
Wild-type	PTC107	*MATa leu2-3,112 his3-11,15 ade2-1 ura3-1 trp1-1 pep4::HIS3 prb1::hisg prc1::hisg*	(Tomimoto et al., 2013)

Table 2.12: Pichia pastoris strains

Strain name	**Collection number**	**Genotype**	**Phenotype**	**Source**
Wild-type	TBD	GS115, His4	Mut+	Thermofisher

2.9.2 Plasmids

Table 2.13: List of plasmids used in S. cerevisiae

Plasmid name	**Description**	**Markers (*E. coli, S. cerevisiae*)**	**Source**
pNM1-tt	Plasmid containing selection marker, His3 integration sites and multiple cloning site. This plasmid was used to construct the remaining (*S.c.* -) plasmids. The plasmid was linearised using EcoRV and PacI restriction enzymes for subsequent plasmid ligations or Gibson assembly	Amp, KanMX4	This study
pGS1	pNM1-tt plasmid that contains P_{TDH3} promoter controlling the expression of the mating alpha factor secretion signal fused to an mNeonGreen fluorescent protein gene followed by a T_{ADH1} terminator.	Amp, KanMX4	This study
pGS2	pNM1-tt plasmid that contains P_{PAB1} promoter controlling the expression of the mating alpha factor secretion signal fused to an mNeonGreen fluorescent protein gene followed by a T_{ADH1} terminator.	Amp, KanMX4	This study
pGS3	pNM1-tt plasmid that contains a synthetic $P_{UASC\text{-}core1}$ promoter controlling the expression of the mating alpha factor secretion signal fused to an mNeonGreen fluorescent protein gene followed by a T_{ADH1} terminator.	Amp, KanMX4	This study

pGS4	pNM1-tt plasmid that contains P_{PAB1} promoter followed by the 5'UTR from P_{TDH3} controlling the expression of the mating alpha factor secretion signal fused to an mNeonGreen fluorescent protein gene followed by a T_{ADH1} terminator.	Amp, KanMX4	This study
pGS6	pNM1-tt plasmid that contains $P_{UASC\text{-}core1}$ promoter followed by the 5'UTR from P_{TDH3} controlling the expression of the mating A factor secretion signal fused to an mNeonGreen fluorescent protein gene followed by a T_{ADH1} terminator.	Amp, KanMX4	This study
pGS7	pNM1-tt plasmid that contains P_{TDH3} promoter controlling the expression of the mating A factor secretion signal fused to an mNeonGreen fluorescent protein gene followed by a T_{ADH1} terminator.	Amp, KanMX4	This study
pGS8	pNM1-tt plasmid that contains P_{PAB1} promoter controlling the expression of the mating A factor secretion signal fused to an mNeonGreen fluorescent protein gene followed by a T_{ADH1} terminator.	Amp, KanMX4	This study

Table 2.14: List of plasmids for P. pastoris

Plasmid name	**Description**	**Markers (*E. coli, P. pastoris*)**	**Source**

pPGS10	Plasmid that contains P_{AOX1} promoter controlling the expression of the mating alpha factor secretion signal fused to an 6xhis-tagged *PaBG1b* (Uniprot ID: BAU51446) gene followed by a T_{AOX1} terminator.	Amp, Zeocin	This study
pPGS11	Plasmid that contains P_{AOX1} promoter controlling the expression of the mating alpha factor secretion signal fused to an 6xhis-tagged *ExG1* (Uniprot ID: P50401) gene followed by a T_{AOX1} terminator.	Amp, Zeocin	This study
pPGS12	Plasmid that contains P_{AOX1} promoter controlling the expression of the mating alpha factor secretion signal fused to an 6xhis-tagged *EnG1* (Uniprot ID: I0KXM4) gene followed by a T_{AOX1} terminator.	Amp, Zeocin	This study
pPGS13	Plasmid that contains P_{AOX1} promoter controlling the expression of the mating alpha factor secretion signal fused to an 6xhis-tagged *BG2* (Uniprot ID: A9F279) gene followed by a T_{AOX1} terminator.	Amp, Zeocin	This study
pPGS14	Plasmid that contains P_{AOX1} promoter controlling the expression of the mating alpha factor secretion signal fused to an 6xhis-tagged *ExG2* (Uniprot ID: I0H8B9) gene followed by a T_{AOX1} terminator.	Amp, Zeocin	This study

pPGS15	Plasmid that contains P_{AOX1} promoter controlling the expression of the mating alpha factor secretion signal fused to an 6xhis-tagged *BX1* (Uniprot ID: Q59277) gene followed by a T_{AOX1} terminator.	Amp, Zeocin	This study
pPGS16	Plasmid that contains P_{AOX1} promoter controlling the expression of the mating alpha factor secretion signal fused to an 6xhis-tagged *BG2* (Uniprot ID: Q11ST3) gene followed by a T_{AOX1} terminator.	Amp, Zeocin	This study
pPGS40	Plasmid design from Li et al., (2017) recreated in this study. Plasmid contains P_{AOX1}-MCS-T_{AOX1} expression site bound by two *loxp* sites (66 and 71). The cre-recombinase gene is regulated by P_{AOX1}. HIS4 is the sequence for homologous recombination. Restriction digested using AfeI and AvrII for subsequent ligations.	Amp, Zeocin	This study
pPGS60	Plasmid contains P_{GAP}-MCS-T_{AOX1} expression site bound by two *loxp* sites (66 and 71). The plasmid also contains cre-recombinase gene regulated by P_{FLD1} and GFP regulated by P_{GAP}. HIS4 is the sequence for homologous recombination. Restriction digested using AfeI and AvrII for subsequent ligations.	Amp, Zeocin	This study

pPGS80	Plasmid contains P_{GAP}-MCS-T_{AOX1} expression site bound by two *loxp* sites (66 and 71). The plasmid also contains cre-recombinase gene regulated by P_{FLD1} and iLOV regulated by P_{GAP}. *URA3* is the sequence for homologous recombination. Restriction digested using AfeI and AvrII for subsequent ligations.	Amp, Zeocin	This study
pPGS-CgHis	Plasmid contains genes CgAKR-1 and HisCar5 expression site regulated by PGAP bound by two *loxp* sites (66 and 71). The plasmid also contains cre-recombinase gene regulated by P_{FLD1} and GFP regulated by P_{GAP}. URA3 is the sequence for homologous recombination.	Amp, KanMX4	This study

Chapter 3 – Comparative assessment of the feasibility of achieving high levels of secretion of the components of a multi-enzyme pathway in *S. cerevisiae* and *P. pastoris*

3.1 Comparison of secretion in *P. pastoris* and *S. cerevisiae*

The optimisation of recombinant protein production or metabolic pathway expression relies on a host capable of suitably high levels of protein production. Moreover, where recombinant proteins are needed in the external medium, the host organism must be capable of appropriate rates of secretion. In this chapter, *S. cerevisiae* and *P. pastoris* are compared in a study to find the most suitable host for extracellular expression of enzymes on a cellulose degradation pathway. To compare the secretion capabilities of the two yeasts, a fluorescent reporter (ymNeonGreen) was utilised. Fluorescent proteins are easily assayable due to their excitation and emission at specific wavelengths, resulting in quantifiable fluorescence. The fluorescent protein was fused to an N-terminal secretion signal allowing extracellular secretion comparison between the two yeasts. Novel cellulase sequences provided by the collaborator Ingenza were extracellularly expressed in the respective yeast hosts to identify which would be most suitable for pathway optimisation.

3.1.1 *P. pastoris* and *S. cerevisiae* gene expression and secretion of ymNeonGreen

To compare the capacity for achieving high levels of secretion in *S. cerevisiae* and *P. pastoris*, YmNeonGreen (ymNG) was expressed in both yeasts and the abundance of the secreted form of this reporter protein product was estimated on the basis of the fluorescence intensity detected in the respective extracellular media. In order to achieve this, the *S. cerevisiae* mating factor alpha pre-pro-signal was fused to ymNeonGreen which allows sercretion of ymNeonGreen from the cytoplasm via the endoplasmic reticulum and Golgi to the extracellular medium of the cells (Brake et al., 1984; Lin-Cereghino et al., 2013). The fusion genes were codon-optimised for each respective organism, cloned into constitutive expression vectors, and transformed into *P. pastoris* and *S. cerevisiae.* For expression in *P. pastoris*, constitutive promtoer p*GAP* was utilised and for *S. cerevisiae*, constitutive promoter p*TDH3* was used.

Initially, the growth of both organisms was compared, since robust growth may be advantageous in protein production. This was followed by monitoring of ymNeonGreen production via fluorescence of cells. The cultures were grown for 24-hours with monitoring of optical density at absorbance 595 nm (OD_{595}) (Figure 3.1) and fluorescence (Figure 3.2) performed in a plate reader (every 14 minutes). *P. pastoris* transformed with YmNeonGreen (*P.p.* –ymNG) was found to have a faster growth rate compared to *S. cerevisiae* transformed with YmNeonGreen (*S.c.* –ymNG) based on OD_{595} measurements (compare the *P.p.*-ymNG and *S.c.* ymNG curves in Figure 3.1). At 24-hours, *S.c.* – ymNG had reached an OD_{595} of approximately 1.2 and *P.p.* – ymNG reached an OD_{595} of 1.6, the latter being approximately 1.3-fold higher (Figure 3.1). The strains *P.p.* –ymNG and *S.c.* –ymNG were also grown in 250 mL shake flasks from OD_{600} 0.6 for a period of 48-hours and the OD_{600} was obtained via a spectrophotometer. *P.p.* – ymNG was again found to grow to higher OD_{600} (Figure S3.1). *P.p.* – ymNG and *S.c.* – ymNG fluorescence measurements were made using excitation and emission wavelengths (λex/λem) of 488/525 nm to measure relative fluorescence units (RFU) for the same 24-hour period. *P. pastoris* and *S. cerevisiae* strains were transformed with vectors lacking ymNeonGreen (Ev) as controls. Both *S.c.* – ymNG and *P.p.* – ymNG strains reported much higher RFU values relative to their respective controls, *S.c.* – Ev and *P.p.* – Ev. The plate reader reached fluorescence saturation at around 56,000 RFU with *S.c.* – ymNG and *P.p.* – ymNG reaching the limit at approximately 14 and 17-hours, respectively (Figure 3.2).

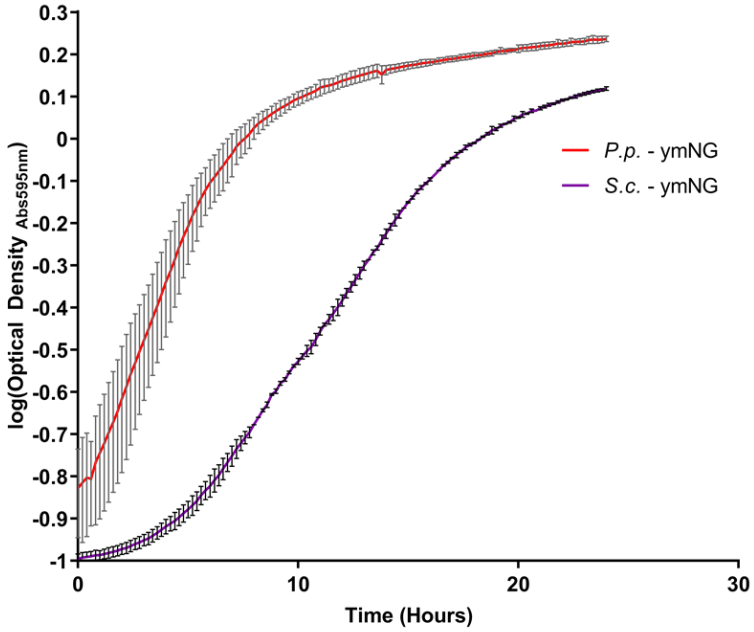

Figure 3.1: Growth comparison between S. cerevisiae and P. pastoris strains expressing ymNeonGreen. ***Growth*** *of P. pastoris (P.p. – ymNG) and S. cerevisiae (S.c. – ymNG)* expressing *mating alpha factor secretion signal fused to ymNeongreen when cultivated in YEPD at 30°C measured as the absorbance (optical density) continuously at 595 nm for 24-hours. The error bars represent standard deviation from 3 replicates.*

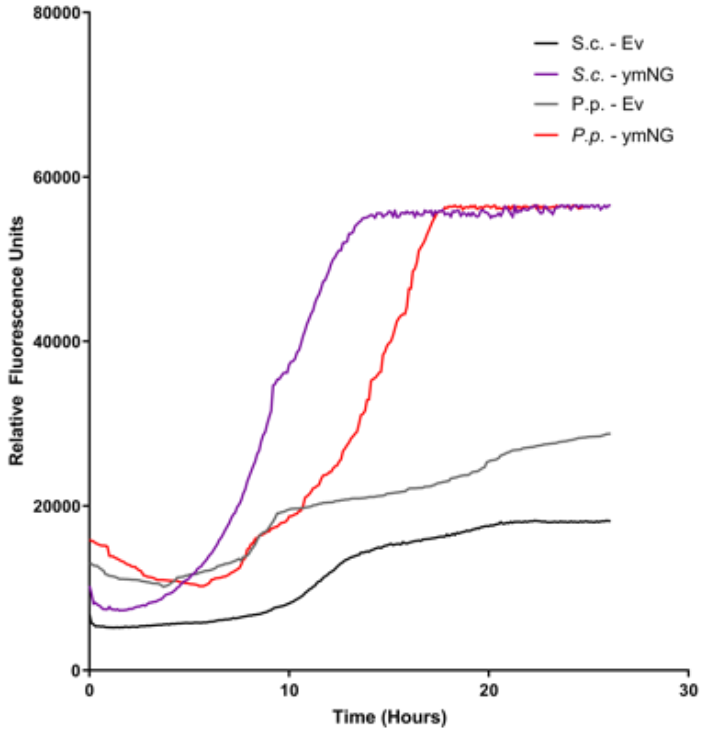

Figure 3.2: YmNeonGreen fluorescence comparison between S. cerevisiae and P. pastoris strains. Fluorescence of P. pastoris (P.p. – ymNG) and S. cerevisiae (S.c. – ymNG) expressing mating alpha factor secretion signal fused to ymNeonGreen when cultivated in YEPD at 30°C measured as relative fluorescence units (RFU) continuously for 24-hours at $\lambda_{ex}/\lambda_{em}$ of 488/525 nm respectively. Strains lacking the ymNeonGreen for P. pastoris (P.p. – Ev) and S. cerevisiae (S.c. – Ev) are shown as controls. Fluorescence saturation is due to the plate reader and not growth cessation.

Both organisms displayed the ability to synthesize ymNeonGreen (Figure 3.2), so the next step was to determine whether extracellular secretion of ymNeonGreen was occurring. In order to achieve that, *P.p.* – ymNG and *S.c.* – ymNG were grown for 24-hours and their fluorescence was detected in the form of 488 nm [excitation (λ_{ex})]/530 nm [emission (λ_{em})] spectra, respectively (Figure 3.1). After 24-hours of growth, the samples were centrifuged, and the supernatant was transferred to a new plate for fluorescence measurement. The extracellular fluorescence of *S.c.* – ymNG RFU was not significantly different compared to *S.c.* – Ev (p value > 0.05), whereas *P.p.* – ymNG produced significantly higher ymNG RFU (p value < 0.05) compared to its control of *P.p.* – Ev, (Figure 3.3) suggesting that *P. pastoris* was capable of ymNeonGreen secretion.

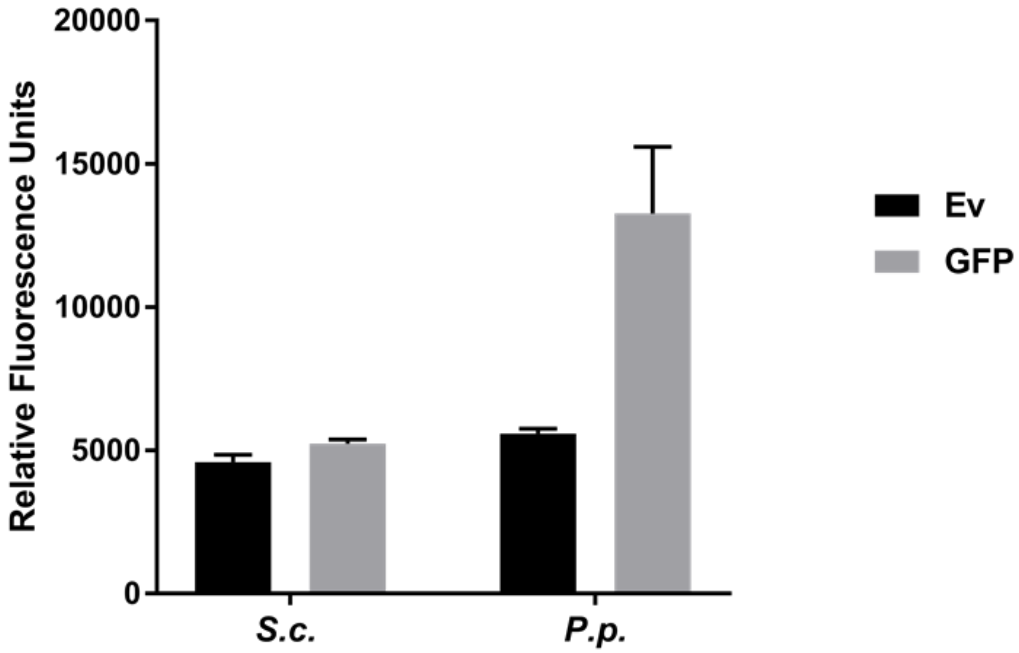

Figure 3.3: Extracellular ymNeonGreen fluorescence comparison between P. pastoris and S. cerevisiae. Fluorescence of extracellular media of P. pastoris (P.p. – ymNG) and S. cerevisiae (S.c. – ymNG) expressing mating alpha factor secretion signal fused to ymNeonGreen when being cultivated for 24-hours in YEPD at 30ºC measured as relative fluorescence units (RFU). The ex/em spectra was 488nm/525nm respectively. The RFU values are lower as this is extracellular media only suggesting a lot of the ymNG has been retained in the cells. The grey bars represent the ymNeonGreen expressing strain and black bars represent their respective empty vector controls and the error bars the standard deviation from 3 biological replicates.

To further confirm that the extracellular secretion detected in *P. pastoris* (Figure 3.3) was a result of secretion, *P.p.* – ymNG was grown for 24-hours and the cell culture

supernatant (1 mL) was isolated from the sedimented cell mass. The 10xhis-tagged ymNG was purified from the isolated supernatant using immobilised metal affinity chromatography (IMAC) via cobalt resin. Five elutions of the purified extracellular ymNeonGreen were made in 300 mM imidazole and analysed using SDS-PAGE gel. From all the elutions, the first two produced visible bands as visualised in Figure 3.4. The molecular weight of the mating alpha factor secretion signal and ymNeonGreen is approximately 9.4 kDa and 26.7 kDa respectively. Both *P.p.* – ymNG and *P.p.* – Ev produced similar bands, suggesting secretion of other native proteins from *P. pastoris*. However, for *P.p.* – ymNG, multiple bands were detected between approximately 25 kDa and 37 kDa in both elution 1 (E1) and elution 2 (E2). The band at approximately 26 kDa was excised and analysed via mass spectrometry which confirmed the protein sequence of YmNeonGreen with successful cleavage of the secretion signal peptide (Figure S3.2). With this finding it is concluded that ymNeonGreen is being secreted by *P. pastoris* (*P.p.* – ymNG) based on the levels of its detection in both intracellular and extracellular conditions.

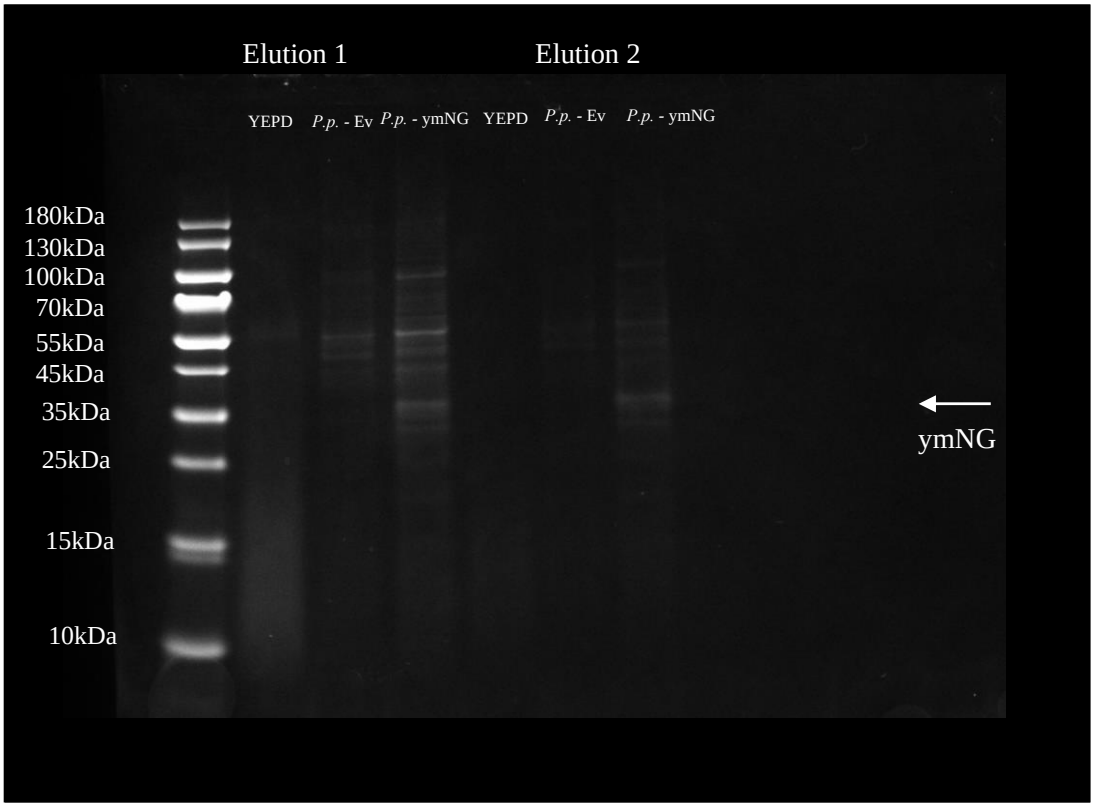

Figure 3.4: SDS-PAGE of extracellular media of P. pastoris expressing ymNeonGreen. P. pastoris (P.p. – ymNG) expressing mating alpha factor secretion signal fused to ymNG was cultivated in 20mL BMGY at 30°C for 24-hours. The 10xhis-tagged ymNG was purified using cobalt resin IMAC. 30 µL was loaded onto the gel. The blank (YEPD) is BMGY media only. P.p. – Ev is P. pastoris lacking ymNG. E1 and E2 are the two elutions when purifying ymNG. The approximate size of ymNG (26.7 kDa) is marked.

The lack of extracellular ymNeonGreen expression observed in *S. cerevisiae* (Figure 3.3), prompted further investigation. *S.c.* – ymNG samples were grown for 24-hours and centrifuged to separate the cell mass and cell culture supernatant. The sedimented cell mass underwent cell lysis to release intracellular proteins. The cell culture supernatant was concentrated to enhance protein purification. As the ymNeonGreen contained a 10x his-tag, it was purified using cobalt resin IMAC and various elutions of both the cell lysis and cell culture supernatant were made in 300mM imidazole. The elutions were visualised on SDS-PAGE (Figure 3.5). No bands were detected from *S.c.* – Ev (control) or *S.c.* – ymNG in any of the elutions (E1, E2, E3) of the cell culture supernatant. In the cell lysate, multiple bands were detected. The first 3 elutions were optimal, containing the greatest concentration of proteins, represented by the brighter bands. *S.c.* – ymNG had similar bands compared to the empty vector strain (*S.c.* – Ev), differing by one band at approximately 30 kDa. As this is similar to the molecular weight of ymNeonGreen (approximately 26.7 kDa), this suggested that ymNeonGreen was only detected intracellularly in *S. cerevisiae*.

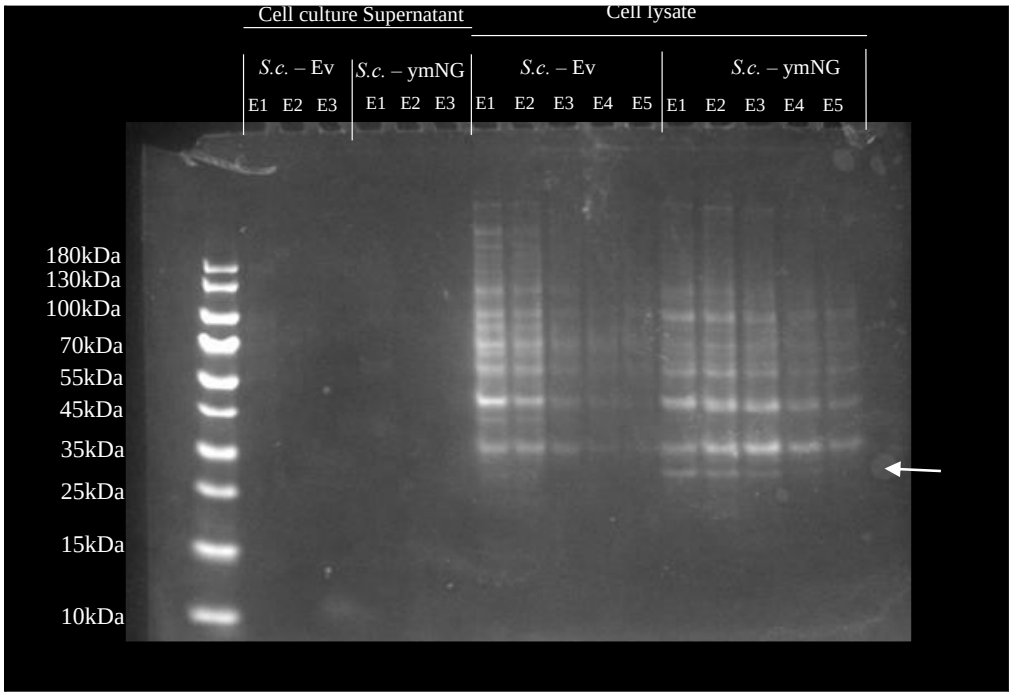

Figure 3.5: SDS-PAGE of S. cerevisiae expressing ymNeonGreen. S. cerevisiae (S.c. – ymNG) expressing mating alpha factor secretion signal fused to ymNeonGreen was cultivated in 20mL YEPD at 30ºC for 24-hours. The 10xhis-tagged ymNG was purified using cobalt resin IMAC. S.c. – Ev is S. cerevisiae lacking ymNeonGreen. E1 to E5 are the elutions when purifying ymNeonGreen to determine optimal elution number. The approximate size of ymNeonGreen (26.7kDa) is marked. Media supernatant is the purification of extracellular media only. Cell lysate is the purification from the cell lysate only.

In the effort to engineer extracellular expression of ymNeonGreen in *S. cerevisiae*, a range of different expression elements were incorporated into further constructs. The expression elements consisted of promoters (P_{TDH3}, P_{PAB1}, $P_{UASC\text{-}core1}$) and their 5'UTRs. These constructs differed in the expression elements (promoters and UTR's) coupled to ymNeonGreen or the secretion signal fused to the ymNeonGreen (Mating-a secretion signal). These changes are highlighted in Table 3.1. These constructs were all individually transformed into *S. cerevisiae.* Initially, their fluorescence was measured for intracellular expression of ymNeonGreen as well as to investigate the ability of the expression elements (promoters P_{TDH3}, P_{PAB1}, $P_{UASC\text{-}core1}$ and their 5'UTRs) to obtain a range of expression of ymNeonGreen. All strains were grown to exponential phase (OD_{600} 0.6) and the ymNeonGreen expression was measured via flow cytometry. The λex/λem wavelengths used were 488 /530 nm respectively. The amount of fluorescence detected per cell was recorded and plotted as a histogram (Figure 3.6).

Table 3.1: Table of plasmid constructs for secretion of ymNeonGreen

Sample	Promoter	5'UTR	Secretion Signal
S.c. – ymNG	*TDH3*	Native	Mating alpha factor
S.c. – pGS2	*PAB1*	*TDH3*	Mating-A factor
S.c. – pGS3	*PAB1*	Native	Mating alpha factor
S.c. – pGS4	*PAB1*	*TDH3*	Mating alpha factor
S.c. – pGS6	*UASC-core1*	Native	Mating alpha factor
S.c. – pGS7	*TDH3*	Native	Mating-A Factor
S.c. – pGS8	*UASC-core1*	*L0*	Mating alpha factor

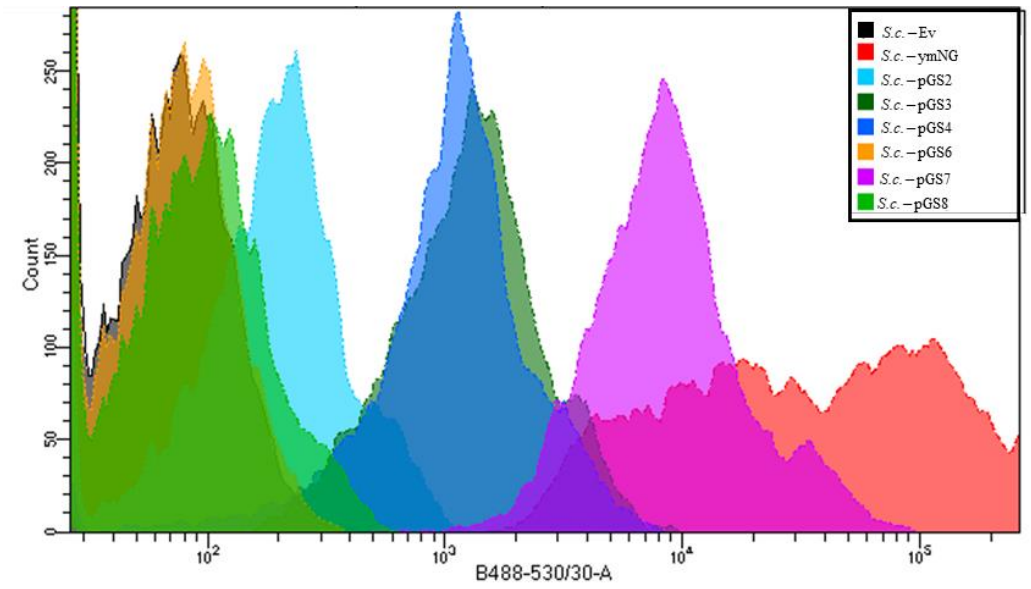

Figure 3.6: Fluorescence measurement from S. cerevisiae strains expressing ymNeonGreen. S. cerevisiae strain PTC107 was transformed with vectors containing ymNeonGreen (pGSx), under various expression controlling elements. S.c. – Ev contains empty vector and the rest of the constructs contain ymNeonGreen regulated by different expression elements where S.c. – ymNG contains TDH3 promoter, S.c. – pGS2 contains PAB1 promoter with Mating A factor with 3C protease site and a GC linker, S.c. – pGS3 contains PAB1 promoter with mating alpha factor signal,, S.c. – pGS4 contains PAB1 promoter with its 5'UTR replaced by TDH3 5'UTR, S.c. – pGS6 contains UASc-core1 promoter with 3C protease site and a GC linker, S.c. – pGS7 contains TDH3 promoter with MAT-A factor secretion signal and S.c. – pGS8 contains UASc-core1 promoter with L0 5'UTR, 3C protease site and a GC linker. Flow cytometry was carried out to measure intracellular fluorescence, exciting at 488 nm and measuring emission at 530 nm. All cells were grown to OD_{600} 0.6 and analysed. The fluorescence of each cell was measured and plotted it as a histogram. 10,000 total cells were measured per strain.

Constructs *S.c.* – pGS6 and *S.c.* – pGS8 produced similar fluorescence levels to that of the empty vector control (*S.c.* – Ev). All of the remaining constructs expressed ymNeonGreen at various levels as a result of the different expression elements. All the constructs that displayed expression of ymNeonGreen (Figure 3.6) were then grown up to 24-hours and centrifuged to isolate cell culture supernatant, which was subjected to $\lambda_{ex}/\lambda_{em}$ spectra of 488 nm/525 nm respectively. However, for all the constructs that were tested, extracellular fluorescence was not significantly different (p value > 0.05) from that obtained with the strain containing the empty vector (*S.c.* – Ev), suggesting a lack of ymNeonGreen secretion into the extracellular media (Figure 3.7). These changes in secretion signal from mating alpha factor (*S.c.* – ymNG) to Mating-A factor (*S.c.* – pGS2/-pGS7), promoters (*S.c.* – pGS2/-pGS3/-pGS6) as well as 5'UTR variations (*S.c.* – pGS4/-pGS8) all did not lead to extracellular secretion.

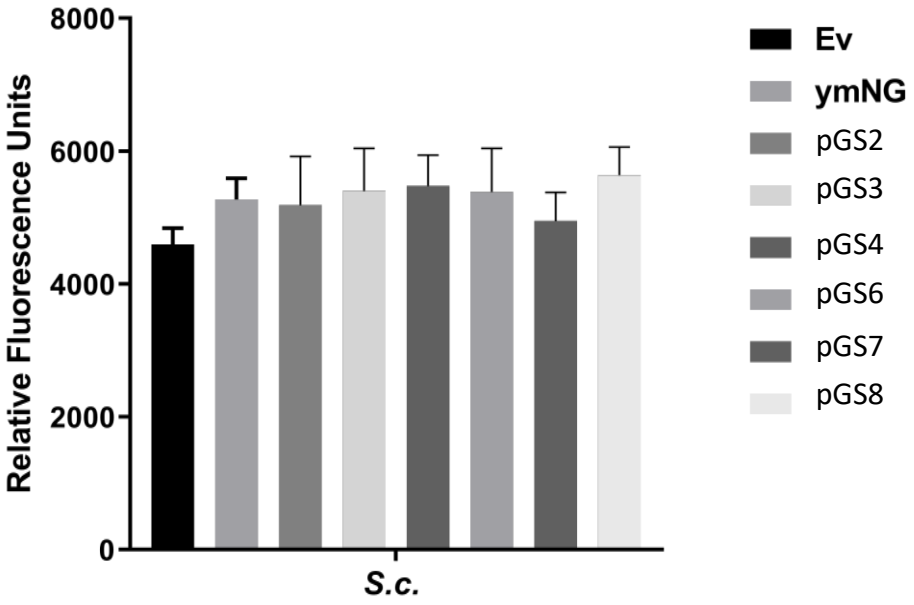

Figure 3.7: Extracellular ymNeonGreen fluorescence comparison between S. cerevisiae *strains. Fluorescence of extracellular media of S. cerevisiae transformed with pGS2, pGS3, pGS4, pGS6, pGS7 or pGS8 when being cultivated for 24 hours in YEPD at 30ºC measured as relative fluorescence units (RFU). The ex/em spectra was 488nm/525nm respectively. The black bars represent the empty vector control and the error bars the standard deviation from 3 biological replicates.*

As the change in expression elements did not lead to extracellular expression of ymNeonGreen, the growth conditions for strain *S.c.* – ymNG were then subjected to further variation in an attempt to obtain a possible optimum condition for ymNeonGreen secretion. This included varying the YEPD growth media pH (4, 4.5 5, 5.5, 6), the incubation temperature (25 ºC, 28 ºC, 30 ºC) and the growth time (18h, 24h, 30h). However, these changes did not result in detection of ymNeonGreen in the cell culture supernatant.

As the amount of expression is important in multi-enzyme pathways, being able to regulate expression in the host organism is essential. The range of promoters led to relative varied expression of ymNeonGreen, ranging from high (*S.c.* – ymNG) to medium (*S.c.* – pGS3/-pGS4) and low (*S.c.* – pGS2) (Figure 3.6). The addition of *TDH3* 5'UTR downstream of the *PAB1* promoter (*S.c.* – pGS4) compared to the wildtype *PAB1* promoter (*S.c.* – pGS3) did not result in secretion of ymNeonGreen (Figure 3.7). This concluded the attempts to try obtaining high levels of secretion of ymNeonGreen in *S. cerevisiae*. As *P. pastoris* had displayed the ability to secrete

ymNeonGreen, it was selected as the organism for attempting extracellular production of a multi-enzyme cellulose-degrading pathway.

3.2 Expression of a cellulose degradation pathway in *P. pastoris*

P. pastoris displayed both an ability to secrete ymNeonGreen as well as higher growth rates compared to *S. cerevisiae* and so was selected as the organism to produce a cellulose degradation pathway. The cellulose degradation pathway that was selected comprised three individual cellulase enzymes: ß-glucosidase, exo-glucanase and endo-glucanase. These enzymes act synergistically to degrade cellulose and a high extracellular concentration would be desirable. The cellulase protein sequences were provided by Ingenza and corresponded to as-yet incompletely characterized or non-characterised enzymes with no previous record of production in *P. pastoris*. The cellulases are listed in Table 3.2 alongside their molecular weight and theoretical isoelectric points (pIs). Within this group of cellulases, there were two ß-glucosidases (BG1, BG2), two exo-glucanases (ExG1, ExG2), one endo-glucanase (EnG1) and one ß-xylanase (BX1). The ß-glucosidase PaBG1b (uniprot: BAU51446) was selected as a control as it has been reported to have high activity and is easily assayed.

Table 3.2: Cellulases to be expressed in P. pastoris. (*) denotes cellulase not provided by Ingenza Ltd.

Referred name	**Cellulase Type**	**Organism origin**	**Uniprot ID**	**Theoretical PI**	**MW (Da)**	**Characterised**
PaBG1b*	β-glucosidase	*Panesthia angustipennis spadica*	BAU51446	5.5	57,509	Yes
ExG1	Exoglucanase	*Cellulomonas fimi*	P50401	5.24	85,349	No
EnG1	Endo-gucanases	*Micromonospora lupini str. Lupac 08*	I0KXM4	8.92	38,961	No
BG1	β-glucosidase	*Sorangium cellulosum*	A9F279	5.88	52,453	No
ExG2	Exo-glucanase	*Actinoplanes missouriensis*	I0H8B9	6.29	40,480	No
BX1	β-Xylanase	*Cellulomonas fimi*	Q59277	6.18	51,209	No
BG2	β-glucosidase	*Cytophaga hutchinsonii*	Q11ST3	5.87	83,660	No

To begin analysis, plasmid constructs were created with *P. pastoris* codon-optimised cellulase sequences fused to mating alpha factor secretion signal, to allow secretion into the extracellular media. All cellulases were fused to a 6xHis-tag. The plasmid constructs (Table 3.2) were individually transformed into *P. pastoris* and grown for 48-hours. The cell culture was centrifuged to isolate the cell mass and the cell culture supernatant. To enhance the purification of the cellulases, the cell culture supernatant was concentrated using 30 kDa filter columns (trapping proteins with molecular weights greater than 30 kDa). Purification was performed using nickel-based affinity (IMAC) columns. The cell culture supernatant was cycled 4 times through the nickel column using a peristaltic pump to allow affinity binding of proteins to the columns. The cellulases were concentrated into 500 µL elutions of 300mM imidazole. SDS-PAGE was used to analyse the content of the fractions (Figure 3.8). The control used was *P. pastoris* transformed with an empty vector (*P.p.* – Ev), showing no readily detectable bands. The *P. pastoris* strain expressing ß-glucosidase PaBG1b (*P.p.* – PaBG1b) produced bands in the first three elutions, suggesting that this enzyme was being secreted. However, *P. pastoris* strains expressing exo-glucanase (P50401) or endo-glucanase (I0KXM4), *P.p.* – ExG1 and *P.p.* – EnG1 respectively, showed no readily detectable bands. Instead, a smearing can be observed for these strains, suggesting possible degradation of both (Figure 3.8A). The ß-glucosidase (A9F27) containing strain *P.p.* – BG1 generated no bands or smearing (Figure 3.8B). Exo-glucanase (I0H8B9) expressing *P. pastoris* strain (*P.p.* – ExG2) showed bands at two sizes, one very intense band between the 70 kDa and 100 kDa marker size (possible dimerization) and a second faint band at around 40 kDa mark, closer to its actual size. The ß-xylanase Q59227 producing *P. pastoris* strain (*P.p.* – BX1) generated very strong bands near the 55 kDa mark and the ß-glucosidase Q11ST3 producing *P. pastoris* strain (*P.p.* – BG2) generated distinct strong bands around the 100 kDa mark (Figure 3.8B), close to the expected sizes of the respective cellulases. The bands from SDS-PAGE (Figure 3.8) of samples *P.p.* – PaBG1b, *P.p.* – ExG1, *P.p.* – BX1 and *P.p.* – BG2 were sent for sequence analysis via mass spectrometry. Their protein sequences were confirmed to be those of the expected cellulases, with successful cleavage of the mating alpha secretion signal (Table 3.3, Figure S3.3, S3.4, S3.5 and S3.6).

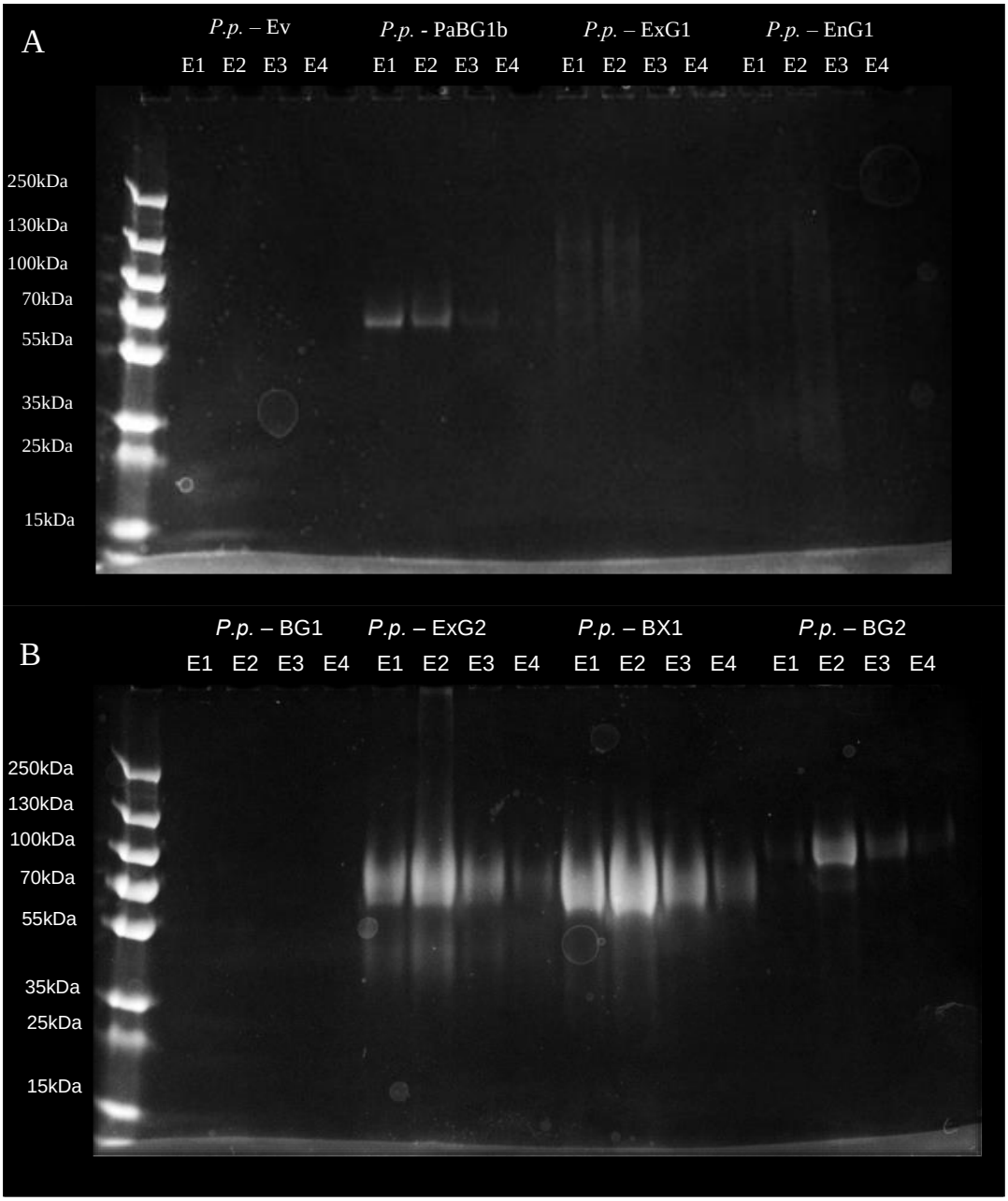

Figure 3.8: SDS-page of extracellular expression of cellulases in P. pastoris. P. pastoris was transformed with cellulases fused to mating alpha factor secretion signal for extracellular expression. The cellulases were as follows; (A) β-glucosidase BAU51446 (P.p. – PaBG1b), exo-glucanase P50401 (P.p. – ExG1), endo-glucanase I0KXM4 (P.p. – EnG1), (B) β-glucosidase A9F279 (P.p. – BG1), exo-glucanase I0H8B9 (P.p. – ExG2), β-xylanase Q59277 (P.p. – BX1) and β-glucosidase Q11ST3 (P.p. –BG2). The strains were grown for 48-hours at 30ºC in BMMY. Cell culture supernatant was isolated by centrifugation and subjected to protein purification and various elutions were made. The control was P. pastoris transformed with empty vector (P.p. – Ev).

Table 3.3: Cellulases protein identification

Referred name	Cellulase Type	MW (Da)	Peptide fragment coverage (%)	Protein Identification Probability (%)
PaBG1b*	β-glucosidase	57,509	40	100
ExG2	Exo-glucanase	40,480	53	100
BX1	β-Xylanase	51,209	61	100
BG2	β-glucosidase	83,660	60	100

The lack of bands or smearing in the supernatant from *P.p.* – ExG1, *P.p.* – EnG1 (Figure 3.8A) and *P.p.* –BG1 (Figure 3.8B) could be attributable to poor expression, protein degradation, or both, so attempts to optimise production were taken to try to obtain secretion of adequate amounts of material. The pH of the growth medium was buffered closer to the theoretical isoelectric point (pI) of the cellulases at pH 5.8 for *P.p.* – BG1, pH 5.2 for *P.p.* – ExG1 and pH 6.5 for *P.p.* – EnG1. The samples were grown for 24, 48 and 72-hours for *P.p.* – BG1 (Figure 3.8A) and for 24 and 48-hours for *P.p.* – ExG1 and *P.p.* – EnG1 (Figure 3.8B). However, there was no detection of bands and therefore an apparent lack of BG1 (Figure 3.9A), ExG1 and EnG1 (Figure 3.9C) secretion. A lower incubation temperature of 25^{o}C also did not yield detectable secretion of BG1 (Figure 3.8A), ExG1 or EnG1 (Figure 3.9C) by *P. pastoris* as confirmed by the absence of bands in all of the elutions. To eliminate the possibility of purification having a negative effect on detection, cell culture supernatant was directly visualised (S) on the SDS-PAGE; however, no bands were detected. As ExG2 from the *P.p.* – ExG2 supernatant was found to be forming two bands at around 45 and 85 kDa (Figure 3.8B), growth parameters were further optimised. The pH was increased as much as possible without affecting growth, to pH 6.5, and a lower incubation temperature of 25^{o}C was used. The strain was grown for 48-hours and cell culture supernatant was analysed again via SDS-PAGE. However, the two bands were detected again at approximately the same size (Figure 3.9B) as previously (Figure 3.8B), suggesting possible dimerization. *P.p.* – BG2 was also subjected to growth parameter optimisation (pH 5.8, 28^{o}C, 48-hours growth) and formed a clear singular band at around 100 kDa, approximating to the expected size (Figure 3.9D).

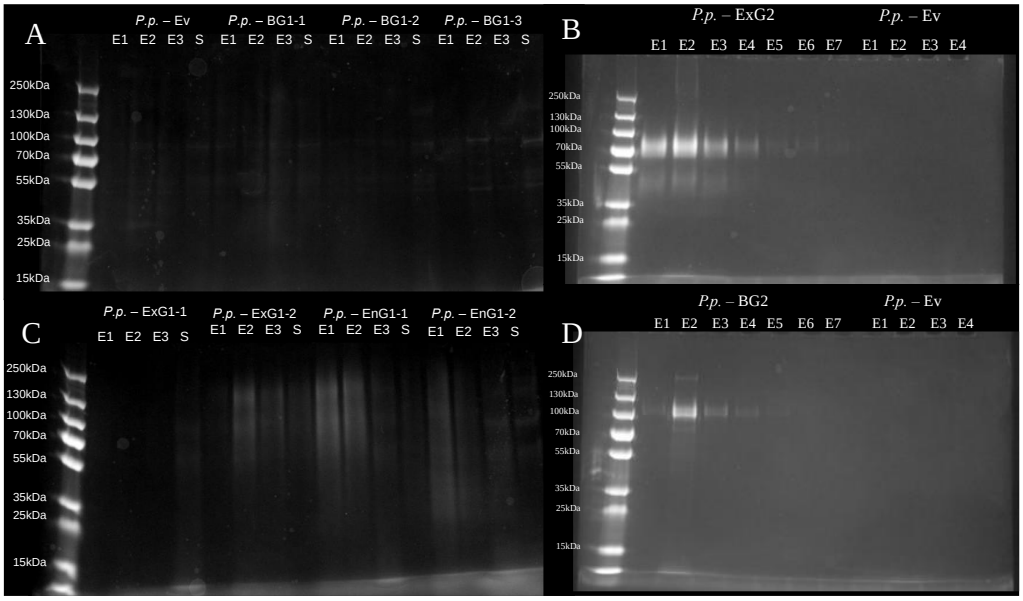

Figure 3.9: SDS-page of further extracellular expression of cellulases in P. pastoris. P. pastoris was transformed with the cellulase fused to mating alpha factor secretion signal for extracellular expression. The cellulases were as follows: β-glucosidase A9F279 (P.p. – BG1), exo-glucanase P50401 (P.p. – ExG1), endo-glucanase I0KXM4 (P.p. – EnG1), exo-glucanase I0H8B9 (P.p. – ExG2) and β-glucosidase Q11ST3 (P.p. –BG2). The strains were grown under various conditions (24-hours-30ºC, 48-hours-25ºC, 72-hours-30ºC) in BMMY pH 5.8 for P.p. – BG1 (A), pH 6.5, 48-hours – 28ºC for P.p. – ExG2 (B), pH 5.2 for P.p. – ExG1 and pH 6.5 for P.p. – EnG1 (C) and pH 5.8, 48-hours-28 ºC for P.p. – BG2 (D). Cell culture supernatant was isolated by centrifugation and subjected to protein purification and various elutions (E) as well as (A,C) cell culture supernatant (S) were visualised. The control was P. pastoris transformed with empty vector (P.p. – Ev).

3.2.1 Exploring the activity of cellulases in extracellular growth media

The cellulases under investigation are required to be secreted into an extracellular medium that is designed to support the growth of *P. pastoris*. If left in that extracellular medium, their ability to degrade cellulose would be expected to be reduced relative to conditions that may be their optimal activity and stability conditions (Table 3.2). To determine the activity of the cellulases in the conditions described here, cellulose-producing *P. pastoris* strains were grown for 48-hours and cell culture supernatant was isolated via centrifugation. Small volumes of the supernatant were added to a number of different assays in order to assess their activity. ß-glucosidases are the final enzymes in the pathway that catalyses the hydrolysis of glycosidic bonds in oligosaccharides to release glucose. A simple assay to measure ß-glucosidase activity is via the catalysis of conversion of the substrate p-nitrophenyl-β-glucopyranoside (pNPG) to β-glucopyranoside and p-nitrophenyl (PNP), which produces a yellow colouration where the intensity of the colour correlates to the amount of PNP present and therefore the activity of the enzyme. By measuring the absorbance of the colouration at 405nm, the activity of ß-glucosidases can be measured.

BG1 in the *P.p.*–BG1 supernatant and BG2 in the *P.p.*–BG2 supernatant were analysed, using the activity of PaBG1b (Li et al., 2017) from the *P.p.*–PaBG1b supernatant as a reference. For the assays, strains were grown for 48-hours, cell culture supernatants were isolated, and 20 µl of each cell culture supernatant was added to 180µl of 5 mM pNPG substrate in pH 5.5 sodium acetate buffer (Figure 3.10). The pNPG activity deduced from the changes in absorbance (at 405 nm) observed using the *P.p.* – BG1 supernatant was not significantly different (p value > 0.05) to that of the empty vector control (*P.p.*–Ev) supernatant, validating the lack of detection of an extracellular protein band described previously (Figure 3.9A). The BG2 activity detected in the *P.p.*–BG2 supernatant was approximately 2-fold higher pNPG activity (p value < 0.05) compared to the *P.p.*–Ev supernatant. The PaBG1b activity in the *P.p.*–PaBG1b supernatant was approximately 20-fold higher activity (p value > 0.05) compared to the *P.p.* – Ev supernatant.

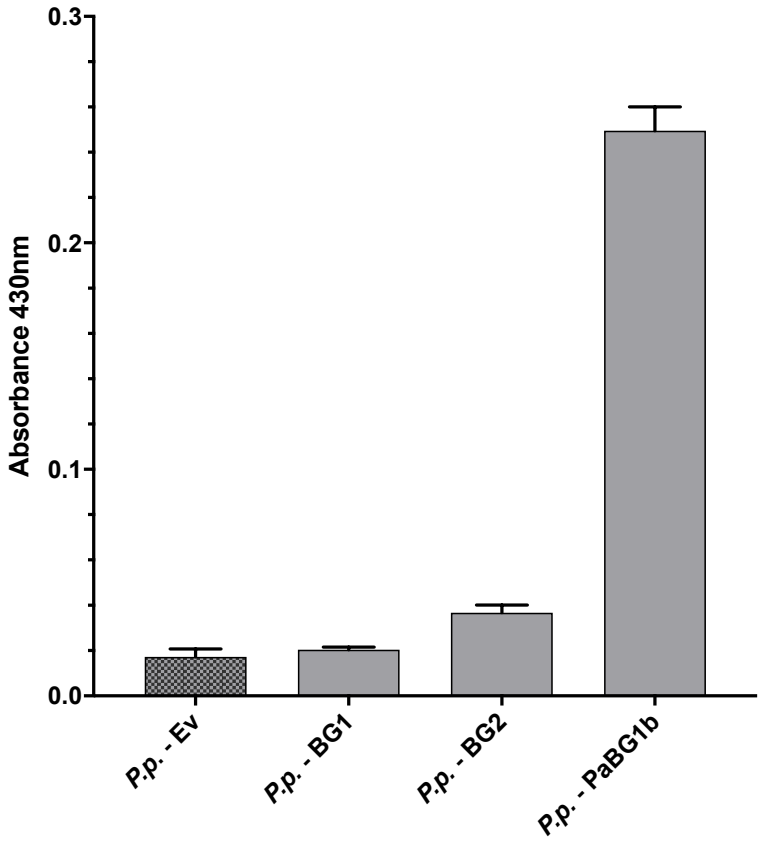

Figure 3.10:Expression of β-glucosidases in extracellular conditions using the pNPG assay. P. pastoris was transformed with the cellulase fused to the mating alpha factor secretion signal for extracellular expression. The cellulases were as follows: β-glucosidase A9F279 (P.p. – BG1), β-glucosidase Q11ST3 (P.p. –BG2) and β-glucosidase BAU51446 (P.p. – PaBG1b). The strains were grown for 24-hours at 30ºC in BMMY pH 5.5. Cell culture supernatant was isolated by centrifugation and 20 µL aliquots were added to 180 µL 5mM pNPG in pH5.5

sodium acetate buffer. The reaction was stopped after 10 minutes by addition of 800 µL 1M sodium carbonate. Absorbance was measured at 430nm. The control was P. pastoris transformed with empty vector (P.p. – Ev). Error bars represent the standard deviation from 3 biological repeats.

Cellulose degradation results in the production of glucose and requires the synergistic actions of all three cellulases. Their expression ratios therefore need to be adjusted to optimise activity. To investigate whether differing the ratio of each cellulase has an effect of glucose production as a result of synergism, the highly expressing and active cellulases ExG1 and BG2 were selected and their ratios (via cell culture supernatant volume) was altered. This change in glucose concentration would be monitored and used as an initial indicator of activity and of the need for optimisation of the enzyme expression ratio. The cellulases producing strains, *P.p.* – BG1 and *P.p.* – ExG2, were grown for 48-hours and the cell culture supernatants were isolated. The volume of cell culture supernatant added to the reaction mixture would be used as the indicator of varying the ratio of the two enzymes.

A filter paper (6mm disc) was used as the source of cellulose in the assay and added to the reaction for 1-hour at 30^{o}C. A glucose standard curve was created using a glucose HK kit (Sigma) that correlated glucose concentration (mg/mL) to absorbance (340 nm) (Figure S3.8) allowing the measurement of glucose concentration in the supernatant samples via absorbance. The equation of the line was obtained from this and allowed to calculate the glucose concentration produced from the various cellulase mixtures from the resulting absorbance data (Figure 3.11). Various combinations from the *P.p.*–BG2 and *P.p.*–ExG2 supernatants were made. However, only the ratios that had a significant difference (p value > 0.05) were displayed in Figure 3.10. The supernatants from *P.p.*–ExG1 and *P.p.*–BG2 were added separately to display the need for the synergistic actions of both enzymes for activity. Alongside the empty vector control supernatant (*P.p.*–Ev, *P.p.*–BG2 and *P.p.*–ExG2 did not result in the formation of glucose after 1-hour. When varying the ratios of the cellulases, an increase in the volume of *P.p.*–BG2 compared to *P.p.*–ExG2 was found to give the highest rate of glucose generation (mg/ml) up to a limit of 4x, after which no significant increase was noticed. Increasing the *P.p.*–ExG2 volume led to a decrease in the rate of glucose production.

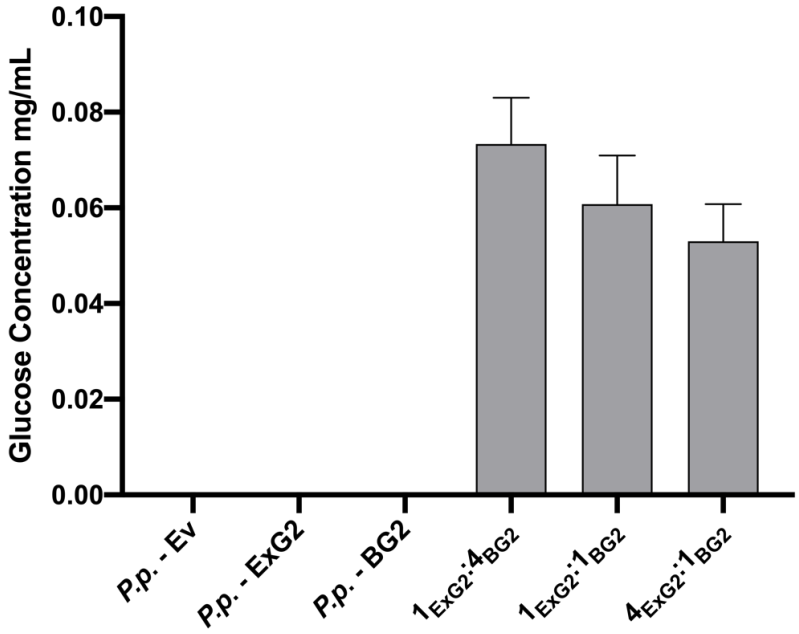

Figure 3.11: Filter paper assay for ratio variation induced glucose formation in P. pastoris expressing cellulases. P. pastoris was transformed with the exo-glucanase I0H8B9 (P.p. – ExG2) and β-glucosidase Q11ST3 (P.p. –BG2) cellulases fused to mating alpha factor secretion signal for extracellular expression. The strains were grown for 48-hours-30°C in BMMY pH 6. Cell culture supernatant was isolated by centrifugation and up to 50 µL was added to reaction mixture alongside 6mm filter paper discs. Glucose concentration was calculated using an equation derived from a glucose standard curve. The control was P. pastoris transformed with empty vector (P.p. – Ev). The error bars represent the standard deviation from three biological replicates.

3.3 Improving Secretion in *P. pastoris*

To improve secretion of the recombinant proteins (cellulases) in *P. pastoris,* some changes were introduced. The first change was the study of the expression systems using constitutive and regulated promoters, respectively. Two *P. pastoris* strains were created containing either the ß-glucosidase (*PaBG1b*) fused to mating alpha factor secretion signal driven by the glycerylaldehyde 3-phosphate dehydrogenase constitutive promoter (*GAP*) (*P.p.–PaBG1b*-P_{GAP}) or by the methanol-induced *AOX1* promoter (*P.p. –PaBG1b*-P_{AOX1}). In both cases, samples were grown for 24-hours, cell culture supernatant was isolated via centrifugation and supernatant was added to the substrate, and the absorbance (430 nm) was monitored (Figure 3.12). Both systems were found to produce activity. However, the methanol-regulated *AOX1* promoter was found to generate approximately a 2-fold higher (p value > 0.05) activity compared to that of the *GAP* promoter.

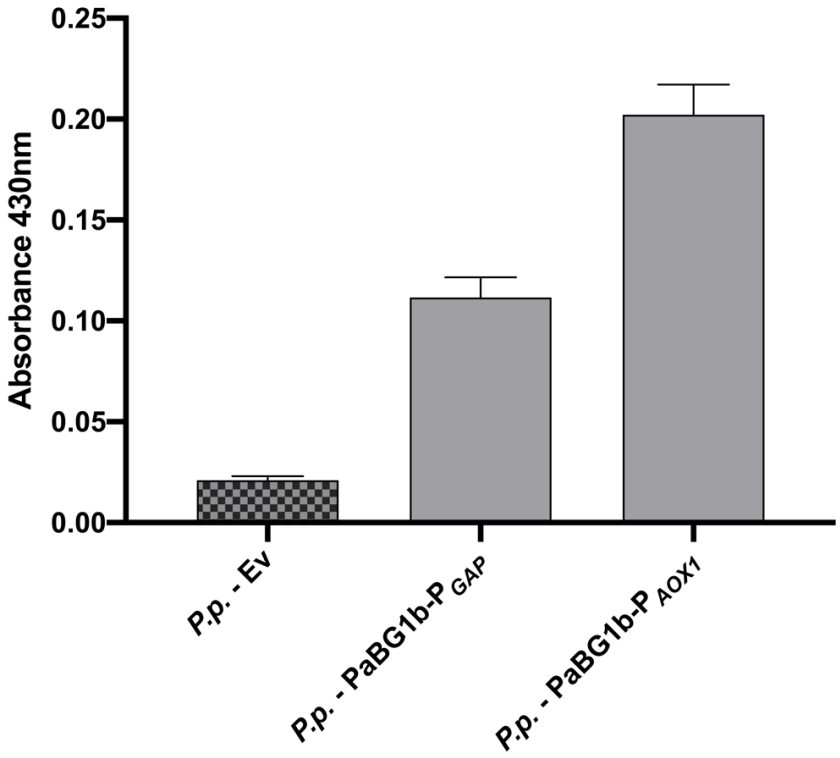

Figure 3.12:Comparison of constitutive against regulated promoter expression. P. pastoris was transformed with the cellulase ß-glucosidase BAU51446 fused to mating alpha factor secretion signal, driven by either the constitutive GAP promoter (P.p. – PaBG1b-P_{GAP}) or by the methanol regulated AOX1 promoter (P.p. – PaBG1b-P_{AOX1}) for extracellular expression. The strains were grown 24-hours at 30ºC in BMMY pH 5.5. Cell culture supernatant was isolated by centrifugation and 20 µL aliquots were added to 180 µL 5mM pNPG in pH5.5 sodium acetate buffer. The reaction was stopped after 10 minutes by addition of 800 µL 1M sodium carbonate. Absorbance was measured at 430nm. The control was P. pastoris transformed with empty vector (P.p. – Ev). Error bars represent standard deviation from 3 biological repeats.

The second change was the secretion signal utilised. The secretion signal plays an important role in transporting the proteins out of the cell and various improvements have been published. To study whether secretion improvements could be achieved, a selection of secretion signals were identified (Table 3.4) that may improve secretion in *P. pastoris*. To compare these secretion sequences, three plasmid constructs were created that contained the ß-glucosidase PaBG1b as a reporter for secretion. *P. pastoris* transformed with mating alpha factor fused to PaBG1b (*P.p.*–PaBG1b-S_{MAT}), the mating alpha factor **Δ**57-70 fused to PaBG1b (*P.p.*–PaBG1b- $S_{MAT\Delta 57\text{-}70}$), the mating alpha factor **Δ**30-43 fused to PaBG1b (*P.p.*–PaBG1b-$S_{MAT\Delta 30\text{-}43}$) and the Pre-Ost1-Pro hybrid secretion signal fused to PaBG1b (*P.p.*–PaBG1b-S_{POP}) were all grown for 24-hours from a starting OD_{600} of 0.6. The supernatants were then analysed using the pNPG assay (Figure 3.13). Mating alpha factor **Δ**57-70 (*P.p.*–PaBG1b-$S_{MAT\Delta 57\text{-}70}$), and **Δ**30-43 (*P.p.*–PaBG1b-$S_{MAT\Delta 30\text{-}43}$) were not found to be significantly different (p value > 0.05) compared to the wild type mating alpha factor secretion

signal (*P.p.* – PaBG1b-S_{MAT}). However, the Pre-Ost1-Pro (*P.p.* – PaBG1b-S_{POP}) was found to support approximately 1.8-fold higher activity (p value< 0.05) compared to the wild type mating alpha factor secretion signal (*P.p.* – PaBG1b-S_{MAT}). Although this enhancement in secretion was smaller than that reported previously for this secretion sequence (Barrero et al., 2018), it was still significantly better than that achieved in this work using the wild-type version. Therefore, this became the preferred secretion signal in future constructs for extracellular expression.

Table 3.4: Secretion signals to improve secretion in P. pastoris

Secretion Signal	Additional info	Source
Mating alpha factor	Wild-type secretion signal	*S. cerevisiae* genome
Mating alpha factor **Δ**57-70	Mating alpha factor secretion signal with deletion of amino acids 57-70	(Lin-Cereghino et al., 2013)
Mating alpha factor **Δ**30-43	Mating alpha factor secretion signal with deletion of amino acids 30-43	(Lin-Cereghino et al., 2013)
Pre-Ost1-Pro	Mating alpha factor Pre and Pro region fused to Ost1 signal sequence	(Barrero et al., 2018)

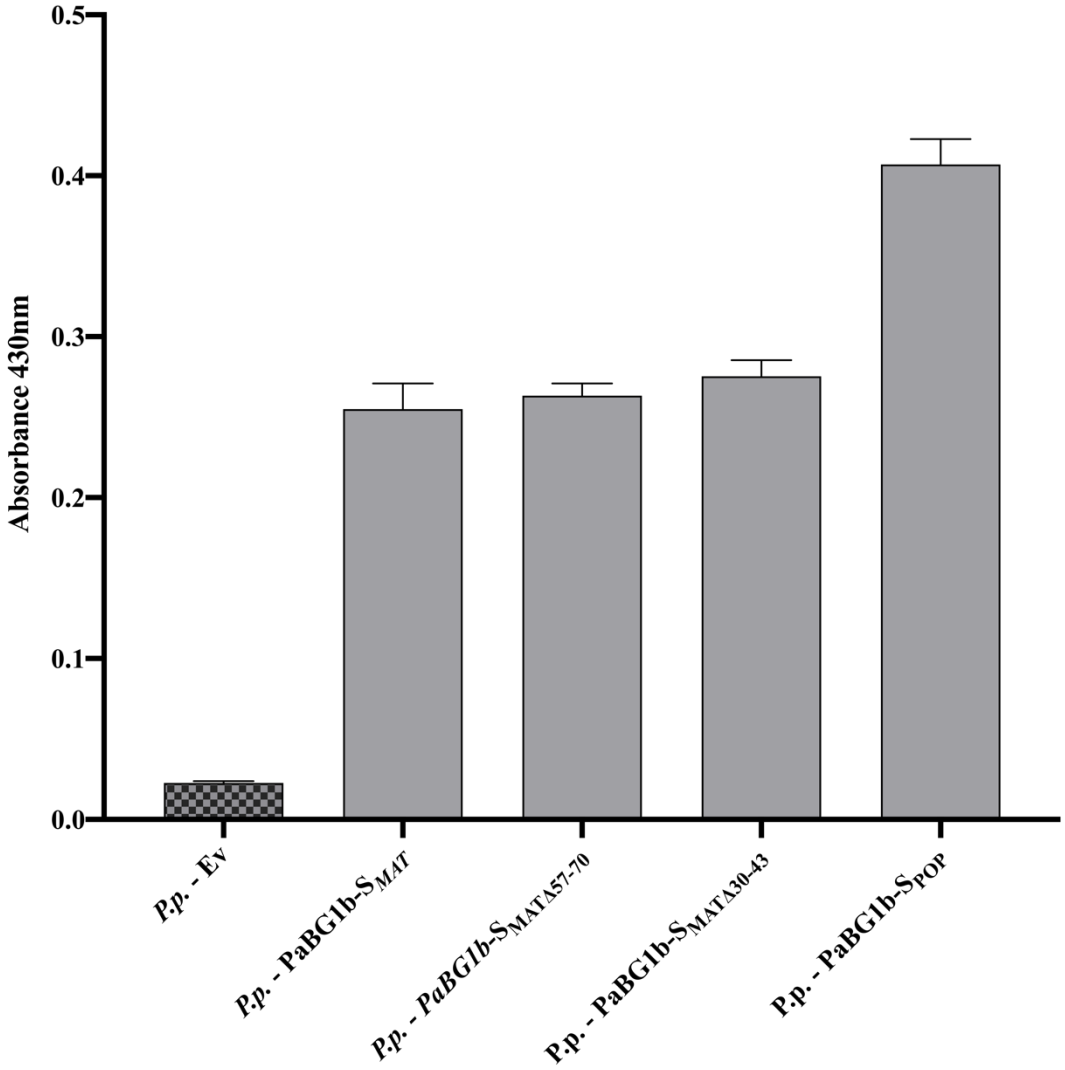

Figure 3.13: Comparison of secretion signals in p. pastoris. P. pastoris was transformed with β-glucosidase PaBG1b where the secretion signal fusion was varied. The strains contained the following: mating alpha factor strain (P.p. – PaBG1b-S_{MAT}), the modified mating alpha factor strain (P.p. – PaBG1b-$S_{MAT\Delta 57\text{-}70}$), the modified mating alpha factor strain (P.p. – PaBG1b-$S_{MAT\Delta 30\text{-}43}$) and the Pre-OST1-Pro hybrid secretion tag (P.p. – PaBG1b-S_{POP}). The control is P. pastoris transformed with PaBG1B containing no secretion tag (P.p. – Ev). The strains were grown 24-hours at 30°C in BMMY pH 5.5. Cell culture supernatant was isolated by centrifugation and 20 µL aliquots were added to 180 µL 5mM pNPG in pH5.5 sodium acetate buffer. The reaction was stopped after 10 minutes by addition of 800 µL 1M sodium carbonate. Absorbance was measured at 430nm. Error bars represent standard deviation from 3 biological repeats.

In summary, *P. pastoris* and *S. cerevisiae* were compared as hosts for extracellular protein production using the reporter ymNeonGreen. *P. pastoris* was found to be able to secrete significant amounts of ymNeonGreen as measured via extracellular fluorescence and SDS-PAGE. After being selected for its ability to express ymNeonGreen, various cellulose genes of interest were expressed using secretion-sequence-fusions. From the 6 cellulases of interest, 3 were found to be synthesised at high levels. These three were further analysed for their activity in the extracellular

growth medium and found to be active and able to degrade specific substrates, including filter paper (cellulose). The next step would be to integrate them into a single yeast strain.

Chapter 4 – Development of a cloning plasmid suited for integration of multi-gene clusters into *P. pastoris*.

The current limitations and lack of research, as well as copyright issues in genomic integration methods (such as CRISPR/cas9) in *Pichia pastoris* encourages the use of alternative integration strategies for *P. pastoris*, especially in the context of biotech industrial research. As CRISPR/cas9 policies prevents the use of the integration method without appropriate licensing agreements, a cre-recombination plasmid from Li et al. (2017) was identified from literature research as a potentially useful tool for multi-gene integration. However, this plasmid was further modified in order to make it compatible with a high-throughput combinatorial strategy for metabolic pathway optimisation. The work in this chapter outlines the development of a plasmid that allows single and multi-gene integration into *P. pastoris.* Furthermore, I aimed to develop a combinatorial strategy utilising recombination of *lox*p sites via cre-recombinase, robotics for high-throughput assaying and fluorescence reporters for colony selection. As time, cost and labour are important factors in the biotech industry, this approach could facilitate high-throughput metabolic pathway genomic integration and optimisation.

4.1 Improvement in marker-recycling plasmids in *P. pastoris*

Antibiotic marker recycling via cre-recombinase is an alternative method to CRISPR that allows for multi-gene pathway integration as well as genomic integration of multiple copies of a single gene into *P. pastoris*. A plasmid that enables recycling of selection markers via cre-recombinase (Chen et al., 2017) was chosen as the starting point for this work.

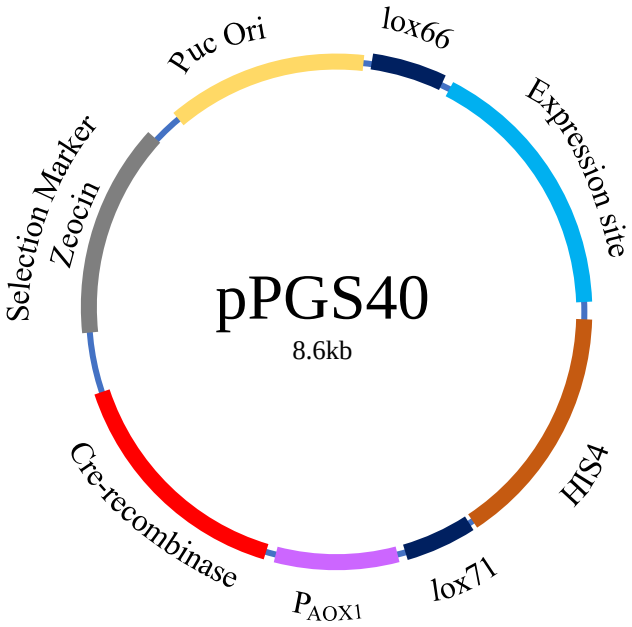

Figure 4.1: Design of marker recycling, cre-recombinase plasmid, pPGS40. The plasmid was designed by Chen et al. (2017) and developed here. The expression site contains a promoter, gene of interest and terminator. Genomic integration occurs via homologous recombination using HIS4. Cre-recombinase is induced by methanol activated promoter AOX1 leading to expression of cre-recombinase, resulting in the recombining of lox66 and lox71 sites and excision of DNA. The expression site and HIS4 remain genomically integrated into the P. pastoris genome.

The plasmid pPGS40 was constructed in this work according to the published design from (Li et al., 2017) (Figure 4.1), using Gibson assembly cloning, and then transformed into *Escherichia coli* TOP10 strain. Colony PCR of *E. coli* transformed with pPGS40 was carried out and positive colonies were identified. Positive colonies were grown overnight and miniprepped to isolate the plasmid, which was visualised on an electrophoresis gel (Figure 4.2A). Multiple bands were observed at over 10 kb, at 8 kb (approximately to the size of the plasmid, 8.6 kb) and around 3 kb. To investigate the reason for these multiple bands, pPGS40 was linearized using restriction enzyme AgeI, cutting at the cre-recombinase gene. Upon linearization and agarose gel electrophoresis analysis, a strong band was detected at around 5 kb size (Figure 4.2A). The linearised plasmid was sent for Sanger sequencing and analysis of the sequences showed that recombination of *lox*66 and *lox*71 sites had occurred to form the recombined mutant *lox*72 site. This resulted in the excision of the DNA sequence in between the *lox*66 and *lox*71 (the expression site and *HIS4* integration site), leaving the recombinant form of the plasmid, named pPGS40r (Figure 4.2B). To overcome this issue, the total plasmid DNA from the miniprep was visualised on an electrophoresis gel and the band close to 9kb (the size of pPGS40) was excised.

However, the process of excising the pPGS40 from the electrophoresis gel reduces the ability of directly transforming the plasmid (pPGS40) into *P. pastoris*, due to the presence of pPGS40r in the sample. The gel electrophoresis excision also adds an extra step in the correct identification and thus the purification of the plasmid pPGS40.

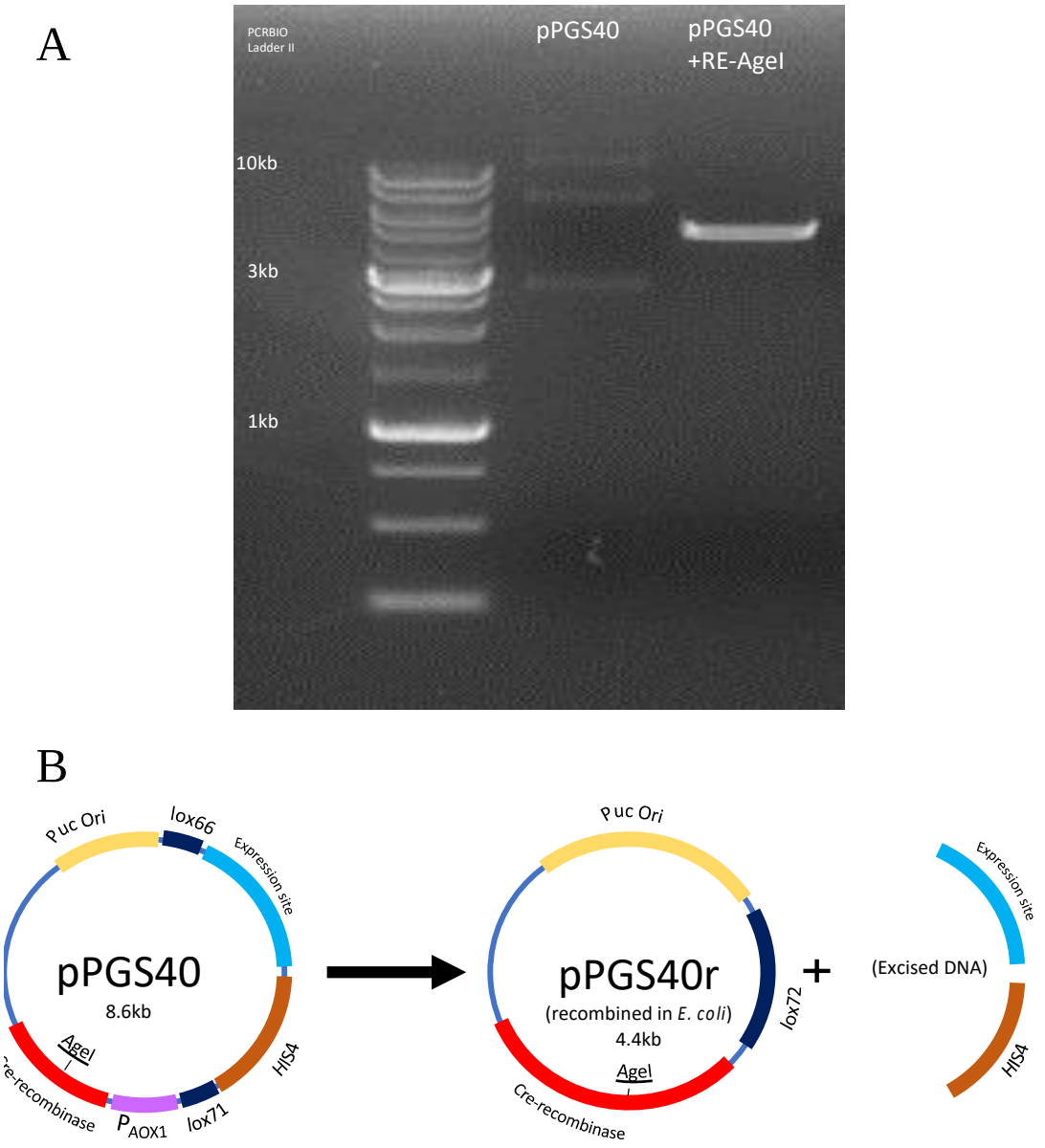

Figure 4.2: pPGS40 recombining in E. coli transformation. pPGS40 was transformed into E. coli Top10 competent cells and positive colonies were selected and cultured overnight. pPGS40 and PPGS40 linearized with AgeI (pPGS40+RE AgeI) were visualised by electrophoresis (A). The band was sequenced and confirmed to show that recombination had occurred (pPGS40r) and DNA between the lox sites had been excised (B).

Pan et al. (2011) suggested that undesired recombination of the plasmid pPGS40 observed in *E. coli* may be a result of the expression of promoter *AOX1* (P_{AOX1}) in *E. coli*. To investigate whether the undesired recombination in *E. coli* could be halted, changes were introduced (Figure 4.3A). A shorter overnight incubation period of *E. coli* transformed with pPGS40 (9hr-Inc) was carried out on the basis that cre-recombinase expression may be induced in the later stages of overnight cell cultures as suggested by Li et al. (2017). The second change was the addition of the *lac* operon operator (lacO) downstream of P_{AOX1} and upstream of cre-recombinase gene (pPGS40+lacO). LacO addition would act as a binding site for the repressor protein LacI resulting in the blocking of downstream expression of P_{AOX1} (Becker et al., 2013). Another change implemented was the substitution of the *AOX1* promoter with the *FLD1* promoter in the hope that P_{FLD1} may not be active in *E. coli* (pPGS40+P_{FLD1}). The promoter *FLD1* was identified as it is also methanol induced, resulting in the selective induced expression of the cre-recombinase in *P. pastoris*. To investigate whether more than one of the above steps may be necessary to prevent the recombination of pPGS40, a combination of all of the steps was also implemented (pPGS40+Combi).

E. coli was transformed with the respective pPGS40 variants (incorporating the changes highlighted) and positive colonies were set up as overnight cell cultures. The cell cultures were miniprepped the following morning to isolate pPGS40 plasmid variants and to visualise them by electrophoresis (Figure 4.3B). The changes implemented produced multiple bands that contained a mixture of pPGS40 and pPGS40r, suggesting recombination was still occurring, as confirmed by Sanger sequencing.

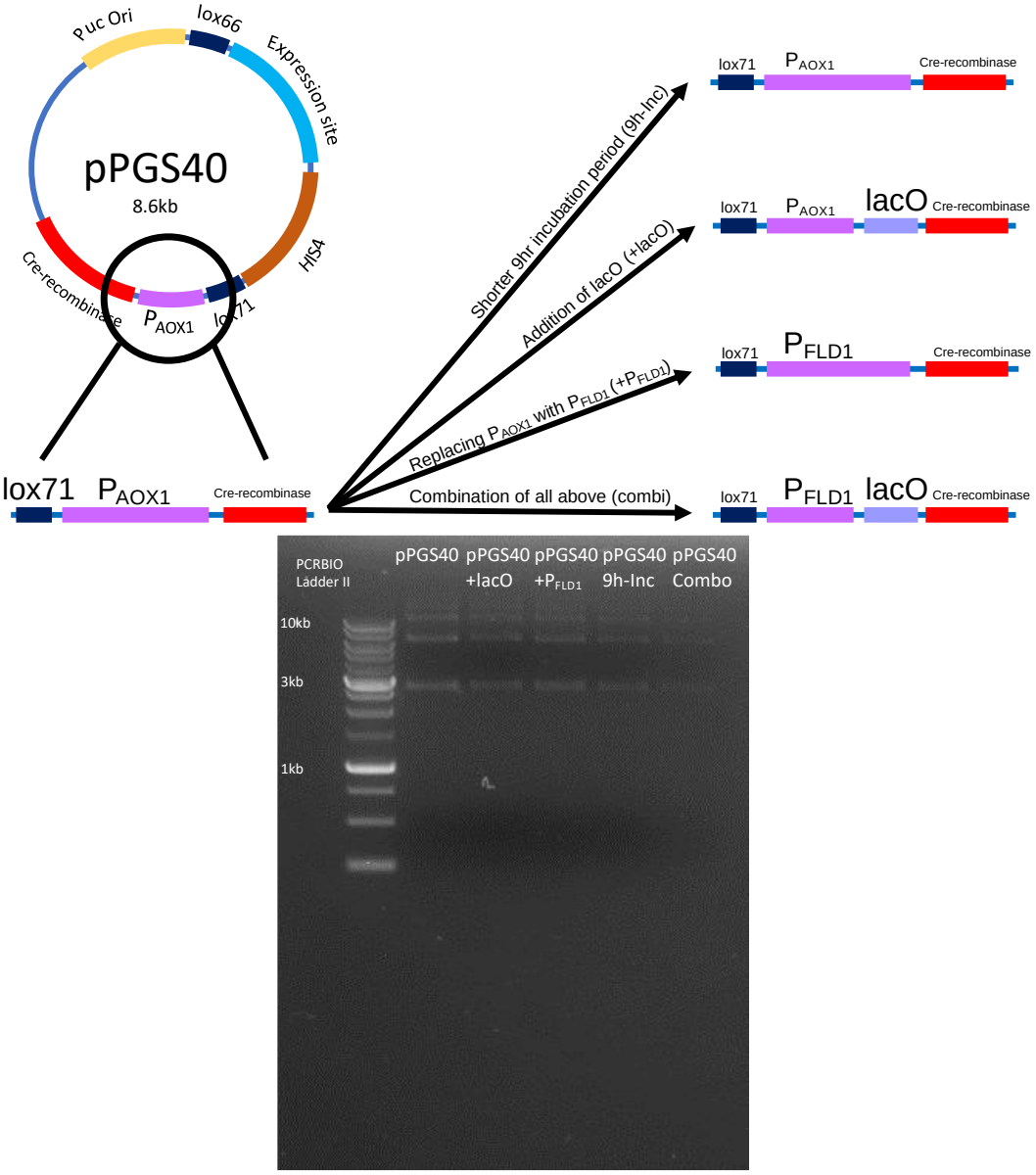

Figure 4.3: Addition of sequence-specific elements to halt recombination in pPGS40. (A) pPGS40 variants were modified to stop recombination occurring in E.coli. The AOX1 promoter was substituted for the FLD1 promoter (pPGS40+P_{FLD1}), the lac operon was added downstream of the AOX1 promoter (pPGS40+lacO), a shorter incubation time (pPGS40+9h-Inc) was tested, and a combination of all the above (pPGS40-Combi) was also trialled. (B) All the sequence modifications were miniprepped and visualised on an agarose electrophoresis gel, producing multiple bands.

To overcome the observed recombination, a bacterial (*E. coli*) transcriptional terminator was sought that would impose strong termination whilst remaining small in size. The intention was to use the terminator to limit non-induced leakiness expression from P_{AOX1}. A synthetic strong terminator of 61 bp (L3SP21) from Chen et al. (2013) was identified, generated as synthetic DNA and placed downstream of the *AOX1* promoter and upstream of the cre-recombinase gene in pPGS40. The new variant was transformed into *E. coli* and the plasmid was extracted from the overnight

cell culture. The plasmid was visualised by agarose gel electrophoresis (Figure 4.4A). The addition of synthetic terminator L3SP21 (T_{L3SP21}) downstream of the *AOX1* promoter (pPGS40 +T_{L3SP21}) resulted in a very distinct band at approximately 9 kb and one at over 10 kb, whereas the band at 3 kb that had been seen previously was no longer present. The plasmid (pPGS40+T_{L3SP21}) was sent for Sanger sequencing. The sequence analysis showed that no recombination of pPGS40+ T_{L3SP21} had occurred, therefore, confirming the absence of pPGS40r in the miniprepped culture.

Further modifications were sought to allow the plasmid to be used in a high-throughput method by utilising fluorescence to allow the detection of successful clones in *P. pastoris.* To minimise the length of the plasmid, the sequence of a 352 bp flavin-based fluorescent protein in iLOV (Christie et al., 2012) was added to the plasmid pPGS40 + T_{L3SP21}. It was added under the expression of the constitutive promoter *GAP* and the terminator *AOX1*tt forming the plasmid (pPGS40 + T_{L3SP21} + iLOV). This plasmid was transformed into *E. coli* and miniprepped. The plasmid was visualised by gel electrophoresis (Figure 4.4A), confirming the bands at 9 kb and over 10 kb, however, no bands at 3 kb were detected. In addition to these modifications (pPGS40 +T_{L3SP21} + iLOV), the *AOX1* promoter (931 bp) was replaced with the *FLD1* promoter (597 bp), resulting in the plasmid pPGS60 (Figure 4.4A, Figure 4.5). Upon linearization of pPGS60 using restriction enzyme AgeI (pPGS60 +RE-AgeI), a single band was observed at the correct size of approximately 10 kb (Figure 4.4B). Sanger sequence analysis and plasmid linearization confirmed that no recombination occurred in overnight cell cultures of pPGS60 (Figure 4.4).

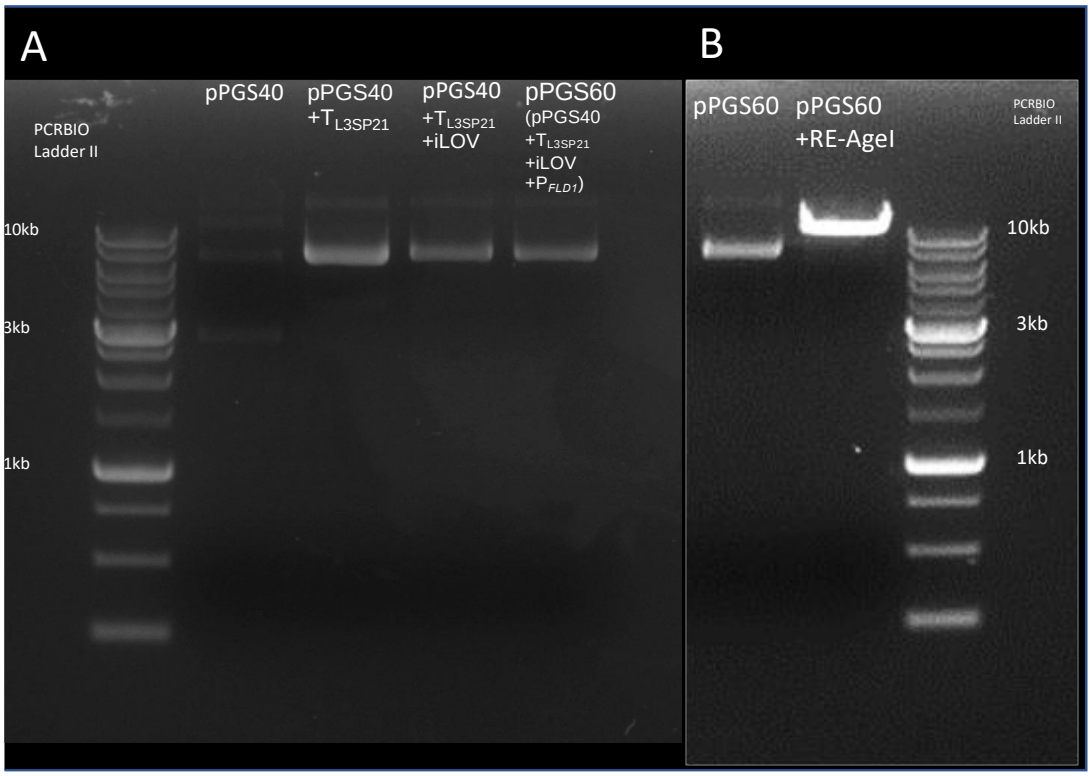

Figure 4.4: Addition of sequence-specific elements to improve pPGS60. (A) pPGS40 is provided as a reference showing recombination in E. coli cell cultures leading to multiple bands. pPGS40 was modified to stop recombination occurring in E. coli with the addition of synthetic terminator L3SP21. Further modifications to improve the plasmid included iLOV (pPGS40 +T_{L3SP21} +iLOV) and a final substitution of AOX1 to FLD1 (pPGS60 +T_{L3SP21} + iLOV +P_{FLD1}) to form the new plasmid pPGS60. (B) pPGS60 was linearized using restriction enzyme AgeI (pPGS60+Re-AgeI) and produced a single band at around 10 kb, close to the size of pPGS60 (9.6 kb).

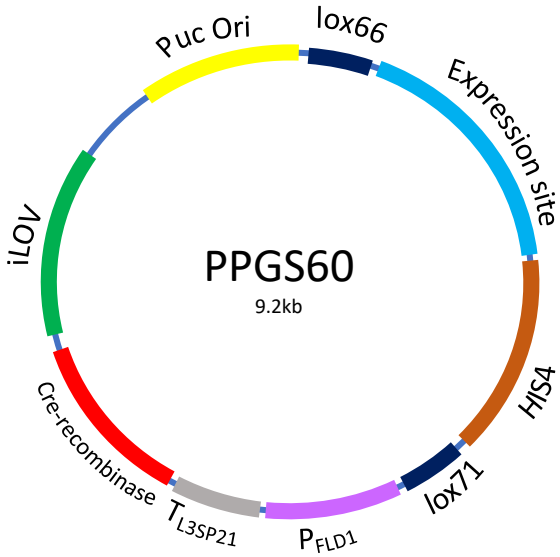

Figure 4.5: Finalised design of the new marker recycling plasmid, pPGS60. The plasmid is a modified variant of that from Chen et al. (2017). Promoter FLD1 (P_{FLD1}) was used to regulate the expression of cre-recombinase. A short synthetic terminator L3SP21 (T_{L3SP21}) from Chen et al. (2013) was added upstream of cre-recombinase to prevent expression in E. coli. iLOV was added (regulated by the constitutive promoter GAP and the terminator AOX1tt) for transformant selection in P. pastoris. The total size of the plasmid is 9.2 kb.

4.2 Addition of fluorescence improves the selection of successful *P. pastoris* transformants

The reporter gene iLOV was incorporated in order to determine the feasibility of transformant selection based on fluorescence, which can be advantageous in a combinatorial strategy, especially for high-throughput robotics. The flavin-based cyan fluorescent protein iLOV has a $\lambda_{ex}/\lambda_{em}$ maxima of 447/500 nm respectively. It was derived from phototropin, the light oxygen or voltage-sensing domain of the plant's blue light receptor by Chapman et al., (2008). It's ability to quickly recover spontaneously after photobleaching makes it ideal candidate for use. To induce fluorescence of iLOV, a light emitting platform was created that generated blue light of 380 to 500 nm, covering the excitation range of iLOV. To detect the fluorescence of iLOV, orange filter specs were used that allowed the passage of wavelength of light over 500 nm, covering the emission range of iLOV.

P. pastoris colonies transformed with a plasmid containing a gene expression cassette require colony PCR in order to confirm the integration of that plasmid. The identification of the *P. pastoris* cells transformed with the target plasmid (positive transformant) is essential for downstream applications (Lin-Cereghino et al., 2005). However, when implementing a combinatorial strategy resulting in the production of hundreds of *P. pastoris* colonies, colony PCR of each *P. pastoris* colony would be a time-consuming additional step that would slow the process for high-throughput screening. Therefore, the use of the iLOV fluorescence was utilised to be able to select a positive *P. pastoris* transformant. *P. pastoris* positive colony selection method was developed in this work that could increase positive transformant selection, reduce lab work input, decrease the amount/number of reagents needed and allow for robotics to select a transformant with a high rate of success (Figure 4.6). According to this method, pPGS60 is linearized at *HIS4* and transformed into competent *P. pastoris* cells (Figure 4.6, Step 1). The cells are then plated on to YEPDA+Zeocin plates and incubated at 30 ^{o}C. After 2-3 days of growth, the plates are exposed to blue light and using orange filter glasses, the *P. pastoris* colonies with genomically integrated plasmid will fluoresce (due to iLOV fluorescence activity) compared to the non-transformed colonies under the blue light/orange filter system. The fluorescing green

colonies are selected and set up as overnight cultures in methanol-containing BMMY media. As a result of the presence of methanol, the promoter P_{FLD1} is induced and subsequently the expression of cre-recombinase is induced (Figure 4.6, Step 2). Consequently, recombination between the *lox*p sites occurs in the genome and unwanted DNA is excised, which includes iLOV, resulting in a loss of fluorescence. The overnight culture is streaked on to YEPDA plates (without the addition of antibiotic selection) and colonies are grown for 2 days. The plates are exposed to blue light and colonies that do not fluoresce are selected using blue light and orange filter glasses (Figure 4.6, Step 3), The non-fluorescing colonies are resuspended in sterile water, and two parallel 5 µl spots from each colony are plated on YEPDA and YEPDA + Zeocin plates. The plates are incubated overnight at 30^{o}C. From the parallel spotting, colonies that grow on YEPDA plates but not YEPDA+Zeocin and do not fluoresce are considered as positive transformants, lacking the entire plasmid region outside of the *lox*p sites. This plasmid region includes the iLOV expression cassette and the antibiotic selection marker (Zeocin) and therefore the excision of these regions results in *P. pastoris'* loss of iLOV fluorescence as well as antibiotic resistance. The positive transformants were screened and confirmed via colony PCR.

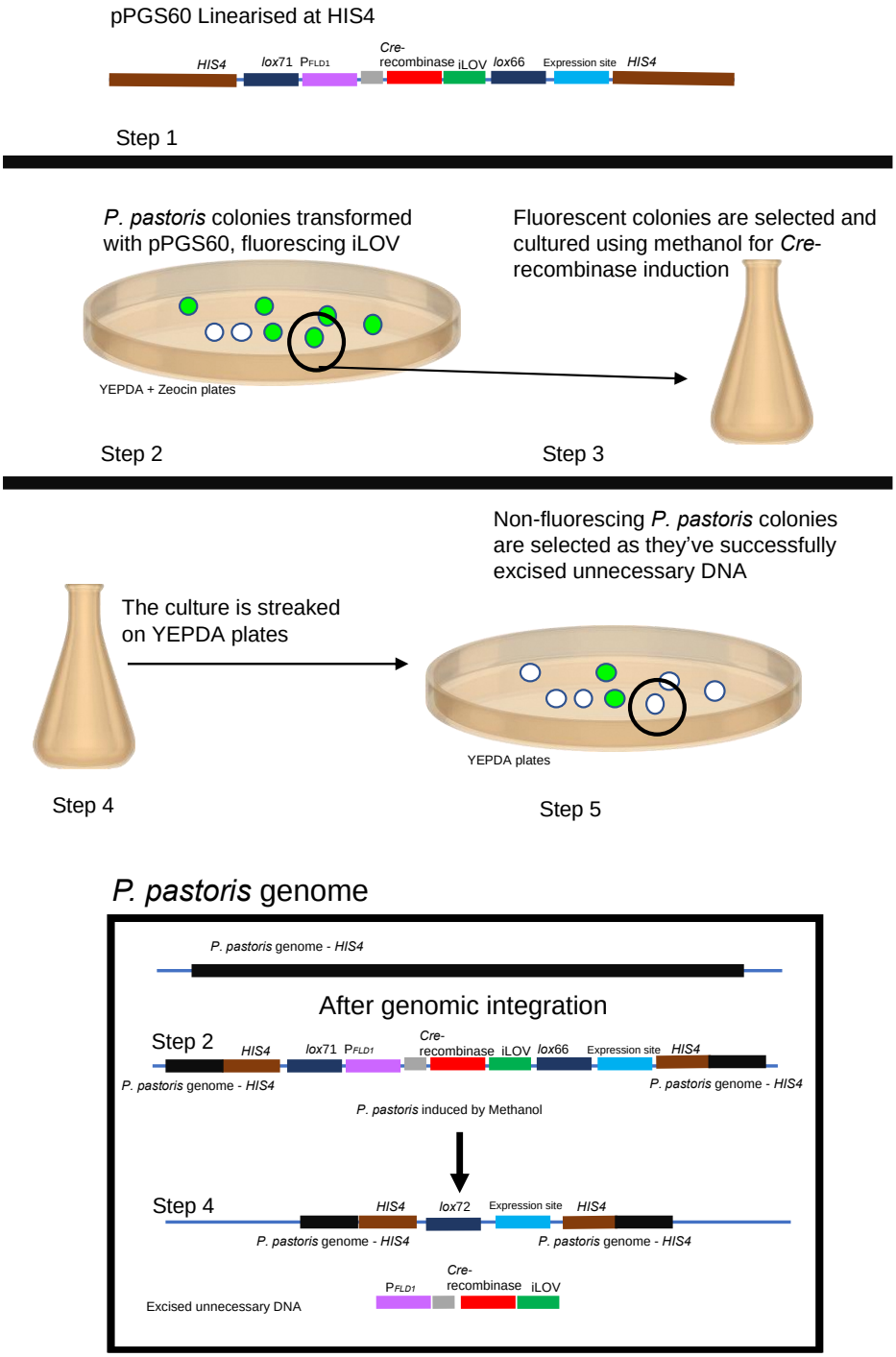

Figure 4.6: Successful transformant selection via iLOV fluorescence. A modified method that allows rapid screening of successful transformants via iLOV. Step 1, competent P. pastoris cells are transformed with PGPS60 linearised by cleavage at HIS4 and incubated at 30°C for 2-3 days. Step 2, the plates are exposed to blue light and by examining the plates using orange filter glasses, iLOV fluorescing colonies can be selected and set as overnight cell cultures at 30°C in BMMY media (Step 3). The expression of cre-recombinase via the methanol-regulated P_{FLD1} *promoter is induced, triggering recombination, excising the unnecessary DNA including selection marker (zeocin) and iLOV. Step 4, the cultures are streaked on to YPD plates and grown for 2-3 days. Step 5, plates are exposed to blue light and using orange filter glasses, colonies that do not fluoresce are picked and spotted as 5 ul spots on YPD and YEPDA+Zeocin plates, respectively. The plates are incubated overnight at 30°C. Colonies that only grew on YPD plates are selected and screened via colony PCR to confirm the correct recombination and integration of the expected expression cassette.*

This method (Figure 4.6) was employed to transform *P. pastoris* with pPGS60 (*P.p.* – pPGS60), checking integration via fluorescence (Figure 4.7). The colonies that were expressing iLOV were identifiable when exposed to blue light and viewed through orange filter glasses (Figure 4.7, compare panels A and B). Some colonies were spotted that did not fluoresce despite blue light exposure (Figure 4.7B) indicating false-positive colonies, being either satellite colonies or possibly as a result of secondary recombination events (Schwarzhans et al., 2016). Upon streaking of cell cultures grown with 0.5 % methanol, a mixture of colonies with or without fluorescence (Figure 4.7C) were detected. A clear visual difference in fluorescence was observed between the two colonies as a result of the excision of iLOV via cre-recombinase. Spotting was carried out, showing the loss of non-essential DNA from pPGS60.

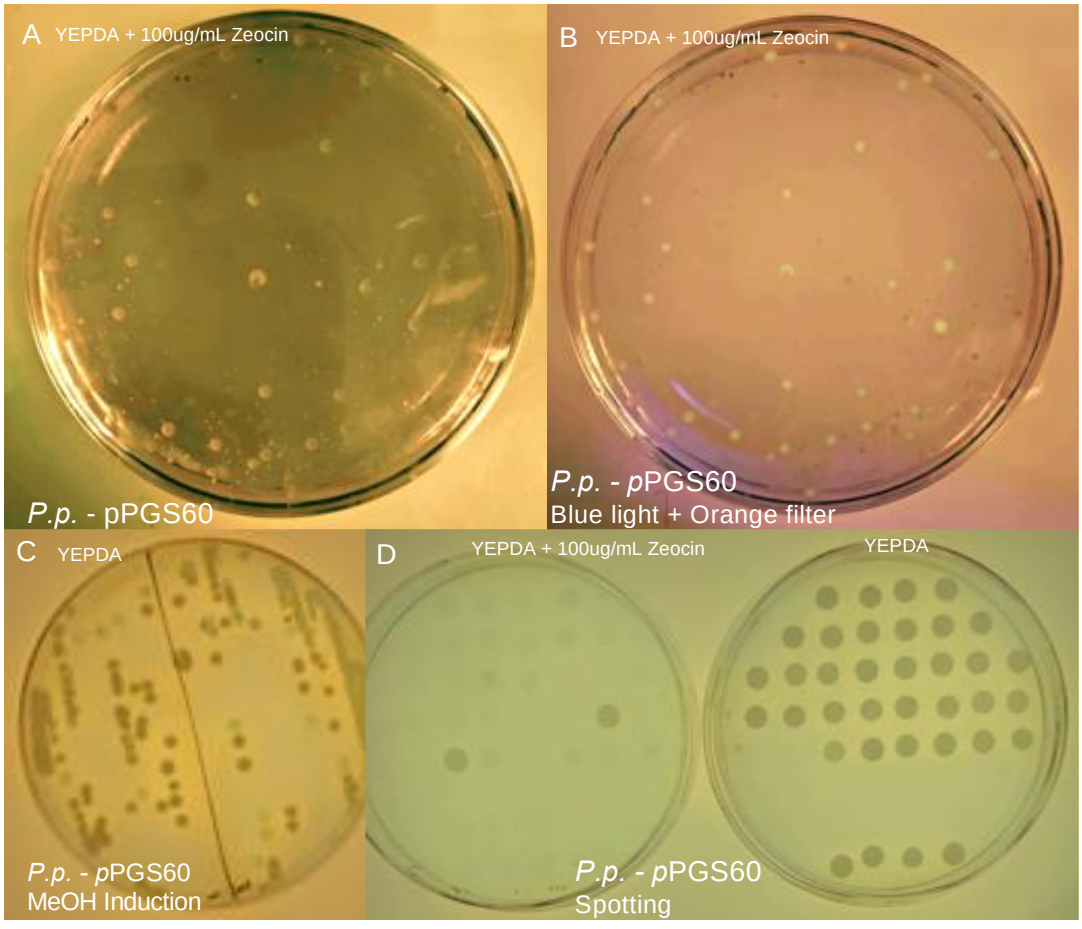

Figure 4.7: iLOV fluorescence method for successful P. pastoris pPGS60 transformant selection. P. pastoris was transformed with pPGS60 (P.p. – pPGS60) and incubated. After 3 days plates full colonies were seen (A) and using orange filter glasses and exposed to blue light (B) where their fluorescence was observed. Fluorescent colonies were selected, set up as overnight cell cultures in BMMY media and plated on YPD plates only. After 2 days of incubation, the plates were exposed to blue light and a lack of fluorescence was checked (C). Using orange filter glasses, colonies that did not fluoresce were selected and spotted as 5 ul spots on YPD and YPD + Zeocin plates (D).

P.p. – pPGS60 colonies were grown in two distinct media as a simple indicator to compare fluorescence levels observed with or without the induction of *cre*-recombinase (Figure 4.8). The colonies were grown in either BMMY (*P.p.* – pPGS60 + Methanol) to observe the effect of recombination and the subsequent loss of iLOV containing DNA or in YEPD (*P.p.* – pPGS60 + Glucose) to determine the effect of a lack of recombination as glucose does not induce the promoter *FLD1* and therefore cre-recombinase is not expressed. The fluorescence was measured as relative fluorescence units (RFU) with $\lambda_{ex}/\lambda_{em}$ maxima of 488 and 525 nm, respectively. *P.p.* – pPGS60 + Glucose produced high fluorescence intensity from the ~20-hour time point. In comparison, *P.p.* – pPGS60 + Methanol produced very little fluorescence at all time points. After 40 hours of culture cultivation, 0.5% methanol was added to *P.p.* – pPGS60 + Glucose culture. After this addition, a decrease in fluorescence was observed.

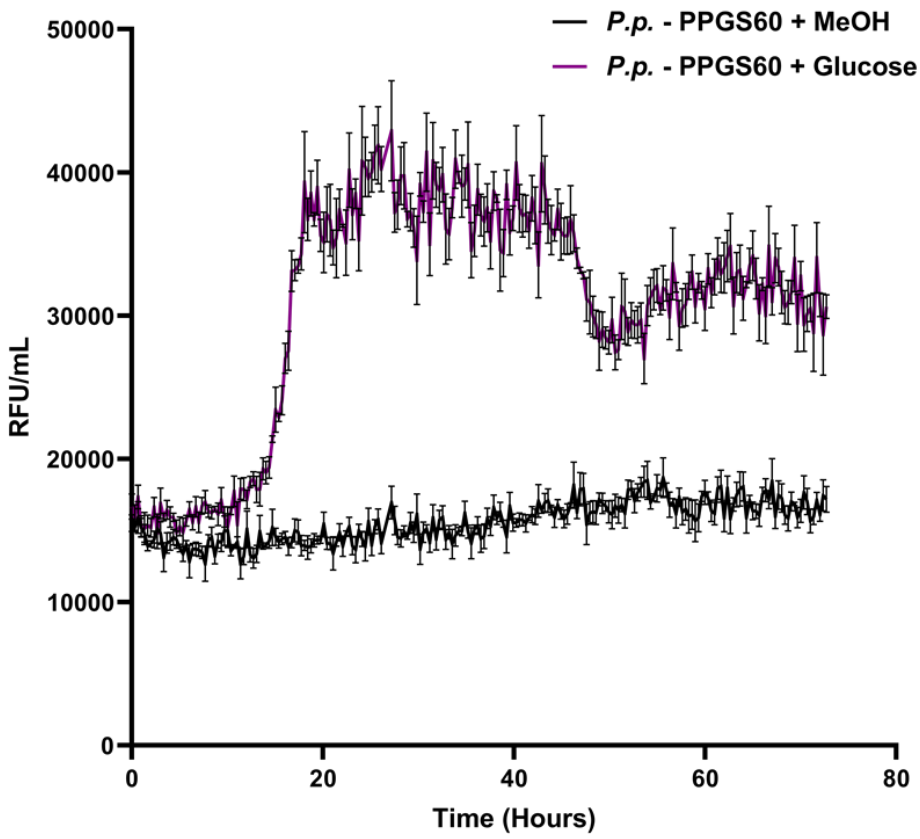

*Figure 4.8: Comparison of pPGS60 transformed P. pastoris in two distinct media. P. pastoris was transformed with pPGS60 (*P.p. *– pPGS60) and grown for 72 hours in either YEP + Glucose media (*P.p. *– pPGS60 + Glucose) or YEP + Methanol media (*P.p. *– pPGS60 + Methanol). The fluorescence was measured as relative fluorescence units (RFU) per mL.*

4.3 Comparison of pPGS40 and pPGS60 for successful transformant selection in *P. pastoris*

The addition of a synthetic terminator and of iLOV into pPGS60 was shown to introduce useful features in the development of a new fluorescence-based transformant selection method (Figure 4.7). To investigate the efficiency of using the pPGS60 plasmid and its associated fluorescence-based selection method for successful *P. pastoris* transformant selection (Figure 4.6), pPGS60 was compared to pPGS40 from Chen et al. (2017). Competent *P. pastoris* cells were transformed with pPGS40 (*P.p.* – pPGS40) or pPGS60 (*P.p.* – pPGS60). Colonies from transformed plates were selected either randomly (*P.p.* – pPGS40) or via the fluorescence method (*P.p.* –pPGS60) and successful/unsuccessful plasmid genomic integration was confirmed via colony PCR and reported as a percentage of rate of successful transformant selection (Table 4.1).

Table 4.1: Comparison of successful P. pastoris *transformant between pPGS40 and pPGS60 plasmid. The percentage is mean from 50 colonies, n=3. * denotes the relative standard deviation.*

		Positive transformant selection success rate (%)	
Step	**Additional Info**	***P.p.* – pPGS40**	***P.p.* – pPGS60**
1. After transformation	*P. pastoris* is transformed with the plasmid and incubated on YEPDA+Zeocin plates for 3 days. Colonies are selected based on fluorescence pPGS60) or randomly (pPGS40)	82% (*4.9)	96% (*1.2)
2. After MeOH induction	Colonies that are confirmed to have the plasmid integrated via iLOV fluorescence (pPGS60) or colony PCR (pPGS40) are grown overnight and then streaked on YPD plates.	46% (*14.8)	92% (*5.4)
3. Spotting	Transformed colonies are parallel-spotted on YEPD and YEPD + Zeocin plates. Those growing on YEPD-only plates are compared.	86% (*3.6)	84% (*3.8)

P.p. – pPGS60 reported a significantly higher success rate of correct transformant selection compared to *P.p.* – pPGS40 (96 % vs 84 % respectively, p-value < 0.05) at step 1 (Table 4.1). The positive *P. pastoris* transformants were then set up as overnight cultures in BMMY media and streaked on to YEPDA plates the following morning. *P.p.* – pPGS60 reported a significantly higher (p value < 0.005) successful transformant selection rate (92 %) compared to *P.p.* – pPGs40 (46 %) at step 2 (Table 4.1). The positive transformants were then spotted on to YEPD and YEPD + Zeocin plates where *P.p.* – pPGS60 had 84 % positive transformant selection rate and *P.p.* – pPGS40 had an 86% positive transformant selection rate, however, no significant difference was observed (p value > 0.05) between the two methods (Step 3, Table 4.1). Although no significant difference was observed, the major advantage that pPGS60 offers is that it does not require colony PCR to confirm the correct transformant until the final stage due to a high probability of the transformant selected being positive as a result of iLOV fluorescence.

To investigate how the addition of fluorescence affects the transformation method to recycle the selection marker, the steps in the two methods are compared in detail in Table 4.2. The key differences include the exclusion of colony PCR confirmations due to the high accuracy of selecting the correct transformant using iLOV fluorescence (Table 4.2, day 3 and day 6). Large *P. pastoris* colonies commonly required for colony PCR (for genomic DNA extraction for cPCR) were unnecessary as colonies could be selected at earlier stages based on the fluorescence. In this instance, positive colonies could be selected after 2 days of incubation instead of 3 days (Table 4.2, day 3 and day 6). A protocol from Kumar (2019) was incorporated here to transform *P. pastoris* cells from the 5 µl spots to simplify and reduce the competence preparation. Commonly, P. *pastoris* cell cultures are grown overnight, diluted the following morning and grown for another 6 hours before being made competent (Lin-Cereghino et al., 2005). This new approach reduced the competence preparation time by a further day (Table 4.2, day 7). Combining the changes resulted in reducing the length of time from 10 days (*P.p.* – pPGS40) to 7 days (*P.p.* – pPGS60). The fluorescence-based selection also allowed automation to be utilised whereas previously the overall procedure relied on extensive manual effort. These changes meant that the transformation of *P. pastoris* using pPGS60 requires fewer colonies selected to

identify positive transformants, fewer colony PCRs (thus reducing workload), and an automation-friendly procedure.

Table 4.2: Comparison of methods between pPGS40 and pPGS60

Day	***P.p.* – pPGS40 Transformation method**	**Day**	***P.p.* – pPGS60 Transformation Method**
1	Transform into *P. pastoris* (Lin-Cereghino et al., 2005)	**1**	Transform into *P. pastoris* (Lin-Cereghino et al., 2005)
4	Select random colonies for transformation confirmation via cPCR	**3**	Select fluorescent colonies
4	Overnight cultures of successful transformants in BMMY	**3**	Overnight Cultures of successful transformants in BMMY
5	Streak colonies from overnight cultures on YEPDA plates	**4**	Streak colonies from overnight cultures on YEPDA plates
8	cPCR of random colonies to select successful transformants (excision of unnecessary DNA) and parallel spot on YEPDA and YEPD + Zeocin plates	**6**	Select colonies that do not fluoresce (excision of unnecessary DNA) and parallel spot on YEPDA and YEPDA + Zeocin plates
9	Select colonies that only grow on YEPDA plates	**7**	cPCR confirmed colonies for assay
9	Grow colonies overnight for the second transformation	**7**	Scrape colonies (Kumar, 2019) and make them competent for second round of transformation via marker recycling
10	Grow for 6 hours and make cells competent and begin the second round of transformation via marker recycling		

4.4 Multi-gene integration analysis of pPGS60

The recycling of the selection marker was expected to allow multi-gene integration into *P. pastoris*. To assess the feasibility of incorporating multiple reporter genes using the iLOV fluorescence-based method, three reporter genes were inserted into pPGS60. The reporters were iLOV, RFP (mCherry) and mCitrine (a yellow fluorescent protein; Higuchi-Sanabria et al., 2016). The fluorescence-based protocol (Figure 4.6) was followed and the reporters were sequentially integrated into *P. pastoris* genome by recycling the zeocin selection marker in the timeframe of 21 days (9 days faster than the non-fluorescence-based pPGS40 method). All three reporter genes were cPCR confirmed to have been co-integrated into *P. pastoris* (as a result of possible bleed-over), creating the strain *P.pastoris* - pPGS60RYi. The strain was imaged using a EmCCD/sCMOS widefield microscope at excitation/emission wavelengths of $\lambda_{ex}/\lambda_{em}$ 560/630 nm for RFP, $\lambda_{ex}/\lambda_{em}$ 500/530 nm for mCitrine, and $\lambda_{ex}/\lambda_{em}$ 470/525 nm for iLOV (Figure 4.9). *P.p.* – pPGS60-RYi fluorescence was detected for all 3 reporters (Figure 4.9B-D). This confirmed that triple gene integration and expression can be achieved using pPGS60 in a comparatively fast timeframe. Although it was already 9 days quicker and less resource-intensive than the pPGS40 method, further improvements were sought to improve the time taken for multi-gene integration.

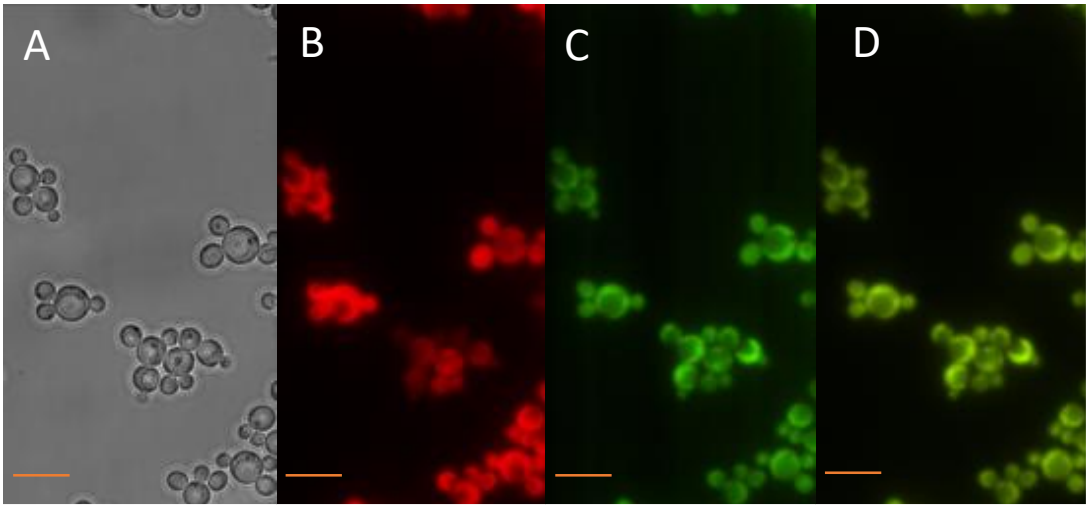

Figure 4.9: Multi-gene integration and expression in P. pastoris using the pPGS60 plasmid. P. pastoris was transformed with RFP, mCitrine and iLOV sequentially using PPGS60 and a strain expressing all three was created (P.p. – pPGS60-RYG). (A-D) Microscopy cell imaging was used to view the expression of all three fluorescent proteins. The excitation/emission wavelengths for RFP, YFP and iLOV was 560/630 nm, 500/535 nm and 470 / 525 nm respectively. The images produced show the expression of (A) brightfield, (B) RFP, (C) iLOV and (D) YFP. Scale bars represent 10 µm at 100x magnification. The strains were colony PCR confirmed to contain each of the three fluorescent reporter as a result of possible bleed-over of the iLOV or YFP.

4.5 Multi-gene integration into the *P. pastoris* genome based on pPGS60

Multi-gene integration could be improved by reducing the time taken for sequential gene integration. To achieve this, an alternative plasmid design (pPGS80) was created that was based on pPGS60 (Figure 4.10). pPGS60 contains the *HIS4* homologous recombination site for integration (Chromosome I) so a different homologous recombination integration site was chosen for pPGS80 (*URA3*, chromosome III). The different chromosomal location also served the purpose to reduce/prevent unwanted recombination between the two plasmids. Moreover, an alternative selection marker in geneticin (G418) was identified for pPGS80, replacing the zeocin selection marker in pPGS60 allowing unequivocal identification of a double transformant. pPGS80 is smaller in size (8.7 kb compared to pPGS60 9.2 kb) due to the smaller size of *URA3*. Initially, pPGS80 was subjected to a single plasmid transformation to determine whether the combination of *URA3* and G418 would be compatible with successful integration into the *P. pastoris* genome.

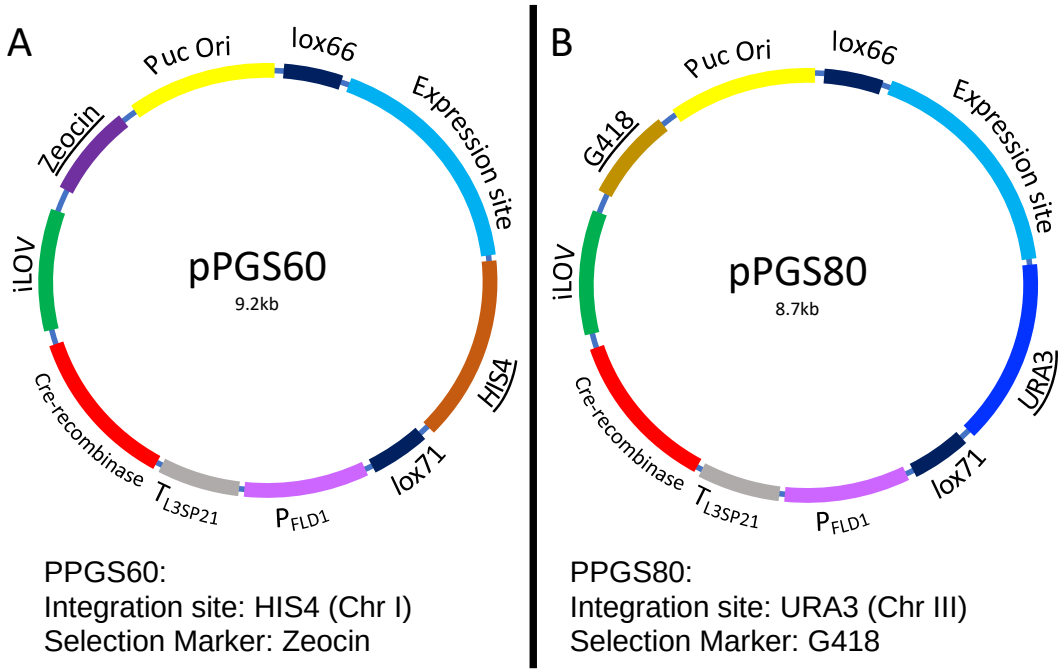

Figure 4.10: Comparison of pPGS80 and pPGS60 plasmid. (A) pPGS60 and (B) pPGS80 plasmid maps. Both plasmids are identical in design and sequence except for two differences (underlined): i) pPGS80 contains URA3 sequence (chromosome III) for homologous recombination integration, pPGS60 contains HIS4 sequence (chromosome I) for homologous recombination integration, and ii) pPGS80 contains geneticin (G418) as a selection marker, pPGS60 contains Zeocin as a selection marker. Both plasmids use cre-recombinase and loxp sites to excise non-essential DNA after genomic integration. Both contain iLOV for fluorescence-based positive transformant selection.

The efficiency of the transformant selection rate for plasmid pPGS80 (Table 4.3) was analysed as previously (Table 4.1). The combined use and co-transformation of the plasmids pPGS60 and pPGS80 was also assessed. *P. pastoris* was transformed with pPGS80 (*P.p.* – pPGS80) or co-transformed with pPGS60 and pPGS80 (*P.p.* – pPGS60+pPGS80). Fluorescence was used to select the suspected positive *P. pastoris* transformants. After the initial transformation (Table 4.3, Step 1), both *P.p.* – pPGS80 (90%) and *P.p.* – pPGS60+pPGS80 (86%) reported a high positive transformant selection rate and were similar to that of pPGS60 (92%) (no significant difference, p value > 0.05). After induction with methanol and re-streaking (Table 4.3, step 2), the percentage success rate for *P.p.* –pPGS80 (84%) and for *P.p.* – pPGS60+pPGS80 (78%) were similar (no significant difference, p value > 0.05). No significant difference (p value > 0.05) was observed in the positive transformant selection rate after the final stage of spotting for *P.p.* – pPGS80 (80%) and *P.p.* – pPGS60+pPGS80 (76%). This indicated that co-transformation of *P. pastoris* using two plasmids can be performed successfully, offering a faster process for double gene integration.

*Table 4.3: Comparison of successful P. pastoris single and double transformants via pPGS80 and pPGS60. The percentage is mean from 50 colonies, n=3. * denotes the relative standard deviation.*

		Positive transformant selection rate (%)		
Step	**Additional Info**	***P.p.* – pPGS60**	***P.p.* – pPGS80**	***P.p.* – pPGS60+pPGS80**
1. After transformation	*P. pastoris* is transformed with the plasmid and incubated on YPD plates for 3 days. Colonies are selected based on fluorescence	96% (*1.2)	90% (*9.6)	86% (*6.2)
2. After MeOH induction	Colonies that are confirmed to have the plasmid integrated via iLOV fluorescence are grown overnight and then streaked on YEPD plates.	92% (*5.4)	84% (*12.6)	78% (*7.7)
3. Spotting	Transformed colonies are spotted on YEPD only, YEPD + Zeocin and YEPD + G418 plates. Those growing on YEPD only plates are compared.	84% (*3.8)	80% (*11.5)	76% (*12.1)

4.6 Improving gene cluster integration using Bi-directional promoters

Significant improvements had been implemented that allowed double gene integration in 7 days (pPGS60+pPGS80). However, additional modifications were now introduced that were intended to allow two gene cassettes to be built into a single plasmid, thus paving the way to enabling four genes to be integrated simultaneously per co-transformation within 7 days. A new combinatorial strategy was assessed that would allow the expression sites of pPGS60 and of pPGS80 to be transcribed from bi-directional promoters instead of mono-directional promoters (Figure 4.11).

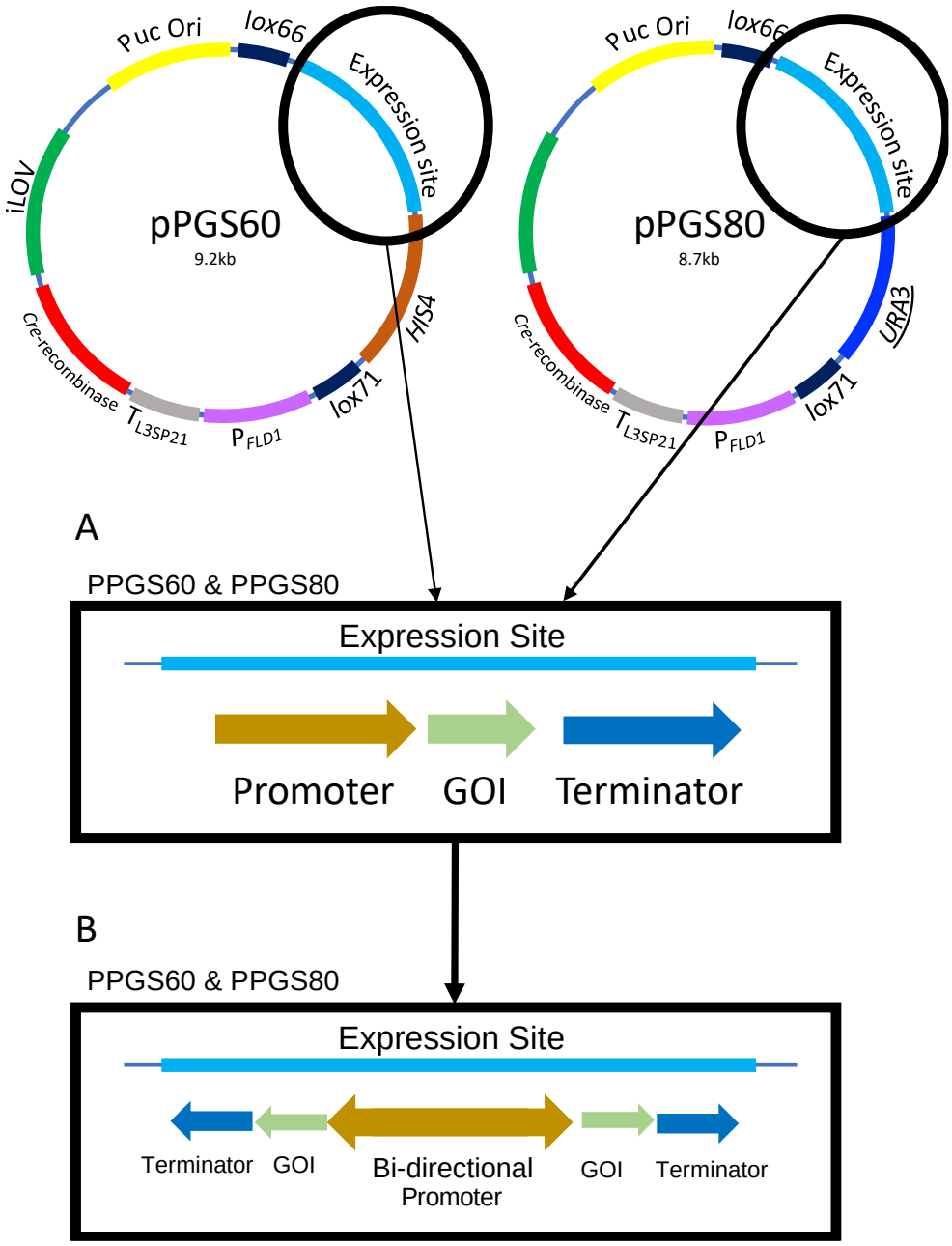

Figure 4.11: Design of pPGS60 and pPGS80 expression site. (A) pPGS60 and pPGS80 both contain expression site that utilises a mono-directional promoter. (B) By utilising a bi-directional promoter in that expression site, two genes per transformation event and so four genes per co-transformation event could occur. A bi-directional promoter can drive the expression of the gene of interest (GOI) in both orientations.

To establish the bi-directional system, two strategies were developed. The first was to use two copies of the *GAP* promoter, one in the forward direction and one in the reverse direction, combined with oppositely oriented genes of interest (GOIs) placed downstream of the respective promoters. The second strategy was to assemble two forward facing cassettes (Figure 4.12).

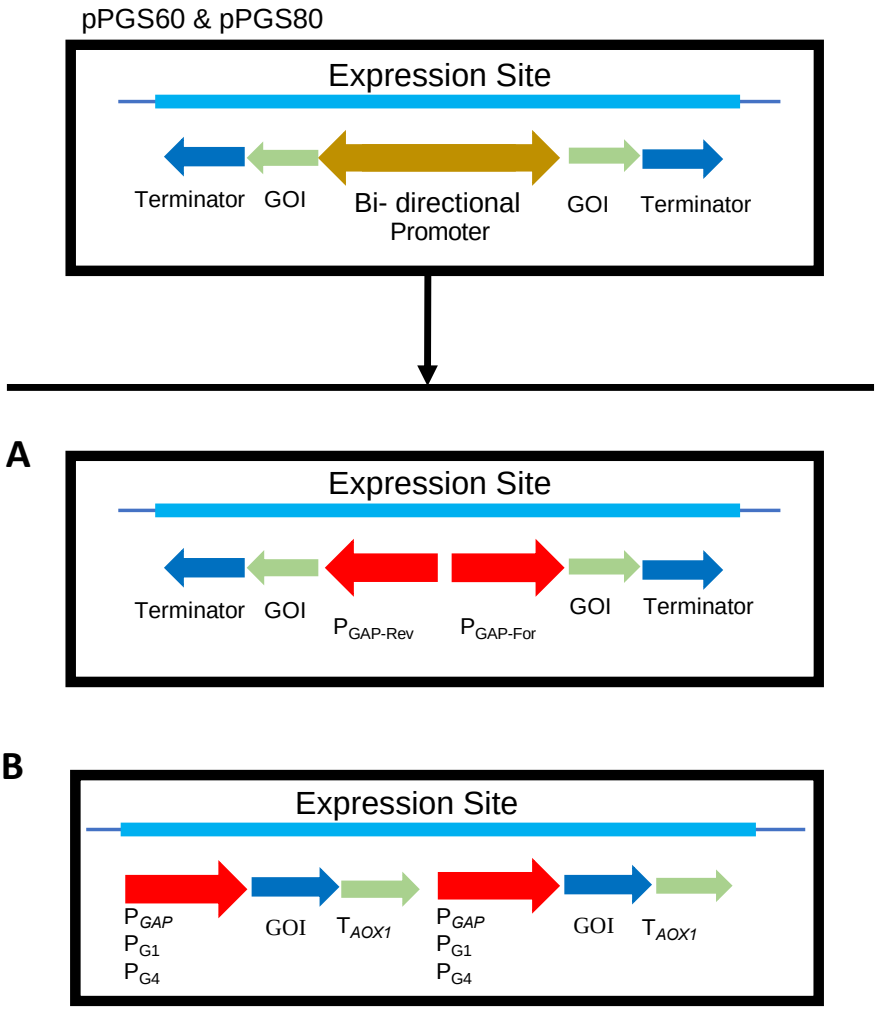

Figure 4.12: Bi-directional promoter strategies. Two strategies will be compared in terms of tuneability and expression. A) GAP promoter variants could be utilised to drive over-expression of recombinant genes whilst varying ratios of expression by using different combinations of promoters. B) Use of a synthetic bi-directional promoter that promotes equally strong transcription in both directions but combining this with a range of 5'UTR sequences that direct different rates of translation.

Two P_{GAP} promoter variants were previously identified (G1, G4; Qin et al., 2011) that manifest higher activities than the original P_{GAP}. The three promoters collectively provided three different rates of transcription initiation. These options were exploited as the basis for varying the ratio of expression of genes in the cellulose degradation pathway (Table 4.4). Combining them in their forward and reverse orientations, allowed the engineering of 6 distinct combinations if we assume that the positioning of the two promoters has no effect on their respective activities (i.e. 3! possibilities; Figure 4.12), or of 9 distinct combinations if the relative positions matter. Moreover, the fusion of forward and reverse facing promoter sequences created a set of bi-directional promoters (Figure 4.12A). The promoters (P_{GAP}, P_{G1}, P_{G4}) were individually amplified together with an overlapping middle sequence that allowed

assembly of each pair of promoters via sequence homology. However, when PCR amplification was carried out, it resulted in no specific bands at the desired size and instead smearing was detected. Various attempts were made to optimise the PCR method (data not shown), but no sequences were amplified.

To overcome this problem, the fragments were amplified as individual fragments (promoter forward and promoter reverse) and added together with the restriction-digested pPGS60 backbone to the Gibson assembly mixture. The Gibson mixture was then transformed into *E. coli* and incubated overnight. However, no colonies were detected in *E. coli* grown on selective media as a suspected result of the lack of generation of the plasmids. Therefore, the approach to use two forward-oriented cassettes as highlighted in (Figure 4.13A) was explored instead. Gibson assembly utilises homologous overlapping regions for assembly between DNA fragments and the plasmid backbone. These overlapping homologous regions would affect the rate at which a fragment is assembled to the plasmid backbone (Casini et al., 2014). Therefore, the 5' and 3'overlapping homologous regions of the promoters (P_{GAP}, P_{G1} and P_{G4}) to the plasmid backbone and gene of interest contained identical overlapping regions. These regions were set to have identical base pairs (25-30 bp) and 60 ^{o}C annealing temperature, in order to avoid any bias towards a specific promoter fragment. This resulted in the random assembly of any of the expected promoter pairings.

Table 4.4: GAP promoter variants for tuneable expression

Promoter	Transcript level of MAT, yEGFP and GAL compared to P_{GAP}	Activity of MAT, yEGFP and GAL compared to P_{GAP}	Description	Source
GAP	1(+/-0)	1(+/-0)	Glyceraldehyde 3-phosphate, a constitutive promoter, present in *P. pastoris* genome	*Pichia pastoris* Genome
G4 (*GAP* Variant)	8.29(+/-1.5)	4.02(+/-4.02)	Glyceraldehyde 3-phosphate variant containing mutations to improve performance	(Qin et al., 2011)
G1 (*GAP* Variant)	61.98 (+/-13.29)	40.62(+/-48.96)	Glyceraldehyde 3-phosphate variant containing mutations to improve performance	(Qin et al., 2011)

The utilisation of two gene cassettes in a single plasmid could lead to the development of a high-throughput method for optimisation of metabolic pathways. A new method was developed that would utilise the two gene-cassettes per plasmid to allow up to two or four gene cassettes to be integrated with randomly assembled expression systems in one transformation event (Figure 4.13). pPGS60 and/or pPGS80 restriction-digested backbones alongside the fragments of gene of interest (GOI) and the *GAP* promoter variants would undergo random Gibson assembly (Figure 4.13A). This random assembly can result in any one of the possible combinations of promoter-GOI cassette, resulting in a range of ratios of expression of the respective genes. The Gibson assembly mixture is then transformed into *E. coli* (Figure 4.13B) and plated overnight. The following morning, all colonies are collected, mixed, and miniprepped to extract the random plasmid combinations (Figure 4.13C). The plasmid is linearized and transformed into *P. pastoris* (Figure 4.13D). Colonies are selected and grown in 24-well plates and assayed using a robotic platform (Tecan freedom Evo in this work) for high-throughput analysis (Figure 4.13E). The highest-yielding colony is selected

and undergoes Sanger sequencing to determine the combination and elements as well as chosen for growth at a larger scale (Figure 4.13F).

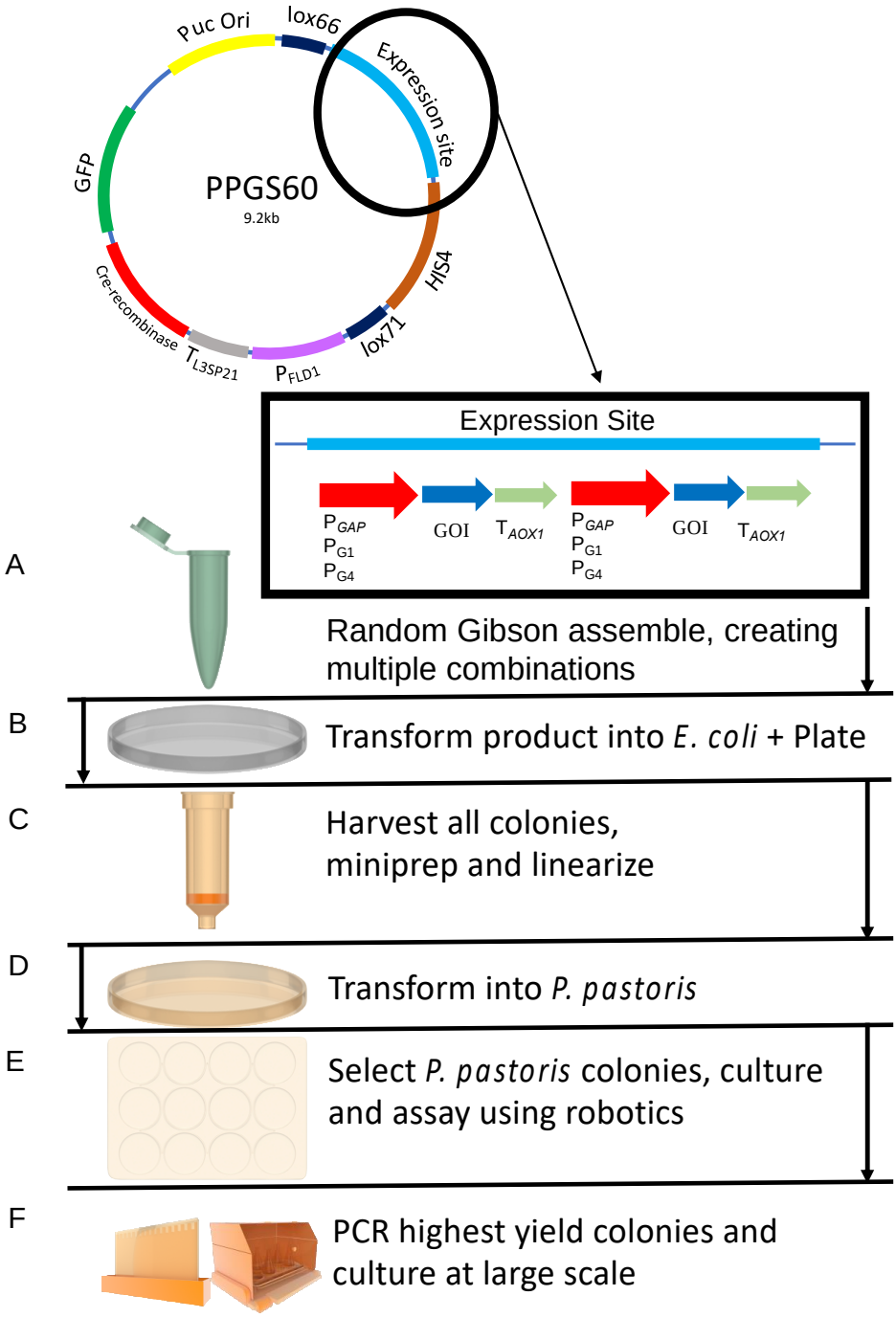

Figure 4.13: A high-throughput method for metabolic pathway optimisation in P. pastoris. (A) cassettes of P_{GAP} *promoter variants are randomly assembled. (B) The mixture is transformed into E. coli competent cells. (C) All colonies are harvested together, miniprepped to extract the randomly assembled plasmids and linearized for P. pastoris integration. (D) The linearized plasmid is transformed into P. pastoris. (E) Non-essential DNA is removed and final colonies are assayed using robotics. (F) The highest-yielding colonies are Sanger-sequenced to determine the optimal combinatorial setup.*

In summary, a cre-recombinase plasmid from Li et al. (2017) was trialled as the basis for the development of a combinatorial multi-gene expression strategy. However, this plasmid manifested properties that ruled it out for use in our intended strategy. Therefore, modifications were introduced into the plasmid to prevent unwanted recombination from occurring, to allow fluorescence-based screening for positive *P. pastoris* transformants, and to develop a new combinatorial multi-gene expression strategy that can be at least partially automated.

Chapter 5 – Combinatorial strategy to optimise metabolic pathways

The optimisation of metabolic pathways is used to maximise the yield of a desired product, which is highly beneficial to the biotech industry. In this chapter, using the combinatorial strategy devised in chapter 4 we optimised metabolic pathways of high value to the biotech industry in *P. pastoris*. As the ferulic acid degradation pathway and cellulose degradation pathway yield high-value products to the biotech industry, we aimed to optimise them using the combinatorial strategy. To achieve this, fluorescent proteins were used to observe the variation in the ratio of gene expression in *P. pastoris*.

5.1 Combinatorial strategy implementation using fluorescent proteins

To assess the capability of the combinatorial strategy outlined in Figure 4.13, the fluorescent proteins (FPs) mCherry (RFP), mTFP1 (TFP) and mCitrine (YFP) were used with the promoters P_{GAP}, P_{G1} and P_{G4} in random combinations using Gibson assembly. The random assembly of those fragments resulted in the creation of various plasmids (up to 27 combinations) in the Gibson mixture which were then transformed into *P. pastoris*. The transformed colonies were randomly selected and cultivated for 48 hours in multiple 24-well plates. Their fluorescence was measured as relative fluorescence units (RFU / mL). The average fluorescence of each of the three FPs for each colony was divided by the average fluorescence of the respective FP of the control strain *P.p.* –RYT (**R**FP, **Y**FP, **T**FP) to calculate the net fluorescence intensity gain or loss for each of the transformed colonies (Figure 5.1). The control strain (*P.p.* –RYT) consisted of the FPs RFP, TFP and YFP under the regulation of the promoter P_{GAP} integrated into *P. pastoris*. The colonies containing a randomly assembled plasmid generated differing amounts of fluorescence intensity, indicating variation in expression as a result of the different promoters being used. For example, colony S3 reported a ~230 % increase in YFP and in TFP fluorescence compared to control YFP and TFP fluorescence. A ~160 % increase in the RFP fluorescence of S3 compared to the control RFP fluorescence was observed making RFP the weakest fluorescing protein of the three in net fluorescence gain. On the other hand, colony S7 displayed a ~200 % increase in RFP fluorescence compared to the control RFP (*P.p.* –RYT), but

only a ~160 % and ~150 % increase in fluorescence for YFP and TFP respectively, compared to the control YFP and TFP (*P.p.* –RYT). These variations observed in FP expression were expected due to the variable expression strength of the promoters, resulting to variability in fluorescence generated by the FPs. The higher fluorescence of most transformants compared to the control is expected as the promoters P_{G4} and P_{G1} are P_{GAP} mutants conferring higher strength. Furthermore, we were able to confirm that the combinatorial strategy outlined in Figure 4.13 is effective in creating variation in gene expression (via the variation in fluorescence of FPs) using Gibson assembly to randomly create plasmids using a single reaction.

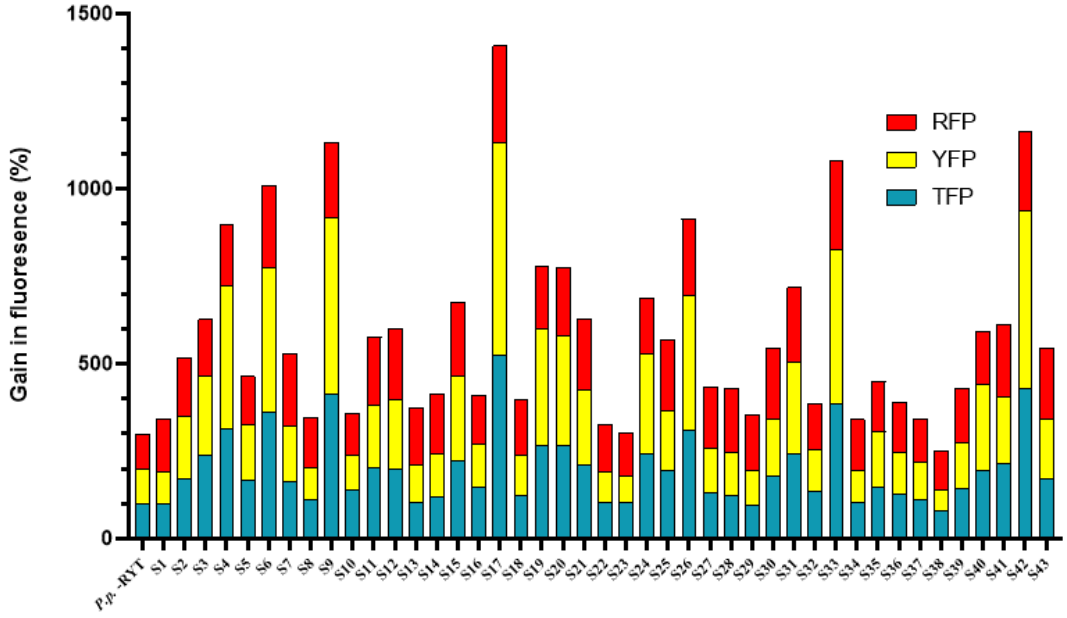

Figure 5.1: Combinatorial strategy using fluorescent proteins in P. pastoris. P. pastoris was transformed using randomly Gibson assembled plasmids containing a total of 3 promoters (GAP, G1 and G4) and 3 fluorescent proteins of mTFP1 (TFP, blue), mCitrine (YFP, yellow) and mCherry (RFP, red). The total sum of fluorescence after 48hours of cultivation in 1ml YPD-2% glucose at 30 ºC, 220rpm divided by the control P. pastoris colony containing RFP, YFP and TFP all regulated by the promoter GAP. The percentage fluorescence is given as an increase in fluorescence intensity compared to P.p. –RYT. The fluorescence was obtained via Tecan plate reader and fluorescence readings have not been normalised to OD_{600}.

5.2 Combinatorial strategy to optimise the cellulose degradation pathway

The combinatorial strategy (Figure 4.13) was applied to the cellulose degradation pathway (Table 3.1) in order to optimise the metabolic pathway for degrading cellulose to glucose (Figure 5.2). Three cellulases (endoglucanase [TrEGI], exo-glucanase [ExG2] and β-glucosidase [PaBG1b]) were assembled in combination with the three *GAP* promoter variants (*GAP*, *G1* and *G4*) via Gibson assembly and transformed into *P. pastoris*. The optimal combination of the promoters and the cellulase proteins would be indicated by the optimal cellulose degradation. Glucose is released as a result of cellulose degradation and is metabolised by *P. pastoris* as the only carbon source, thus enabling growth. Therefore, the rate of cellulose degradation is reflected in the rate of growth of the colonies (as measured using OD_{595}). The transformed cellulase-containing *P. pastoris* colonies were inoculated into 1 mL BMGY in a 24 well plate and grown for 24 hours at 30 ^{o}C. These colonies were then added to a fresh 24 well plate containing BGCY (cellulose-containing media) at a starting OD_{595} of 0.6. The colonies were left to grow for 24 hours and their OD_{595} was measured. The growth rate was calculated using OD_{595} and the results are highlighted in Figure 5.3. The empty vector control was grown on 2 % glucose (*P.p.* -Ev +Glucose) or cellulose (*P.p.* -Ev) as the positive and negative controls respectively. *P.p.* -Ev +Glucose observed the highest growth rate as expected, in comparison to *P.p.* -Ev which observed almost no growth as a result of the lack of a carbon source present. The strains expressing the cellulases in different ratios (S1-S40) produced a range of growth rates (Figure 5.3). Sanger sequencing was used to identify the promoters present in the genome of the strains exhibiting the highest growth rate (S38, S39, S40). It was found that both the endoglucanase and the exoglucanase were regulated by the promoter P_{G1} whereas the β-glucosidase was regulated by the promoter P_{GAP} (S39 and S40) or P_{G1} (S38). The strains S7, S8 and S9 contained the promoter P_{GAP} for all three cellulases (Table S5.2). Overall, the data show that the observed variation in the growth rate of these respective strains is linked to the promoter types present upstream of the cellulase genes.

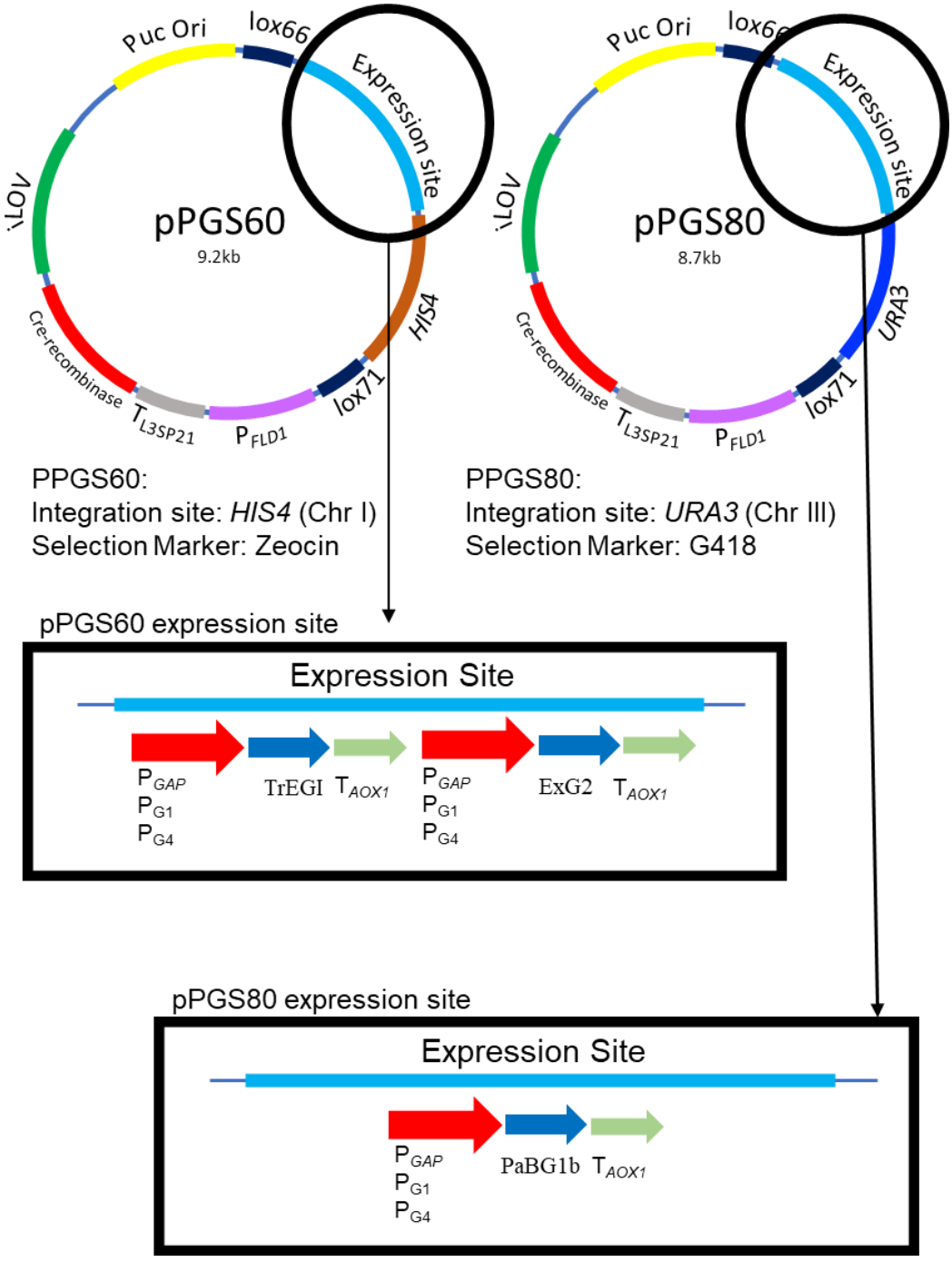

Figure 5.2: Plasmid map of pPGS60 and pPGS80 containing cellulase sequences. The plasmid pPGS60 expression site contains the Endoglucanase (EnG, TrEGI) sequence and exoglucanase (ExG2, I0H8B9) regulated by either one of the promoters pGAP, PG1 or PG4. The plasmid pPGS80 expression site contains the β-glucosidase (BGL, PaBG1b) sequence regulated by either one of the promoters P_{GAP}, P_{G1} or P_{G4}. All three cellulase sequences (TrEGI, ExG2, PaBG1b) contain the secretion signal Pre-Ost1-Pro fused upstream of their start codon.

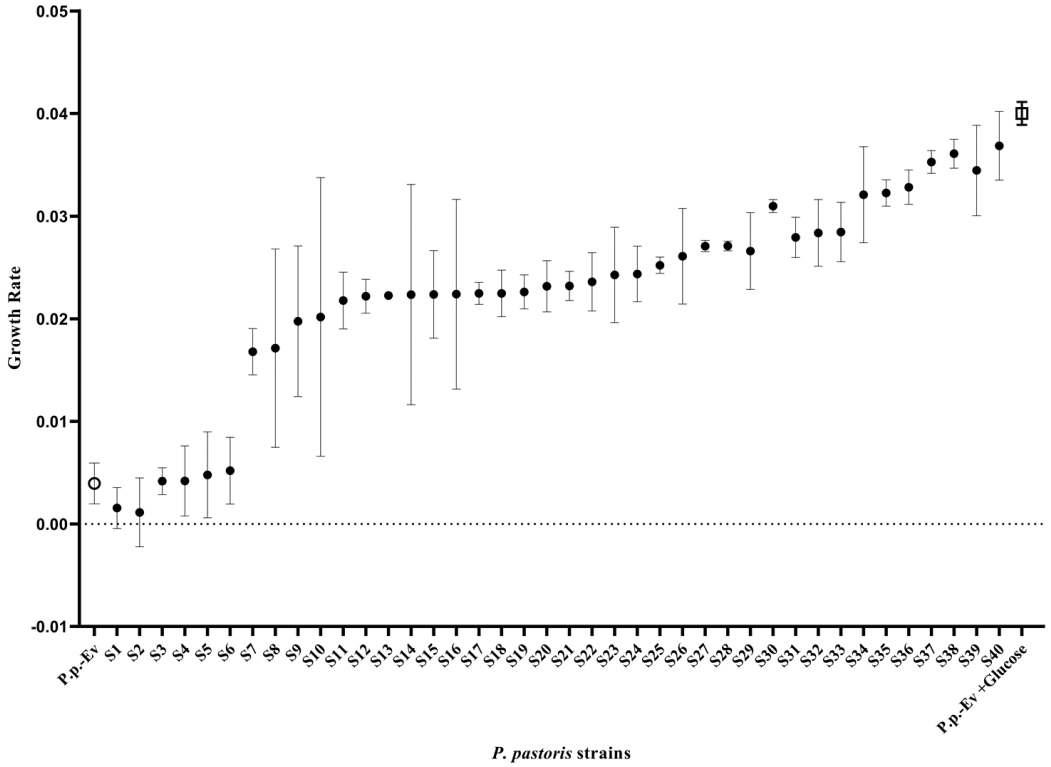

Figure 5.3: Growth rate (K) of P. pastoris containing cellulases regulated by different promoters. P. pastoris was transformed with randomly assembled plasmid containing three cellulases (endoglucanase, exoglucanase and β-glucosidase) regulated by any one of three promoters (GAP, G1 and G4). The strains were grown from an OD_{595} of 0.6 in cellulose (6mm disc shaped filter paper). The growth rate (k) was measured using non-linear regression. The error bars represent two technical replicates of each colony. The empty vector control was grown on 2 % glucose (P.p. -Ev +Glucose) and on cellulose (P.p. -Ev) as the positive (square) and negative (circle) controls respectively. Error bars represent two repeats.

5.3 Expression of the ferulic acid metabolic pathway

As previously indicated, the combinatorial strategy developed and outlined in chapter 4 could be successfully employed to vary the expression of genes in a metabolic pathway. To further investigate the applications and the potential of this strategy, the optimisation of another high-value metabolic pathway was sought; the ferulic acid metabolic pathway. This pathway results in the conversion of the compound ferulic acid to coniferyl alcohol. Ferulic acid can be released from lignocellulosic biomass (wheat arabinose xylan) via the xylanase containing the feruloyl domain (CE1). Wheat arabinose xylan was used as the lignocellulosic biomass as it contains a high content of ferulic acid (2 % wt/wt) (Tramontina et al., 2020). Ferulic acid is converted to coniferyl aldehyde via the carboxylic acid reductase (HisCar5). The coniferyl aldehyde is then further converted to the high-value target product coniferyl alcohol

via the aldo-keto reductase (CgAkr-1) (Figure 1.6). The secretion of the feruloyl esterase (CE1) into the extracellular medium is necessary for releasing ferulic acid from wheat arabinose xylan. The CE1 sequence (29.13 kDa) containing a 6xhis-tag was fused to the Pre-Ost1-Pro secretion signal and transformed into *P. pastoris* (*P.p.* –CE1). The strain was then grown in 100 mL BMGY for 48 hours at 30 ºC, 250 rpm and centrifuged to obtain the supernatant. The cell culture supernatant was eluted through a nickel based Histrap column using a peristaltic pump. Ten 1 mL elutions were obtained using buffer C (Chapter 2.4.1). The first 5 elution fractions were analysed on SDS-PAGE (Figure 5.4). Strong bands were detected in elution fraction 1 (E1), 2 (E2) and 3 (E3), approximately at the expected size (30 kDa), indicating CE1 was being secreted. The bands at this size were excised and sent for protein mass-spectrometry analysis with which the CE1 sequence was confirmed (Figure S5.1).

To assess the feasibility of catalysing ferulic acid degradation in the extracellular medium, *P. pastoris* strains containing the enzymes carboxylic acid reductase (*P.p.* – HisCar5) and acid-keto reductase (*P.p.* – CgAkr-1) were also fused to Pre-Ost1-Pro secretion signal and analysed as described above (Figure 5.4). HisCar5 (*P.p.* – HisCar5) was found to be secreted (bands at 130kDa) and its sequence identity was confirmed via protein mass spectrometry (Figure S5.2). However, CgAkr-1 (*P.p.* – CgAkr-1) bands were not detected suggesting no secretion had occurred. Several attempts to optimise the secretion process were made, including pH and temperature modifications, but no secretion was observed.

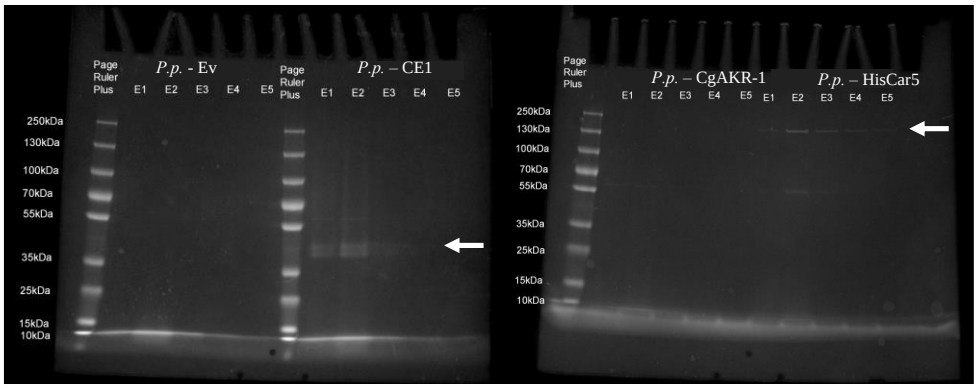

Figure 5.4: SDS-PAGE of extracellular media of P. pastoris expressing genes from ferulic acid degradation pathway. P. pastoris expressing Pre-Ost1-Pro secretion signal fused to a feruloyl esterase (P.p. –CE1), a carboxylic acid reductase (P.p. –HisCar5) and an Aldo-keto reductase (P.p. –CgAKR-1). The strains were cultivated in 100mL BMGY-2% glucose at 20 ºC for 48 hours. Cell culture supernatant was isolated by centrifugation and subjected to protein purification and various elutions were made in buffer C. The control was P. pastoris transformed with an empty vector (P.p. – Ev). The white arrows indicate the presence of CE1 (A) and HisCar5 (B).

5.3.1 TLC for detection of ferulic acid, coniferyl aldehyde and coniferyl alcohol

To be able to analyse the data from the ferulic acid pathway, thin-layer chromatography was initially utilised. The substrate (ferulic acid), the intermediate product (coniferyl aldehyde) and the final product (coniferyl alcohol) were all dissolved in 50 % BMGY 50 % acetonitrile mixture (1 mM final concentration) and spotted (5 uL) on TLC Silica gel 60 F_{254}. The plates would fluoresce under UV light (254 nm exposure). An absence of fluorescence from the silica gel plates would be observed as dark colouration where compounds that absorbed at 254 nm were present. All three compounds were visualised under UV lighting exposure (254 nm) (Figure 5.5). Ferulic acid identification was further visualised using 1 % ferric chloride staining, resulting in browning of the spots. The blank control (50 % BMGY 50 % acetonitrile only) did not produce any darkening indicating absence of other compounds. All three compounds were fluorescent (Figure 5.5) however, the separation between the three compounds was poor when all three compounds were present in the same mixture. Calculating the concentrations of coniferyl aldehyde and coniferyl alcohol would be a difficult process as it would require the calculation of the intensity of ferulic acid after ferric chloride staining. Therefore, HPLC-MS was chosen as a complementary method for compound identification as it allows the clear separation of the compounds as well as the detection of low compound concentrations.

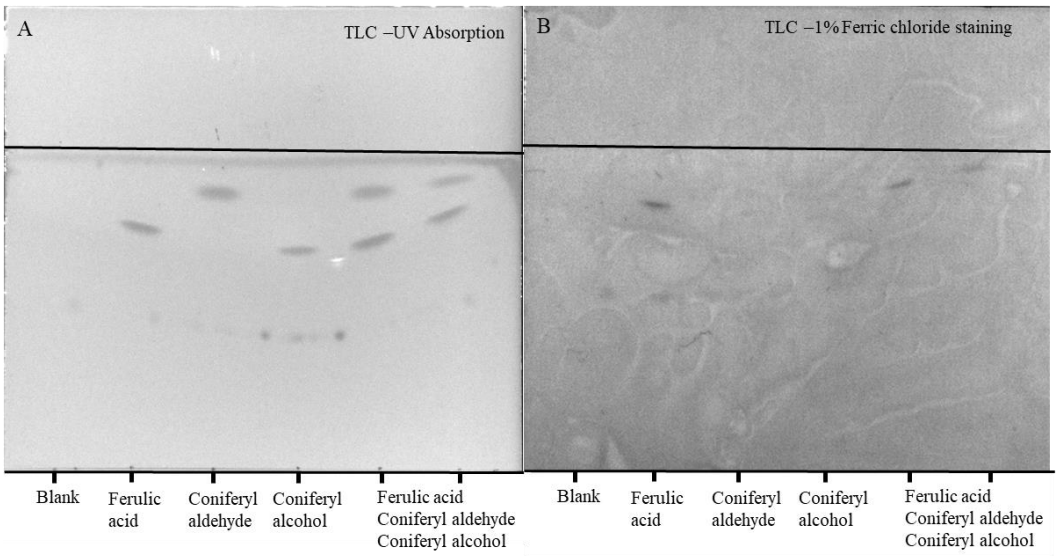

Figure 5.5: TLC analysis of ferulic acid, coniferyl aldehyde and coniferyl alcohol. The compounds ferulic acid, coniferyl aldehyde and coniferyl alcohol (1mM in 100 % acetonitrile) were spotted on TLC plates and developed with chloroform, methanol and formic acid (85:14:1 respectively). The blank is acetonitrile only. The TLC plate was revealed under UV light (A) and stained with 1 % ferric chloride (B).

5.3.2 Preliminary LC-MS data for detection of ferulic acid, coniferyl aldehyde and coniferyl alcohol

The combinatorial strategy method developed in Figure 4.13 was utilised to create strains transformed to contain the pathway allowing them to metabolically convert ferulic acid (from wheat arabinose xylan) to coniferyl alcohol. *P. pastoris* strains were constructed to determine the feasibility of using HPLC-MS for the detection of the three compounds as well as to obtain preliminary data on the production of those compounds. *P. pastoris* was transformed with pPGS60 plasmid containing CgAKR-1 and HisCar5 (*P.p.* –CgCar) (Figure 5.7A). *P.p.* -CgCar contained *GAP* promoter-regulated genes HisCar5 and CgAkr-1.

Another strain was created by transforming *P. pastoris* with pPGS80 plasmid containing a secretion signal bound CE1 (*P.p.* –CE1) (Figure 5.7B). *P.p.* –CE1 contained the Pre-Ost1-Pro secretion signal fused to the CE1 gene for extracellular expression (Figure 5.7) under the regulation of *GAP* promoter to determine if ferulic acid could be produced by the expression of CE1.

Both strains were grown for 24 hours in 100 mL BMGY, pH 6 at 30^{o}C, 250 rpm. Wheat arabinose-xylan was added to *P.p.* –CE1 at a final concentration of 2 g/L to observe the release of ferulic acid via the CE1 gene. Ferulic acid was added to *P.p.* –CgCar at a final concentration of 5 mM to observe the metabolic conversion of ferulic acid to coniferyl alcohol. The cell cultures were then grown for a further 24 hours, centrifuged at 3400 rpm and cell culture supernatant was isolated. Metabolite extraction from all samples was carried out via liquid-liquid organic phase extraction and HPLC-MS analysis followed. Ferulic acid and coniferyl alcohol were used as references at a concentration of 5 mM in 50% MeOH and their peaks were identified. In the strain *P.p.* –CE1, a peak was detected at the retention time similar to that of the ferulic acid control (Figure 5.4A-B), suggesting the production of ferulic acid. This was further confirmed by the absence of a peak at the corresponding retention time for the empty vector control (*P.p.* –Ev). Ferulic acid was detected in both *P.p.* –Ev and *P.p.* –CgCar. However, no peak was detected for Coniferyl alcohol compared to the control (Figure 5.6D). To investigate the reason for the lack of coniferyl alcohol or coniferyl aldehyde production in the *P. pastoris* strains, HPLC analysis using a diode

array detector was carried out. The diode array detector measures absorbance at specific wavelengths. Coniferyl aldehyde, coniferyl alcohol and ferulic acid all absorb at 280 nm. Coniferyl aldehyde and ferulic acid also absorb at 340 nm whereas coniferyl alcohol does not, allowing for discrimination between coniferyl alcohol and the other two compounds in a cell culture sample.

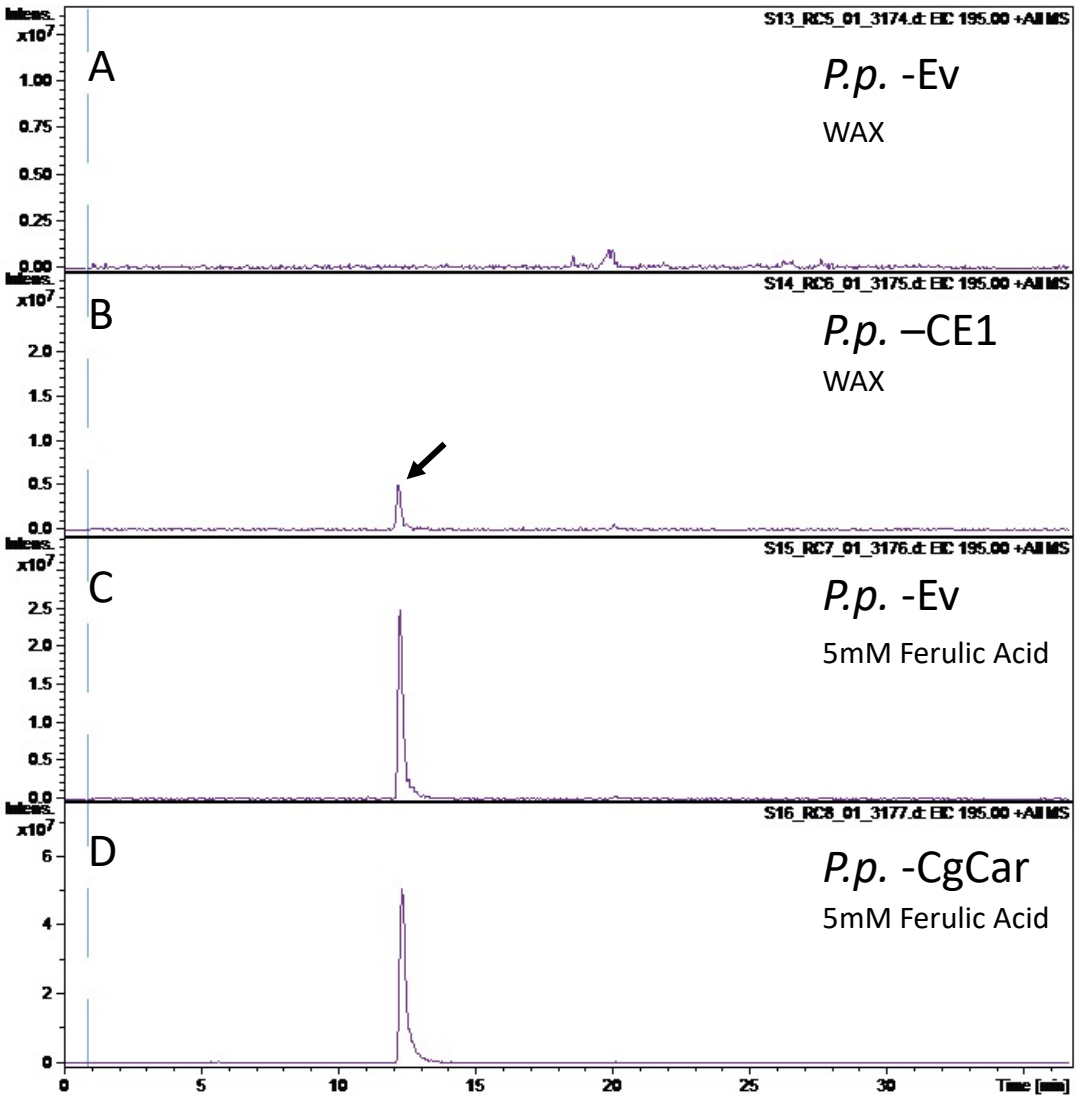

Figure 5.6: HPLC-MS chromatogram of Ferulic acid identification in P. pastoris strains. (A)P.p. –Ev control containing Wheat arabinose-xylan (WAX) (2g/l). (B) P.p. –CE1 with the addition of WAX (2g/l). The black arrow indicates the presence of ferulic acid. (C) P.p. –Ev culture with the addition of 5 mM ferulic acid. (D) P.p. –CgCar culture with the addition of 5 mM ferulic acid.

5.3.3 Investigation of coniferyl alcohol production in *P. pastoris*

To investigate the absence of coniferyl alcohol production in *P. pastoris*, HPLC-DAD was utilised. The controls were set at 1 mM concentration in a mixture of 50 % BMGY and 50 % acetonitrile (Figure 5.7). The chromatographic run was performed for the

controls using a diode array detector (DAD) as all three compounds absorbed UV at wavelength 280 nm. Both ferulic acid and coniferyl aldehyde also absorbed at wavelength 340 nm whereas coniferyl alcohol did not. This confirmed the presence of coniferyl alcohol in the control sample as identified by the peaks at 280 nm and the absence of the peaks at 340 nm wavelengths at the corresponding retention times (rt) (Figure 5.7).

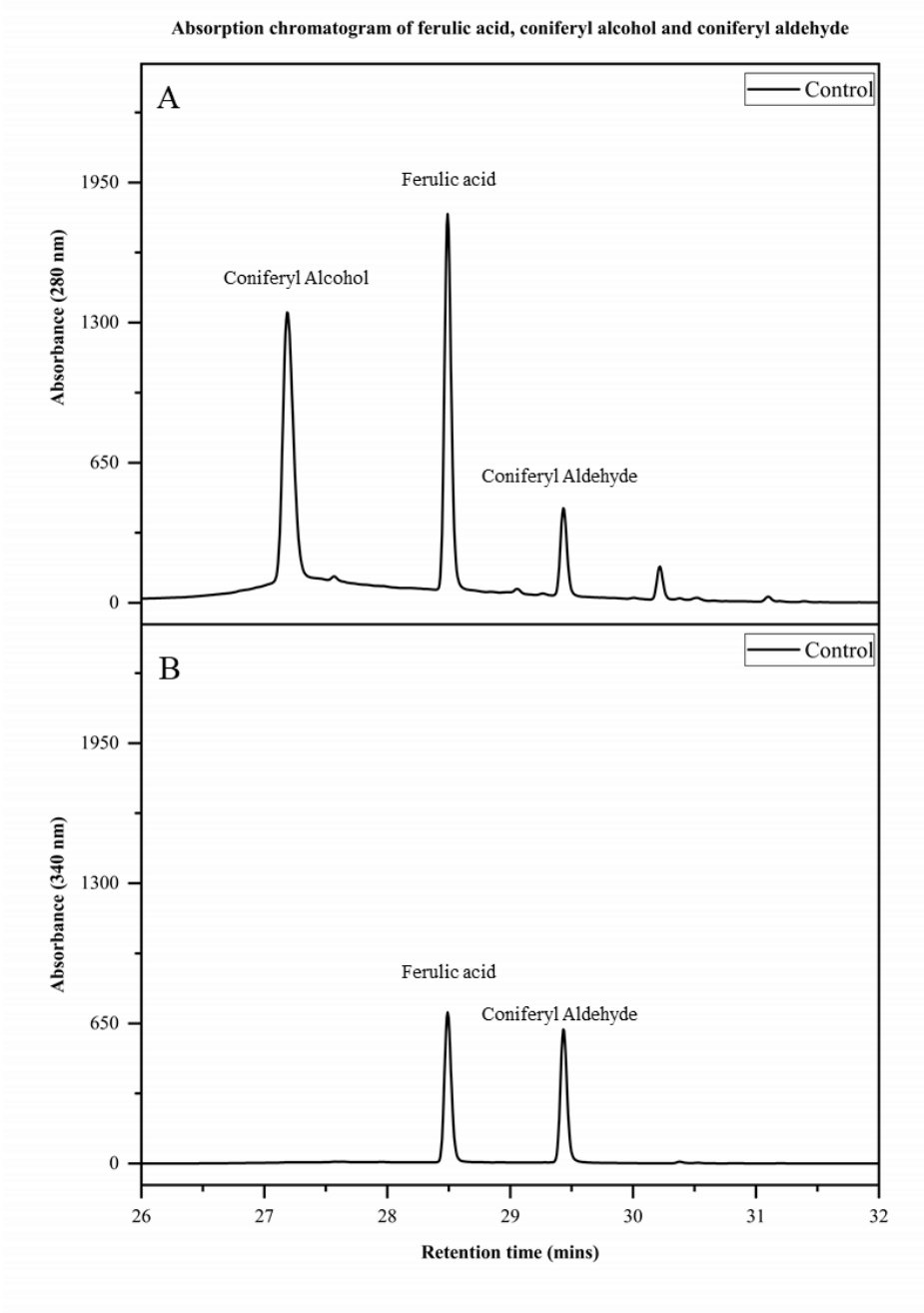

Figure 5.7: HPLC chromatogram of ferulic acid, coniferyl alcohol and coniferyl aldehyde. The compounds ferulic acid, coniferyl alcohol and coniferyl aldehyde were mixed at 0.33mM final concentration in 50 % BMGY and 50 % acetonitrile. The absorptions of these compounds were measured at wavelengths 280 nm (A) and 340 nm (B).

P. pastoris was co-transformed with the plasmids pPGS60 containing the two genes for the enzymes HisCar5, CgAkr-1 and the plasmid pPGS80 containing the genes for enzyme CE1 fused to a secretion signal (Figure 5.8), creating the strain *P.p.* –CgHis. The strain was grown in 50 mL BMGY media at 30 ^{o}C and 250 rpm. After 24 hours, ferulic acid (1 mM final concentration) was added to the cell cultures. The cultures were then grown for a further 72 hours and 1 mL samples were taken at various time points. Timepoints were set at 8 hours, 24 hours, 48 hours and 72 hours after the addition of ferulic acid. Compound extraction at these timepoints was carried out as described in Materials and Methods (section 2.1.2.3). *P.p.* -Ev and *P.p.* -CgHis cultures containing 1 mM ferulic acid generated a single peak at the timepoints of 8, 24, 48 and 72 hours (Figure 5.9). The peak shifted slightly from the control compound ferulic acid peak (rt 28.49 minutes, Figure 5.8) at timepoints 8 hours (rt 28.04 minutes) and 48 hours (rt 28.07 minutes). However, these peaks were also detected at wavelength 340 nm, indicating that the peaks may not be due to the presence of coniferyl alcohol (Figure 5.10). Similar findings were also observed for *P.p.* -Ev (Figure 5.6). To investigate whether the 1 mM concentration of ferulic acid is toxic to *P. pastoris*, resulting in conversion of ferulic acid to another less toxic compound, cultures were set up with 0.5 mM final concentration of ferulic acid. When running these through HPLC, a single peak at both 280 and 340 nm was observed for both *P.p.* -Ev and *P.p.* -CgHis in all time points. This indicates that at this low concentration of 0.5 mM, ferulic acid was not converted with readily detectible efficiency to coniferyl alcohol.

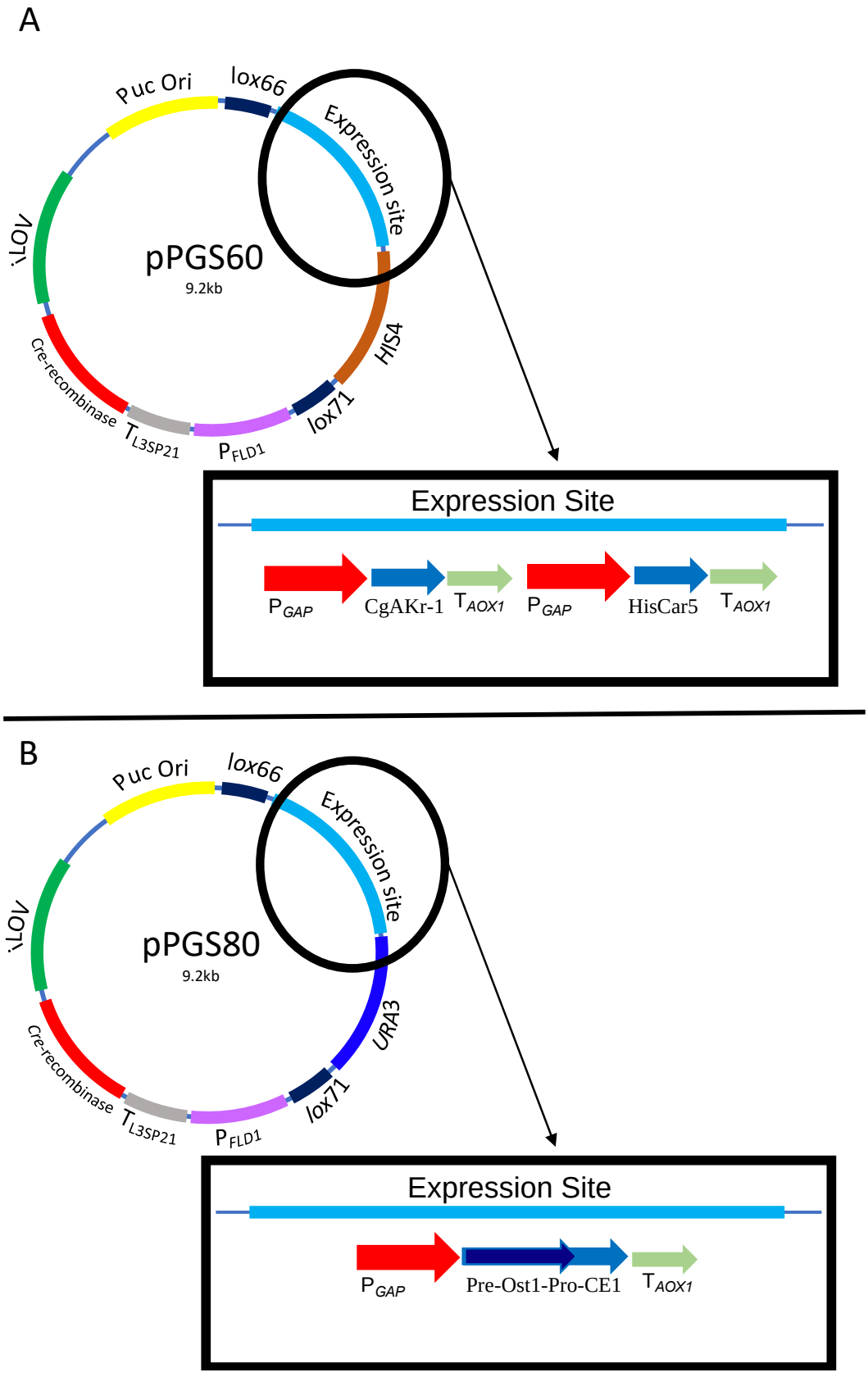

Figure 5.8: Plasmids containing CgAKr-1, Hiscar5 and CE1. The plasmid pPGS60 expression site contained the CgAKR-1 and HisCar5 sequence regulated by P_{GAP}. The plasmid pPGS80 expression site contained the sequence CE1 bound to secretion signal sequence Pre-Ost1-Pro regulated by P_{GAP}. The plasmids were co-transformed into P. pastoris and the markers were recycled.

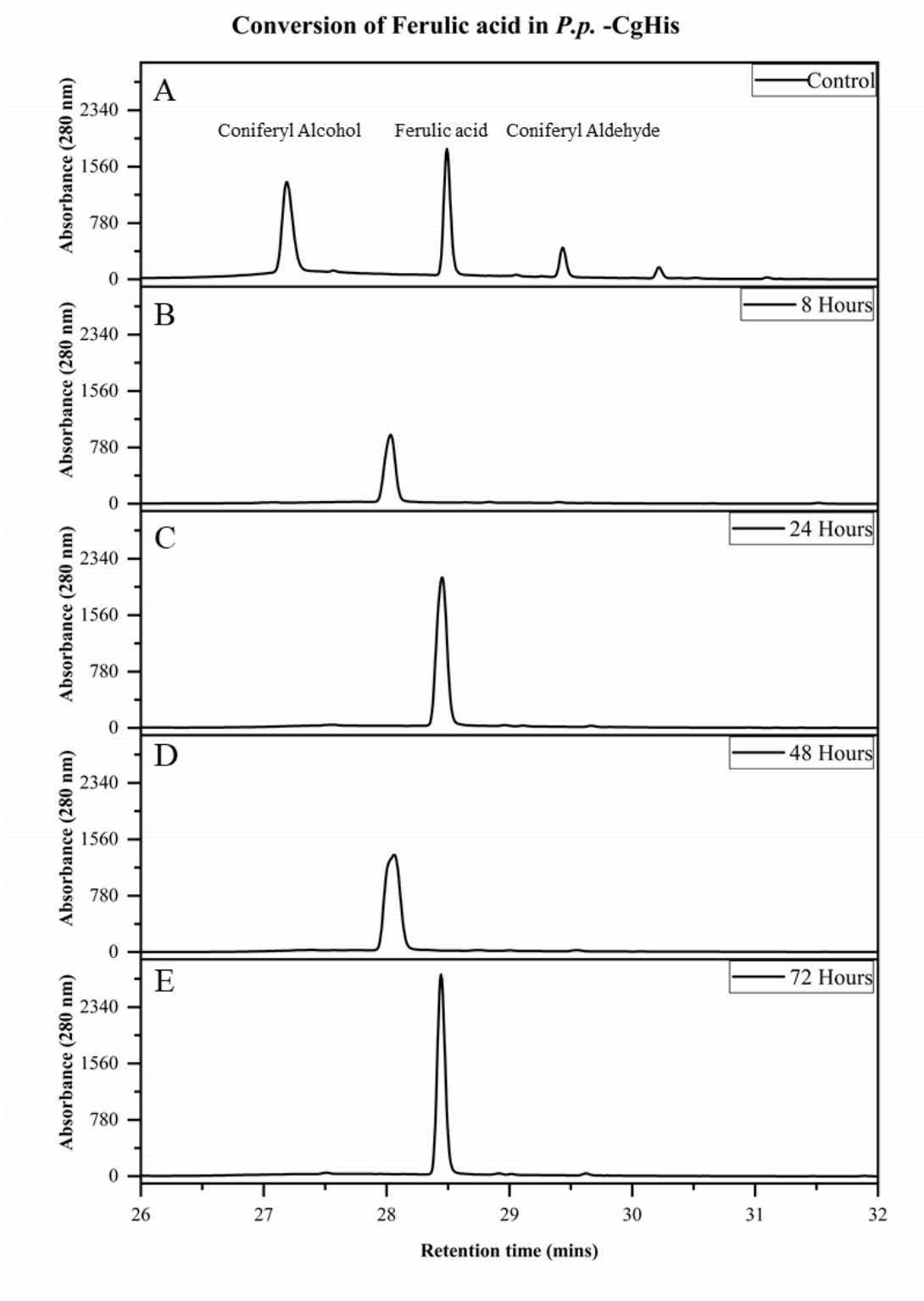

Figure 5.9: HPLC chromatogram of conversion of ferulic acid in P.p. – CgHis at absorption 280nm. The strain P.p. -CgHis was grown for 24 hours in BMGY. The compound ferulic acid was then added at a final concentration of 1mM. Samples were taken and metabolites extraction carried out using acetonitrile at time points of 8 hours (B) 24 hours (C) 48 hours (D) and 72 hours (E). The absorbance was measured at 280 nm. The controls (A) were 0.33mM final concentration of coniferyl alcohol, ferulic acid and coniferyl aldehyde in 50 % BMGY and 50 % acetonitrile.

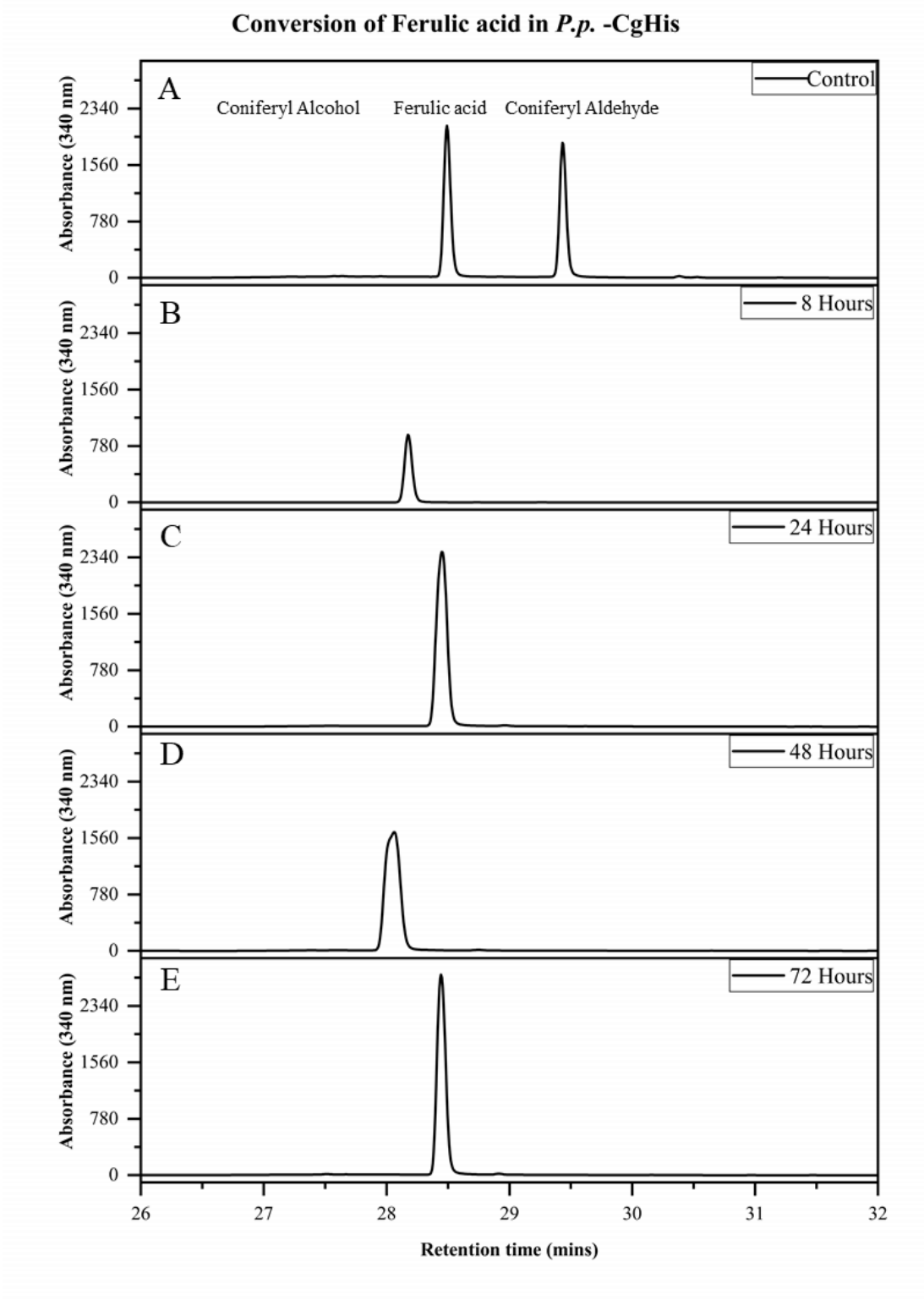

Figure 5.10: HPLC chromatogram of conversion of ferulic acid in P.p. – CgHis at absorption 340nm. The strain P.p. -CgHis was grown for 24 hours in BMGY. The compound ferulic acid was then added at a final concentration of 1mM. Samples were taken at and metabolites extraction carried out using acetonitrile at time points of 8 hours (B) 24 hours (C) 48 hours (D) and 72 hours (E). The absorbance was measured at 280 nm. The controls (A) were 0.33mM final concentration of coniferyl alcohol, ferulic acid and coniferyl aldehyde in 50 % BMGY and 50 % acetonitrile.

In order to investigate whether enzymatic release of ferulic acid from wheat arabinose in the presence of *P.p.* –CgHis could result in production of coniferyl alcohol, cell cultures of *P.p.* –CgHis and *P.p.* –Ev were grown for 24 hours before the addition of wheat arabinose (2 g/L). The absorbance of samples was measured via HPLC and peaks were generated at rt 28.99, rt 28.79 and rt 28.49 minutes for the *P.p.* -CgHis sample at time points 24, 48 and 72 hours respectively but not in the empty vector strain sample *P.p.* -Ev (Figure 5.11). Furthermore, these peaks were similar to the control peaks (ferulic acid, rt 28.49 minutes, Figure 5.11A and B) indicating the possible production of ferulic acid by *P.p.* –CgHis. Even though *P.p.* –CgHis may be able to produce ferulic acid, the presence of the peaks in both 280 nm and 340 nm for all the timepoints (Figure 5.9 and Figure 5.10 respectively) suggests a lack of conversion of ferulic acid to coniferyl alcohol. These findings indicate that either *P.p.* –CgHis was unable to convert ferulic acid to coniferyl alcohol or the compounds coniferyl aldehyde and coniferyl alcohol may be converted to different compounds by *P. pastoris*.

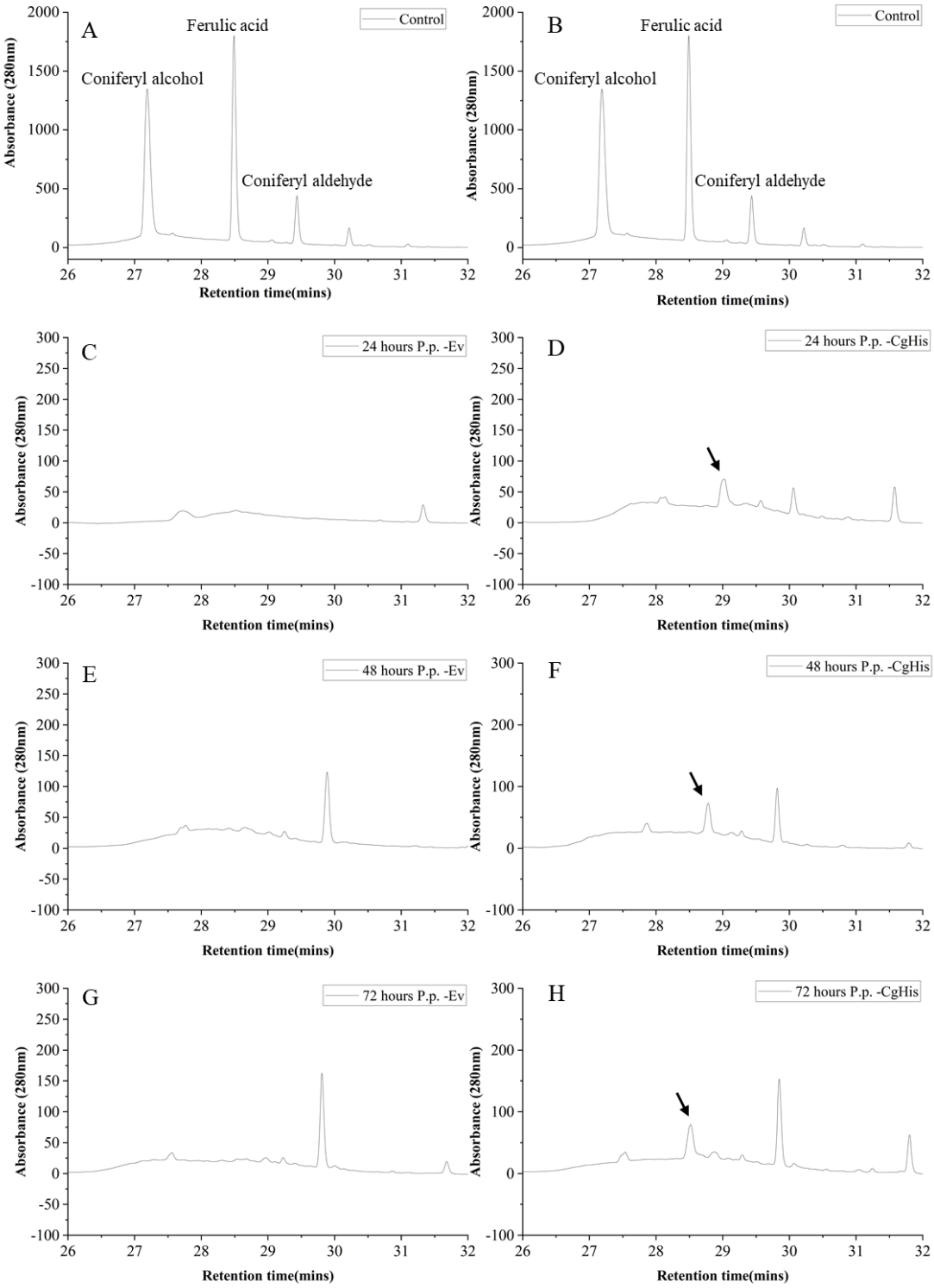

Figure 5.11: HPLC chromatogram of conversion of wheat arabinose in P.p. – CgHis. The strains P.p. -CgHis and P.p. -Ev were grown for 24 hours in BMGY. Wheat arabinose was then added at a final concentration of 1 g/l. Metabolite extraction was carried out using acetonitrile at time points of 24 hours (C,D) 48 hours (E,F) and 72 hours (G,H). The absorbance was measured at 280 nm. The controls (A,B) were 0.33mM final concentration of coniferyl alcohol, ferulic acid and coniferyl aldehyde in 50 % BMGY and 50 % acetonitrile. The black arrow represents the peaks that indicate the presence of ferulic acid.

To determine whether the compounds coniferyl aldehyde or coniferyl alcohol were being converted to a different compound by *P. pastoris*, the strains *P.p.* -CgHis and *P.p.* -Ev were cultured with the addition of these compounds separately. The samples were analysed via HPLC. In more detail, coniferyl alcohol and coniferyl aldehyde were added to *P.p.* –CgHis at 1 mM concentration and sampling time points were set at 12, 24, 48 and 72 hours. Ferulic acid, coniferyl alcohol and coniferyl aldehyde were added in combination to *P.p.* –CgHis at a final concentration of 0.33 mM each, in order to observe their potential conversion to different compounds. The *P.p.* –CgHis cultures containing coniferyl aldehyde did not lead to the production of coniferyl alcohol as the peaks were observed at both 280 and 340 nm wavelengths (Figure 5.12). At the time point 8 hours, the coniferyl aldehyde peak had shifted from rt 29 minutes to rt 28 minutes indicating that coniferyl aldehyde was converted prior to that time point. Similar findings were observed in the respective *P.p.* –Ev (Figure S5.3).

In the sample taken from the *P.p.* –CgHis culture with added coniferyl alcohol, the peak shifted from rt 27 minutes to rt 28 minutes after the 8-hour time point (Figure 5.13). This peak (rt 28 minutes) was similar to the one observed in the control sample for ferulic acid (Figure 5.6, rt 28 minutes). The control compound coniferyl alcohol (Figure 5.12) generated a peak at 280 nm but not at 340 nm. However, when sampling *P.p.* –CgHis with added coniferyl alcohol at time point 8 hours, the peak had become visible at both wavelengths 280 nm and 340 nm. Similar results were also found in the respective *P.p.* –Ev (Figure S5.4). This suggests that even if coniferyl alcohol is produced by *P.p.* –CgHis, it may be converted to another compound prior to that time point.

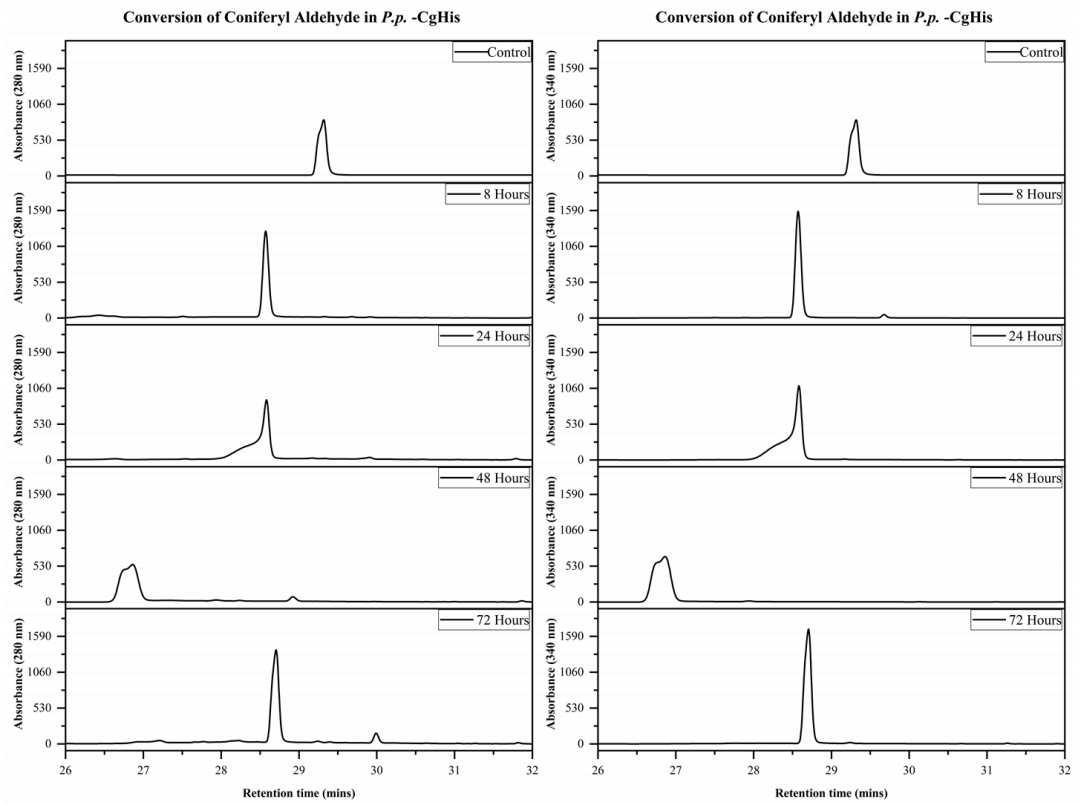

Figure 5.12: HPLC chromatogram of conversion of coniferyl aldehyde in P.p. – CgHis. The strains P.p. -CgHis and P.p. -Ev were grown for 24 hours in BMGY. Coniferyl aldehyde was then added at a final concentration of 1mM. Metabolite extraction was carried out using acetonitrile at time points of 24, 48 and 72 hours. The absorbance was measured at 280 (left) and 340 (right) nm. The controls (A) were 1mM final concentration of coniferyl aldehyde in 50 % BMGY and 50 % acetonitrile.

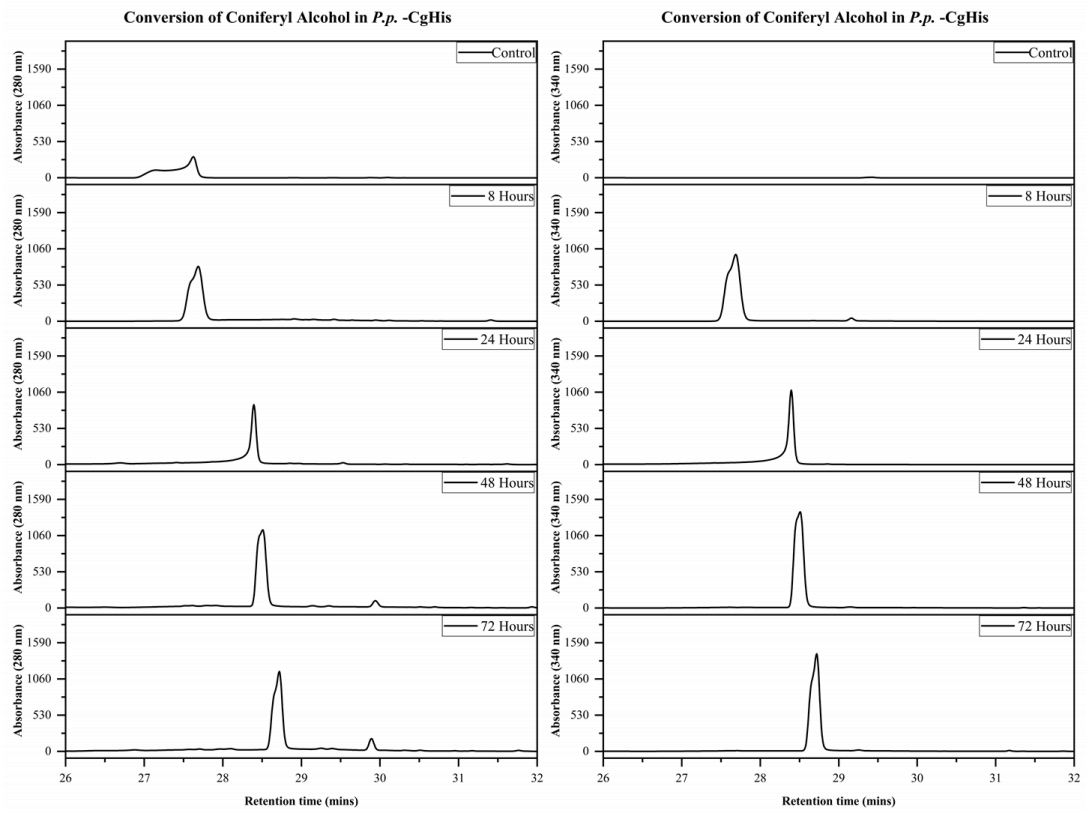

Figure 5.13: HPLC chromatogram of degradation of coniferyl alcohol in P.p. – CgHis. The strains P.p. -CgHis and P.p. -Ev were grown for 24 hours in BMGY. Coniferyl alcohol was then added at a final concentration of 1mM. Metabolite extraction was carried out using acetonitrile at time points of 24, 48 and 72 hours. The absorbance was measured at 280 (left) and 340 (right) nm. The controls (A) were 1mM final concentration of coniferyl aldehyde in 50 % BMGY and 50 % acetonitrile.

The conversion of coniferyl alcohol in the *P.p.* –CgHis samples appeared to have occurred before the 8-hour time point. Therefore, to investigate the time point at which the conversion may occur, *P.p.* –CgHis culture containing coniferyl alcohol was subjected to a shorter time point sampling protocol. *P.p.* –CgHis culture samples were taken 0.5, 1, 2 and 3-hours after the addition of coniferyl alcohol. The compounds were extracted and analysed via HPLC (Figure 5.14). Coniferyl alcohol was already being converted to a second unknown compound at time point 0.5-hours as indicated by the presence of a second peak (rt 29 minutes). At time point 1-hour post coniferyl alcohol addition, three peaks were observed (Figure 5.14). The first peak at around rt 27 minutes corresponds to coniferyl alcohol since no peak was observed at the respective time at 340 nm. The second peak (rt 28 minutes) was observed at both 280 and 340 nm, indicating the conversion of coniferyl alcohol into another compound. The third peak was generated at rt 29 minutes and was present at both 280 and 340 nm also confirming the conversion of coniferyl alcohol into another compound. At

timepoints 2 and 3-hours, the absorbance of the samples at 280 nm changed, with the first peak (coniferyl alcohol, rt 27 minutes) and the third peak (unknown, rt 29 minutes) decreasing in size whilst the second peak (unknown, rt 28 minutes) increasing in size. This indicates the conversion of coniferyl alcohol at 3-hour time point, with the formation of the single peak at rt 28 minutes (present in both 280 nm and 340nm). Similar findings were observed in the respective *P.p.* –Ev (Figure S5.5) These findings suggest that coniferyl alcohol is rapidly converted to a different compound.

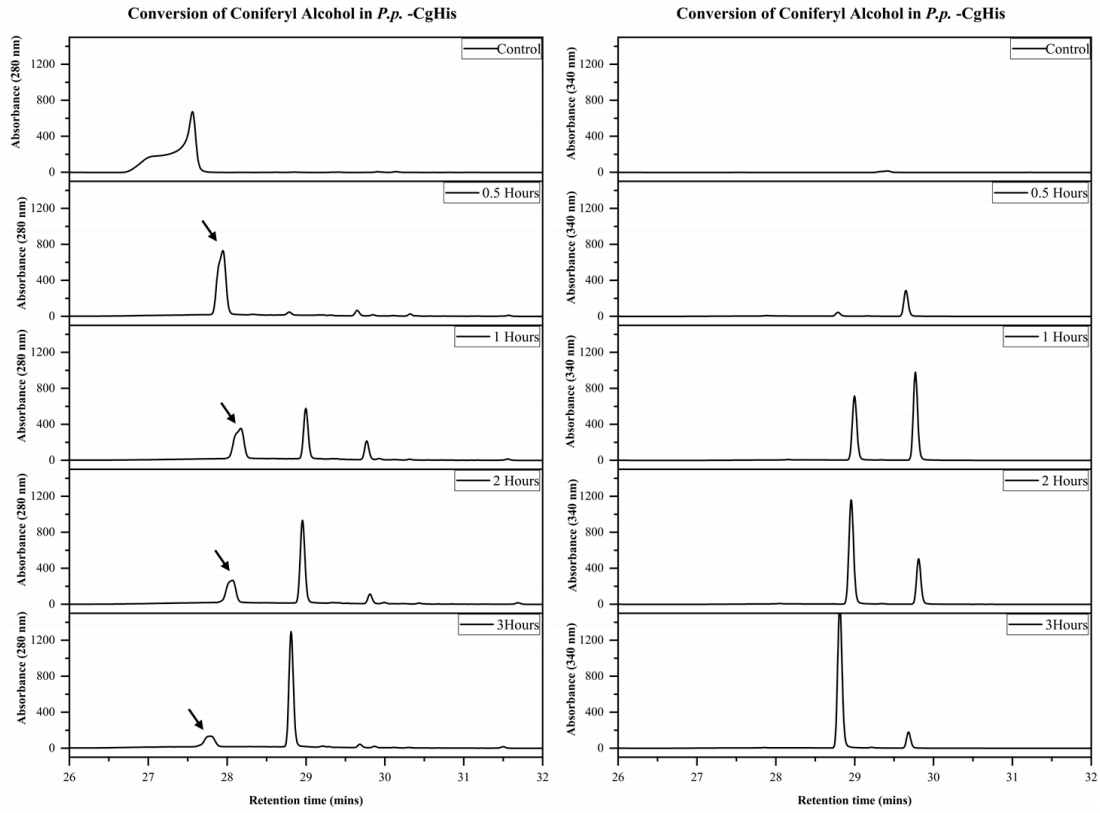

Figure 5.14: HPLC chromatogram of degradation of coniferyl alcohol in P.p. – CgHis. The strains P.p. -CgHis and P.p. -Ev were grown for 24 hours in BMGY. Coniferyl alcohol was then added at a final concentration of 1mM. Metabolite extraction was carried out using acetonitrile at time points of 0.5, 1, 2 and 3 hours. The absorbance was measured at 280 (left) and 340 (right) nm. The controls (A) were 1mM final concentration of coniferyl alcohol in 50 % BMGY and 50 % acetonitrile. Black arrow represents the peak corresponding to coniferyl aldehyde.

To investigate whether the compounds (coniferyl alcohol and coniferyl aldehyde) were being converted to ferulic acid or to carboxylic group containing acids in *P.p.* –CgHis, TLC analysis using ferric chloride was carried out. The *P.p.* –CgHis HPLC

samples incubated with either ferulic acid, coniferyl alcohol or coniferyl aldehyde from timepoints 0.5 hours to 24 hours were spotted as 5 uL spots on TLC plates (Figure 5.15). The controls ferulic acid (FA), coniferyl aldehyde (CA) and coniferyl alcohol (Col) were made to 1 mM concentration in 50 % BMGY and 50 % acetonitrile. The TLC plate fluoresced under UV_{254nm} light. The presence of the compounds was indicated by the darkening of the plate as a result of the compounds absorbing at 254 nm and preventing the TLC plate from fluorescing. Dark spots were observed for all the samples as expected under UV_{254nm} light exposure (Figure 5.14A). Upon staining with 1 % ferric chloride solution, the controls FA appeared to have dark spot formation whereas the controls CA and Col did not (Figure 5.14B). Dark spots in the samples from *P.p.* – CgHis with the addition of CA, Col or mixture all three of FA, CA and Col were observed. The intensity of those dark spots increased over time after the addition of compound, suggesting the possible conversion of CA and Col to a carboxylic acid containing compound.

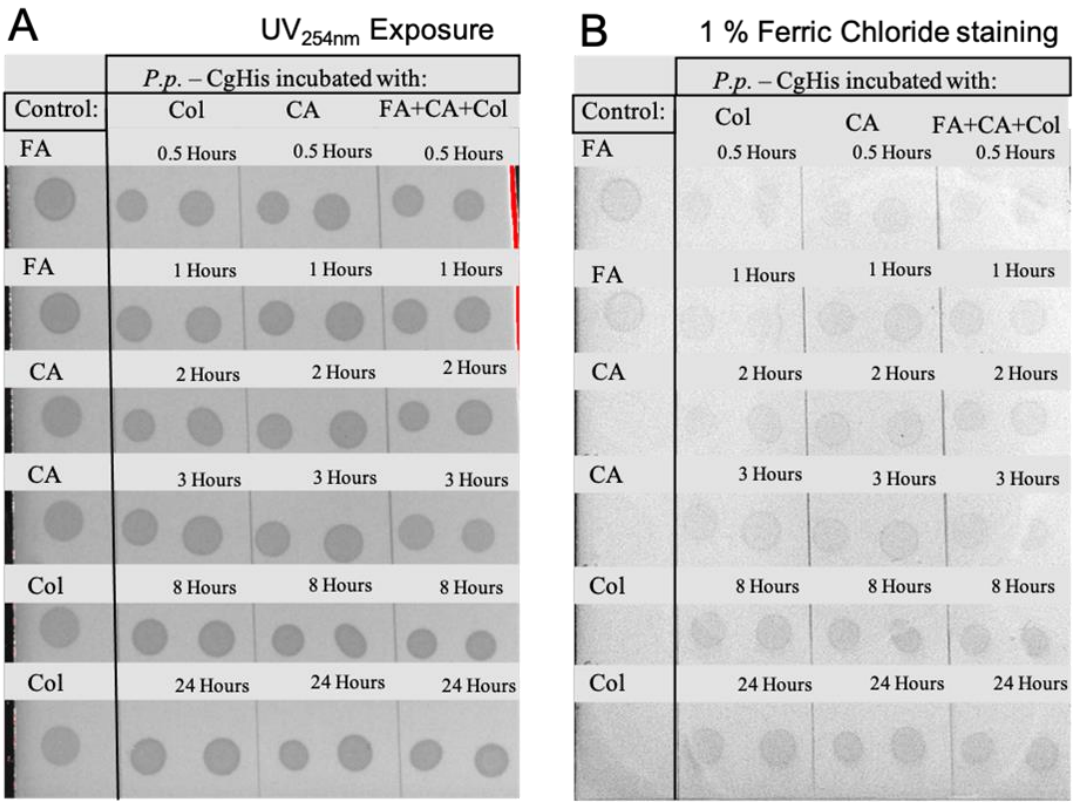

Figure 5.15: TLC analysis of ferulic acid, coniferyl aldehyde and coniferyl alcohol. The P. pastoris strain P.p. – CgHis was grown for 24 hours. The compounds coniferyl alcohol (Col), Coniferyl aldehyde (CA) or Ferulic acid (FA) were added to the cultures. Samples were taken at various time-points and spotted on the TLC plate. (A) The TLC plate was subjected to UV_{254nm} light exposure where compounds absorbing at that wavelength resulted in the darkening of the plate. (B) The TLC plate was stained using 1 % ferric chloride which resulted in the darkening of compounds containing carboxylic acid group. The controls were FA, CA and Col as 1 mM spots in 50 % BMGY and 50 % acetonitrile.

In summary, the combinatorial strategy for metabolic pathway optimisation previously described was used to optimise various pathways in *P. pastoris.* Initially, fluorescent proteins were used to highlight the potential of the combinatorial strategy via the observation of variation in fluorescence expression of the FPs. Furthermore, the strategy was successfully used to optimise a cellulose degradation pathway indicating an optimal combination of the expression of cellulase enzymes. The combinatorial strategy was then used to optimise ferulic acid degradation pathway, however, a lack of conversion of ferulic acid to coniferyl alcohol was observed. Further investigations were carried out via HPLC which suggested that *P. pastoris* may be converting coniferyl alcohol and coniferyl aldehyde to alternative compounds. It was not possible to investigate these alternative compounds any further because of the pandemic-related lockdown.

Chapter 6 – Discussion

6.1 An optimal organism for recombinant protein production

Recombinant protein production is of great interest to the biotech industry due to the need and value of proteins in a wide range of commercial applications. More than 170 recombinant proteins (2018) have been produced worldwide to be used for medicinal purposes in humans (Pham, 2018). Recombinant proteins have also been used in the biotech industry to maximise manufacturing technologies in order to reduce cost-of-goods such as the use of cellulase proteins to produce bioethanol (Puetz and Wurm, 2019). The production of recombinant proteins in the form of enzymes is highly desirable as they can be used in metabolic pathways to degrade substrates into valuable products (Jozala et al., 2016). Secretion of proteins is of particular interest as it eliminates the time consuming and expensive cell lysis steps. Therefore, it is important to obtain a microbial host that is capable of high levels of protein production and secretion. *S. cerevisiae* and *P. pastoris* are two yeasts that have capabilities to secrete proteins via their specialised secretory pathways (Hou, K. E. J. Tyo, et al., 2012). Compared to other eukaryotes and prokaryotes, these two yeasts have been amongst some of the preferred microorganisms in the biotech industry. This has been a result of being entitled the 'Generally regarded as safe (GRAS)' status, having a plethora of genetic modification research available and historically being reliably used in industry for recombinant protein production. (Niazi, 2002).

In chapter 3, I investigated the capabilities of the two yeasts *S. cerevisiae* and *P. pastoris* for their ability to produce and secrete proteins by using the fluorescent reporter ymNeonGreen. The fluorescent reporter ymNeonGreen was selected because it is easily and conveniently assayable (fluorescent-based assays) due to its stability and fluorescence emission intensity. The results in chapter 3 showed *P. pastoris* to be more suitable as a host for recombinant protein production between the two yeasts. The results in this chapter indicated that compared to *S. cerevisiae, P. pastoris* was: (1) able to grow to higher optical density (OD_{600}), (2) better at secreting ymNeonGreen, and (3) able to cleave the alpha-factor mating secretion signal tag.

6.1.1 *P. pastoris* as an optimal host for recombinant protein production

Optimal conditions or results can be considered conditions that are favourable to the biotech industry which may include the temperature of culturing, obtaining highest possible yield or products from the strain, reducing the number of consumables as well as other parameters used. The use of optimisation can refer to making a process or task as effective as possible. *P. pastoris* was found to achieve higher OD_{600} compared to *S. cerevisiae* (Figure 3.1, Figure S3.1). A higher OD_{600} is beneficial to the industry as, under optimal conditions, it can result in a higher concentration of protein being produced. Strategies are sought to improve cell density of cultures to improve recombinant protein production and it is a key part of the biotechnology sector (Yang and Zhang, 2018). A study by Vu, (2009) looked at optimising growth conditions for a strain of *S. cerevisiae*, reporting batch culture dry cell weight of 36.5 g/L and fed-batch culture cell dry weight of 187.63 g/L. Another study by Eugene Raj et al. (2002) reported dry cell weight of *S. cerevisiae* strains as high as 140 g/L. However, in comparison, a study by X. Li et al. (2013) reported 394 g/L fed-batch dry cell weight for *P. pastoris* strains as a result of optimising conditions (that would provide highest dry cell weight). It has been reported that *P. pastoris* can achieve high dry cell weight (>100 g/L) in a simple low-cost medium (Liu et al., 2016). Furthermore, *P. pastoris* has been reported to reach up to 13 g/L dry cell weight even in 96 deep-well plates (0.6 mL volume) which could be particularly useful at the initial optimisation stage (Kaushik et al., 2020).

Both yeasts reported intracellular expression of YmNeonGreen (ymNG) (Figure 3.2), but, whilst *P. pastoris* showed secretion of ymNG into the media, *S. cerevisiae* did not (Figure 3.3). A study by (Li et al., 2002) looked at secreting GFP or GFP fusions using different secretion signals in *S. cerevisiae*. However, the authors were unable to obtain extracellular expression of GFP. Confocal studies revealed that GFP was mostly retained in the endoplasmic reticulum (ER)/Golgi compartments. Improvements and optimisations can be made to secrete GFP in *S. cerevisiae*, such as those made by (Huang and Shusta, 2008). In this case, the authors found that S. cerevisiae transformed with yeast-enhanced-GFP cultured at 20 ^{o}C resulted in secretion of the GFP. However, at 20 ^{o}C, the growth of the yeasts would be slower and so would

require longer cultivation periods for high yields, something which is not ideal for industrial use. The secretion of fluorescent proteins is important as it can be used as a reporter for the levels of recombinant protein being produced. Duellman et al. (2015) were able to quantify the levels of recombinant proteins being produced by fusing mCherry to a recombinant protein, bypassing the need for western blot or SDS-PAGE methods. Furthermore, the use of this fluorescent protein fusion allows the high-throughput screening of secretion optimisation parameters in assay development.

S. cerevisiae has also been reported to produce hyperglycosylated proteins that may impact the structure or size of the recombinant proteins. This may have been a limiting factor in enabling the secretion of ymNG in this work. A study by Tran et al. (2017) investigated the production of a human granulocyte-macrophage colony-stimulating factor and found its molecular weight to range between 30-67 kDa in *S. cerevisiae.* This suggests extensive glycosylation in *S. cerevisiae* compared to *P. pastoris* (28-43 kDa). When comparing the total secreted protein/L the authors found it was 4.4-fold higher in *P. pastoris* compared to *S. cerevisiae*, suggesting that *P. pastoris* is a better host for recombinant protein secretion. These studies highlight that *P. pastoris* has better protein secretion ability than *S. cerevisiae,* findings that align with the work presented in this thesis (Chapter 3).

The extracellular expression of ymNG in *P. pastoris* was found to be low compared to its intracellular expression (Figure 3.3). Similar findings were also reported by Barrero et al. (2018) when observing the expression of an RFP and a monomeric GFP. The authors used *P. pastoris* to secrete these fluorescent markers using the mating alpha factor, where a low ratio of extracellular to intracellular expression was observed. The authors reported that fluorescent proteins may be arrested along the secretory pathway. This occurs because fluorescent proteins are either failing to cross the ER membrane as a result of protein folding prior to post-translational translocation, or they may be aggregating in the ER lumen as the pro-region of the mating alpha-factor may have a tendency to self-associate. This potentially indicates a reason why ymNG fails to secrete in *S. cerevisiae* or is secreted at low levels in *P. pastoris*. Approaches were implemented to optimise extracellular expression using a range of tuneable expression elements in *S. cerevisiae.* Various promoters and 5'UTR

combinations were implemented alongside the two secretion signals and although variation in intracellular expression was observed, no secretion was detected. The combination of expression elements (5'UTR and promoters) suggested that it was possible to vary the expression of genes by either using different promoters or by combining different 5'UTRs. The use of 5'UTRs is effective in varying expression in *S. cerevisiae* (Hoshida et al., 2017). Hoshida et al. (2017) studied the introns within the 5'UTR and found that alterations in their sequence led to a 50-fold range in the expression of genes. The results presented in this thesis combined with the data from literature indicate that *P. pastoris* is a more optimal host for extracellular expression between these two yeasts and could be utilised for metabolic pathway integration.

6.1.2 Expression of cellulases in *P. pastoris*

The expression of the cellulose degradation pathway was carried out in *P. pastoris* based on the data indicating it as a better host for extracellular secretion. Seven cellulase proteins had been identified (Table 3.2) and produced in *P. pastoris*. However, only three cellulases were found to secrete into the cell culture medium (Figure 3.8). These cellulases were the exo-glucanase ExG2, the ß-glucosidases BG2 and PaBG1b. Since extracellular expression is important in being able to degrade cellulosic biomass, further optimisation was performed to improve the secretion of the three remaining cellulases, however, this proved unsuccessful. For reasons not well understood, *P. pastoris* may not secrete proteins even if they contain a secretion signal (Lin-Cereghino et al., 2013b). However, there may be several reasons for the absence of cellulase secretion. It has been reported that msGFP folds poorly in the ER lumen as a result of the formation of non-native disulphide bond formation (Fitzgerald and Glick, 2014) and because these cellulases contained multiple cysteines this may result in poor protein folding. It has also been reported that the deletion of the vacuole sorting receptor (Vps10) domain in *S. cerevisiae* prevented the diverting of msGFP to the vacuole, resulting in higher concentrations of secretion products. This approach could also be utilised in *P. pastoris* to increase recombinant protein secretion. These possibilities suggest that amino acid mutations or alterations in the cellulase sequences could be applied to improve secretion, however, it would require a thorough investigation into effects of the mutations in the cellulase sequences to assess whether the cellulases still retain activity.

To assess the feasibility of using the secreted cellulases for cellulosic biomass degradation, the extracellular medium was isolated and evaluated for various factors. An important consideration is that secreted cellulases might need to be used directly in the extracellular medium. It was therefore important to determine the cellulase activities under these exact conditions. The secreted cellulase media containing two of the three cellulases (exoglucanase and ß-glucosidase) necessary for cellulose degradation were added in varying ratios to solutions containing amorphous cellulose (similar to lignocellulosic biomass). The expected outcome would be the generation of varying amounts of glucose, based upon mixtures at different ratios of the two cellulases. However, only low concentrations (up to 0.08 mg/mL) of glucose were detected (Figure 3.10). One possible reason is the lack of endo-glucanase, because the endo-glucanase in this work (EnG1) failed to be generated extracellularly (Figure 3.8). Eveleigh et al., 2009 reported glucose concentrations of over 1 mg/mL when using the combination of all three cellulases. EnG1 has a theoretical PI of 8.92, significantly higher than that of the culture media pH at 6, suggesting that the low pH might be the reason for the smearing observed and the lack of extracellular expression (Figure 3.8) Another reason for the low glucose concentration observed may be due to the low activity of the ß-glucosidase (BG2). PaBG1b reported high activity towards pNPG in this work, finding also previously reported by (Y. Li et al., 2017). In comparison, BG2 showed 10-fold lower activity (Figure 3.10). It was found that an increase in the BG2 ratio in comparison to ExG2 increased the concentration of glucose produced. This may be a result of the low activity of BG2 (Figure 3.10) as well as the type of substrate used. The cellulose substrate used contains some regions of free ends and amorphous cellulose that ExG1 is still able to act upon and release cellobiose for BG2 to degrade. However, as the majority of the substrate contains crystalline cellulose, there is a need for endoglucanase. From the cellulases, only the endoglucanase can hydrolyze the internal glycosidic bonds in amorphous cellulose releasing shorter chains for exoglucanase to hydrolyse to cellobiose (Teter et al., 2014). Another possibility may be the generally low amounts of cellulases secreted by these strains. Therefore, it was evident at this stage that further work would be required to improve secretion of the cellulases.

6.1.3 Improved extracellular secretion in *P. pastoris*

Extracellular production of cellulases is necessary for efficient cellulose degradation, therefore their secretion must be optimised. Here, I aimed to compare the use of two different promoter systems in terms of their expression of secreted proteins. The use of two distinct promoter systems was compared by using the constitutive promoter Glyceraldehydes-3-phosphate dehydrogenase (P_{GAP}) and the methanol-induced promoter alcohol oxidase I (P_{AOX1}). The results indicated that promoter P_{AOX1} had almost a 2-fold higher activity compared to the promoter P_{GAP} (Figure 3.12). Similar findings were observed by Várnai et al. (2014) where it was found that the activity and secreted concentration of *Ta*Cel$_5$A (an endoglucanase) to be higher when under the control of P_{AOX1} compared to P_{GAP}. Both the promoters P_{AOX1} and P_{GAP} are regarded to have strong expression via inducible or constitutive regulation, respectively (Türkanoğlu Özçelik et al., 2019).

The *GAP* promoter is a constitutive promoter and the *AOX1* promoter is an inducible promoter activated by the addition of methanol as the carbon source. Methanol is toxic to *P. pastoris* due to the formation of formaldehyde and hydrogen peroxide via methanol oxidisation, resulting in quick degradation of methanol under the alcohol oxidase pathway (Pérez and González, 2017). A study by Adi Santoso et al. (2012) compared the concentration of methanol induction on the growth of *P. pastoris* and found that concentrations over 0.5 % methanol to affect growth negatively. The use of promoter P_{AOX1} to drive the synthesis of recombinant protein was found to result in up to 30 % of the total cellular protein produced in *P. pastoris* being recombinant. Attempts have been made to study the promoter regions responsible for the strength of the promoter, in order to determine how to create stronger variants. One of the regions responsible for P_{AOX1} strength was its 5'UTR, which was identified by Staley et al. (2012). The authors made various deletions to this 5'UTR region, all resulting in decreased expression. The *AOX1* 5'UTR was replaced by the P_{GAP} 5'UTR, resulting in half the activity compared to wildtype P_{AOX1}. This indicated the importance of the 5'UTR for expression in P_{AOX1}. However, the authors did not study the impact of replacing the P_{GAP} 5'UTR with that of the P_{AOX1} 5'UTR to determine whether the

expression of P_{GAP} could be improved, which could be significantly beneficial in a constitutive expression system.

Another group (Yang et al., 2018) studied modifications of the P_{AOX1} poly (dA:dT) tract and found it to have increased expression approximately 3.5fold. A comparison of those modified promoters and the substituted 5'UTRs (P_{AOX1}, P_{GAP}) would be an interesting direction for future work to determine how the expression of the recombinant promoters could be further increased.

Although P_{AOX1} is tightly regulated, it still requires methanol for induction. As methanol is very flammable and toxic towards humans, it results in inflammation of the eyes, visual failure and stomach disturbances. As little as 3g of orally induced methanol may be enough to cause blindness (Moon, 2017b). With larger batch cultures at the industrial level, the safe storage and use of methanol may become a significant hazard, so reduced usage can lead to positive health and safety and economic benefits. The strongly induced expression using $P_{AOX1,}$ and methanol requires that transformed *P. pastoris* strains are initially grown to a high OD_{600} using glycerol or glucose as carbon source and then switched to methanol as the carbon source to increase recombinant protein production Therefore, although P_{AOX1} may be a stronger promoter, P_{GAP} may still be the preferred choice. *P. pastoris* is also known to grow to a higher OD_{600} under glycerol or glucose as a carbon source compared to methanol (Chang et al., 2006; Vanz et al., 2012). As both carbon sources have their own limitations, to overcome this Wu et al. (2003) suggested the combination of both promoters to drive expression. P_{GAP} would express protein production in the initial growth stage and P_{AOX1} would allow for production under methanol induction. Using this combinatorial strategy, the authors reported 2-fold higher expression of recombinant protein. As PaBG1b was found to be expressed under both promoter systems (Figure 3.12), this combination may allow for an increase in its production levels.

The secretion signal is another factor in determining the level of extracellular production of recombinant proteins. Although using secretion signals may result in varying levels of proteins detected extracellularly, the intracellular levels may remain

constant. Therefore, it is important to ensure that as much of the recombinant protein is secreted as possible. To study this, several secretion signal sequences were identified from the literature aiming at the improvement of the concentration of secreted recombinant proteins. G. P. Lin-Cereghino et al. (2013) studied the effects of mutations on the mating alpha-factor secretion signal and found some mutations lead up to a 50 % increase in recombinant protein secretion levels. In this work, these mating alpha-factor mutants from G. P. Lin-Cereghino et al. (2013) were compared using PaBG1b as the reporter protein. It was found that there was no significant difference between the three secretion signals which was contrasting to the authors (G. P. Lin-Cereghino et al., 2013) findings. The mating alpha-factor is known to consist of the pre and pro region. The pre region interacts with the signal recognition particle and the pro region slows down and ensures proper folding of the protein occurs. The authors (G. P. Lin-Cereghino et al., 2013) aimed to reduce the folding time by deleting amino acids from the pro-region that interact with the recombinant protein slowing down secretion. The pro-region consists of two flexible alpha-helices that are required to interact with the recombinant protein. However, other factors may compensate for the observed difference as the PaBG1b was codon optimised whereas G. P. Lin-Cereghino et al. (2013) did not specify if the recombinant proteins in their studies were codon optimised. (Aggarwal and Mishra, 2020) found that some deletions in the pro region of mating alpha-factor resulted in the termination of granulocyte colony-stimulating factor secretion. However, when the authors codon optimised this gene, the same deletions did not restrict its secretion. This may be a reason for the absence of secretion difference observed between the wildtype and mutant mating alpha factors. PaBG1b in this work was codon optimised so it may be translated faster compared to its non-optimised form. Therefore, the mutation in mating alpha-factor to improve secretion by reducing folding time may not be enough to improve the secretion of PaBG1b significantly.

Although these secretion signal mutants did not improve secretion rates of PaBG1b, another secretion signal, Pre-Ost1-Pop, showed positive results. Pre-Ost1-Pop was found to increase the secretion activity of PaBG1b by 1.8-fold (Figure 3.13). The Pre-Ost1-Pop signal is a hybrid secretion signal created by Barrero et al. (2018). The authors combined an Ost1 signal sequence from *S. cerevisiae* into the mating factor

alpha pre-pro-region. The addition of Ost1 promotes cotranslational translocation into the endoplasmic reticulum (ER). Some proteins may fold in the cytosol, preventing transport through the ER. Therefore, using Ost1 to translocate the proteins into the ER prior to folding, may result in a higher rate of protein entry into the ER ready to fold and undergo secretion. This may be a reason for the increased secretion rate seen by PaBG1b (Figure 3.13). This may also be a reason that no increase in secretion is detected when using the mating alpha-factor mutants. If PaBG1b begins folding in the cytosol, it may be unable to cross into the ER therefore the folding process modifications of the mating alpha factors do not result in increased secretion.

Although the hybrid Pre-Ost1-Pro secretion signal was found to increase the activity of PaBG1b (1.8-fold compared to wildtype mating factor), it was not comparable to their reported findings of between 10-20 fold (Barrero et al., 2018). The authors had used E2-crimson protein (RFP) and it was highlighted that a majority of the E2-crimson folds prior to posttranslational translocation and that it tended to aggregate in the ER lumen. Therefore, utilising the Ost1 signal prevented these issues resulting in significantly higher protein secretion. Upon secretion of PaBG1b, it was found to not be aggregating when analysed by SDS-PAGE (Figure 3.8), suggesting that aggregation was not an issue that needed optimisation, therefore, the secretion did not increase as greatly when compared to Barrero et al. (2018).

A combination of the modified (dA:dT) *AOX1* promoter (Yang et al., 2018) as well as the Pre-Ost1-Pro secretion signal (Barrero et al., 2018) may improve secretion further and would be an interesting approach to determine whether a collective effort may result in further increases to secretion. Since secretion is a key element for extracellular production, changes not relating to promoter or secretion signals have also been studied. These changes involved the deletion or addition of genes to aid secretion. The gene *SEC1* was identified in being responsible for the final stage of protein secretion in yeast which involved the secretory vesicle containing the protein to be secreted, fusing with the cell membrane. The gene *SLY1* was also identified to aid in secretion involved in the protein transportation from endoplasmic reticulum to Golgi (Halachmi and Lev, 1996). These two genes were co-overexpressed by C. Li et al. (2017) to aid in secretion and found that a single gene copy overexpression of each

gene led to 64% higher secretion of phytase protein in *P. pastoris*. The authors used a recyclable vector to integrate the overexpression of *SEC1* and *SLY1* and also used this plasmid to integrate multiple gene copies of phytase to improve extracellular expression. This plasmid also allowed the integration of multiple gene copies by being able to recycle the selection marker via cre/*lox*p recombination system.

In summary, *P. pastoris* was identified as the better host for expression and extracellular secretion of cellulases based on its ability to secrete mNeonGreen and exhibit faster growth (OD_{600}). A number of novel cellulases sequences were identified and expressed in *P. pastoris.* The extracellular secretion of these cellulases was improved with the utilisation of a hybrid secretion signal Pre-Ost1-Pop. The next step was to perform further optimisation of the expression of cellulases in *P. pastoris* using a high-throughput integration method that would allow for the integration and combinatorial extracellular optimisation of this metabolic pathway.

6.2 Improved plasmid for high-throughput and multi-gene integration

The cre/*lox*p plasmid from C. Li et al. (2017) was identified to use for multi-gene and multicopy single gene integrations (Figure 4.1) and I aimed to determine its applicability in creating a combinatorial strategy. The plasmid contained the *HIS4* sequence as the site of integration via homologous recombination, offering the advantage of allowing promoter variants to be easily compared and integrated. It also contained a cre-recombinase *lox*p system to excise unnecessary DNA once it was integrated into *P. pastoris*. After creating the plasmid pPGS40 via Gibson assembly, it was analysed by agarose gel electrophoresis. However, multiple bands were detected at around 3kb, 8kb and over 10kb (Figure 4.2). The restriction enzyme AgeI was used as it cut the plasmid in the cre-recombinase gene. Upon linearisation and evaluation, it was found that the plasmid contained a single band at 6kb suggesting recombination between the two *lox* sites had already occurred. This recombination was validated by Sanger sequencing, confirming the presence plasmid pPGS40r.

The issue of recombination occurring in *E. coli* prior to integration in *P. pastoris* was also reported by Pan et al. (2011). The authors suggested that it could be due to

baseline (uninduced) expression of promoter P_{AOX1} in *E. coli*. However, it was not an issue for downstream processes for the authors. The authors analysed the bands by agarose gel electrophoresis and extracted the non-recombined plasmid. However, for a high-throughput approach, this would not be a feasible solution. Therefore, a change in the promoter P_{AOX1} was carried out to try and reduce baseline expression and the promoter selected to replace it was P_{FLD1} (formaldehyde dehydrogenase). The advantage that this promoter offered was that it was also induced by methanol (required for recombinase excision in *P. pastoris*) as well as being 332bp shorter in size compared to P_{AOX1} whilst having similar expression strength (Shen et al., 1998). The substitution of P_{AOX1} by P_{FLD1} did not stop recombination occurring (Figure 4.3) suggesting that either the P_{FLD1} may also exhibit leaked expression in *E. coli* or that other elements resulted in the translation of cre-recombinase. However, P_{FLD1} is not as commonly used or researched as the P_{AOX1}, possibly due to the patent on the promoter which expired in July 2019 (Cregg, 2004), so it was not possible to suggest whether recombination was also occurring due to leaked expression similar to that observed for P_{AOX1}.

Another strategy was to implement a shorter incubation time to reduce the amount of the cre-recombinase protein produced in the cells, reducing the possibility of recombination occurring. However, that did not reduce the recombination (Figure 4.3). Research by D. Li et al. (2017) also highlighted the undesired *lox* recombination observed in *E. coli*. The authors implemented the use of introns in between the promoter and cre-recombinase gene region as *E. coli* cannot process introns in mRNA (Agaphonov and Alexandrov, 2014). Specifically, the authors inserted the lac operon operator lacO which was bound by LacI in *E. coli* only, preventing the downstream expression from occurring. This solution was implemented into pPGS40 with the addition of the lacO sequence in between the P_{FLD1} and cre-recombinase gene (Figure 4.3). However, multiple bands were still detected and so recombination was still occurring. The *E. coli* strain used in this work was TOP10 which does not contain a lac repressor necessary for binding to the lacO. To address this, the *E. coli* TOP10F' strain was tested, but it still resulted in recombination occurring (data not shown). It may be possible that as this was a high copy plasmid, so even slight expression of cre-recombinase was enough to begin recombining the plasmids.

A combination of all the above methods was implemented to try and achieve a cumulative effect, however, it also did not stop the recombination completely (Figure 4.3). An alternative strategy was utilised whereby an *E. coli* terminator was placed upstream of the cre-recombinase gene and downstream of P_{FLD1}. A library of synthetic short *E. coli* terminators had been designed and characterized by Chen et al. (2013) among which was the terminator L3S2P21. This was reported as the strongest synthetic terminator whilst remaining short in size. Upon its addition and analysis on electrophoresis gel (Figure 4.4) only two bands were detected for this plasmid (pPGS40 +L3S2P21) at around 8kb and over 10kb in comparison to pPGS40 which displayed multiple bands including at 3kb. An additional fragment of flavin-based fluorescent protein, iLOV, was added whilst swapping the P_{AOX1} for P_{FLD1} to reduce the size of the plasmid creating the final desired plasmid (pPGS60, Figure 4.6) which did not recombine in *E. coli*. Upon linearization using the restriction enzyme AgeI, the size was confirmed to be around 10 kbp and the plasmid was verified via Sanger sequencing. This allowed the purification of the plasmid pPGS60 without requiring electrophoresis gel purification and therefore a high-throughput approach.

A plasmid previously shown to be used for metabolic pathway integration was chosen for use in a high-throughput integration approach, however, unwanted recombination in *E. coli* prevented it from being used as a high-throughput system. By employing a number of systematic approaches, I was able to identify modifications to the plasmid that prevented unwanted recombination in *E. coli* as well as creating a faster and more efficient method for selecting positive transformed colonies in *P. pastoris*. The resulting system enabled high throughput metabolic pathway integration. The resulting system paved the way to utilise a combinatorial strategy for a high-throughput approach. This provided variation in expression of genes within a metabolic pathway allowing the selection of an optimal expression system.

6.3 High-throughput combinatorial strategy

To create a high-throughput approach for cloning genes into pPGS60, iLOV was added to the final plasmid (pPGS60, Figure 4.5) allowing a faster, more efficient method of identifying colonies that had been transformed. Utilising the mechanism of this recombination system, iLOV would fluoresce when exposed to blue light λ_{ex} (400 - 525 nm). The *P. pastoris* competent cells were transformed using the plasmid

pPGS60. When the colonies were exposed to blue light, the iLOV emission was observed using orange filter glasses. As the iLOV expression was under the control of constitutive promoter P_{GAP}, it would continually express the fluorescent protein (Figure 4.7). The pPGS60-containing colonies were grown for 72 hours using either glucose or methanol as a carbon source. The colonies containing glucose as a carbon source expressed iLOV, resulting in the observed fluorescence units after 24 hours of growth (~38,000 RFU / mL) compared to using methanol as a carbon source where an increase in iLOV fluorescence was not detected (~15,000 RFU / mL) (Figure 4.8). This application of fluorescence has been reported to be a useful method to select positive transformed colonies by others (Kondo & Yumura, 2019; Wong & Truong, 2010). These authors utilised fluorescent proteins (FP) either as free FPs or FPs bound to the protein of interest, as a way of selecting positive transformants, highlighting the reduction in labour and costs when using this approach, while improving efficiency. However, the fluorescent protein would continue to be expressed in those systems, which may impact cell growth or production of the protein of interest by diverting/sharing cellular resources towards the FP. Here, iLOV would be excised through the recombination system after positive colony selection, and thus not affect the colonies in downstream applications.

The addition of iLOV improved positive transformant colony selection compared to the original method published by C. Li et al. (2017) (Table 4.1). Another improvement in positive colony transformation was observed compared to the method from C. Li et al. (2017) after methanol induction streaking (Table 4.1). Upon cultivation of positive colonies in methanol and re-streaking on YPDA plates, the colonies were selected based on the lack of fluorescence. As a result of methanol as a carbon source inducing the expression of P_{FLD1}, the expression of *cre*-recombinase was activated. This would result in the recombination of the two *lox*p sites in *P. pastoris*, excising unnecessary DNA, and resulting in the loss of the iLOV cassette and therefore fluorescence from iLOV. The colonies that had undergone possible successful recombination would not fluoresce and were selected for the next step (Figure 4.6). Here, the colonies were spotted on plates with or without the antibiotic selection and as iLOV fluorescence was not utilised, no noticeable difference between the two methods was observed (Table 4.1).

The combination of these events decreases the method costs (fewer cPCRs) and time (positive colonies could be selected early via iLOV fluorescence) (Table 4.2). The final stage of the method used by C. Li et al. (2017) and originally published by J. Lin-Cereghino et al. (2005) required cells to be grown overnight and then refreshed the next day to OD_{600} 0.6 before making them competent. However, a study by Kumar, (2019) highlighted the use of scraping *P. pastoris* cells from plates where *P. pastoris* had been spotted and directly using them to make competent cells for subsequent transformation. This method from Kumar, (2019) was incorporated into this method (Figure 4.6) allowing competent cell preparation from the spotted plates shortening the experimental procedure by 2 days compared to the method by C. Li et al. (2017) (Table 4.2).

Overall, the cumulative changes reduced the method from a 10-day process to a 7-day process. This is extremely useful for metabolic pathway integration for example, when increasing integration of the gene of interest to 4 copy number or integration of a metabolic pathway containing 4 genes, it will take up to four rounds of integration and so 40 days from the method from C. Li et al. (2017). However, it requires only up to 28 days in the modified method presented here. This modified method also offers the advantage of the ability for high-throughput and automation as colonies could be directly selected for assay using their fluorescence via an automated colony picker.

The modified method was tested by using fluorescent proteins, successfully integrating RFP, iLOV and YFP (Figure 4.9). This was also highlighted by C. Li et al. (2017) whereby the integration of up to six copies of a single phytase gene was carried out. However, either of these methods would not be applicable for integrating a full metabolic pathway. For example, an increase in the number of variables affecting the expression of genes, would result in an increase in the number of strains created. Subsequently, transformations of those strains required to achieve all combinations of the expression elements would also increase. For example, three different promoters controlling three genes would require a total of 27 individual strains (3^3) to be made. To overcome this issue, a second plasmid was designed. It was based on pPGS60 but

differed at the homologous site of integration sequence and the antibiotic selection marker.

The plasmid pPGS80 was created (Figure 4.10) featuring *URA3* as the site of integration and G418 as the antibiotic selection marker. Nett & Gerngross, (2003) had previously reported the successful use of *URA3* as an integration marker in *P. pastoris*. However, the authors used *URA3* as the auxotrophic selection marker, resulting in slower growth compared to antibiotic usage. Here, as it was planned to use *URA3* only for site of integration, it would not affect the growth of the transformants as much as it would if it was used for an auxotrophic marker (J. Lin-Cereghino et al., 2005;Nett & Gerngross, 2003). Another advantage of using *URA3* (pPGS80) was that it was located on chromosome III and *HIS4* (pPGS60) was located at chromosome I, significantly separating the site of integration. The final plasmid was tested for its efficiency and it was found to have similar efficiency in the use of this modified method as pPGS60 (Table 4.3).

The main aim of this secondary plasmid (pPGS80) was to use it in combination with pPGS60, through co-transformation, which would allow 2 genes to be integrated in a single reaction. This would significantly reduce the time taken to integrate multiple genes and also allow the integration of metabolic pathways and their optimisation. Therefore, the positive transformant efficiency (percentage of positive transformants after transformation) of both plasmids was studied in this modified method. Using a combination of the plasmids pPGS60 and pPGS80 allowed a similar success in identifying positive transformants (Table 4.3). This co-transformation was able to produce a strain containing four genes in fourteen days. This was significantly faster compared to the method by C. Li et al. (2017) which estimate forty days. Both plasmids contained cre-recombinase induced by methanol. However, an alternative approach may be to remove the cre-recombinase cassette from the second plasmid, allowing a shorter plasmid containing only the *lox*p sites. The mode of action of cre-recombinase from the first plasmid will ensure the second plasmid also undergoes *lox* recombination. Another option would be to integrate cre-recombinase induced by methanol into a wildtype *P. pastoris* strain and then subsequent integrations of

plasmids containing only *lox*p sites as a way to prevent cre-recombinase activity completely in *E. coli* as well as reducing the plasmid size significantly.

The aim of creating a combinatorial strategy to optimise metabolic pathways could therefore begin with the utilisation of this double plasmid co-transformation using pPGS60 and pPGS80. A method was hypothesised that would create a complex combinatorial assembly from a single reaction (Figure 4.13). Initially, a bi-directional promoter system was hypothesised allowing forward and reverse orientation genes to be inserted into a single plasmid. A study by Vogl et al. (2018) resulted in the creation of a library of synthetic bi-directional promoters that could be used as they had a varied range of expression. However, these bi-directional promoters were based on methanol induced promoters so were all required to be induced by methanol which may be an unfavoured in biotechnology industrial science (Moon, 2017b).

Constitutive bi-directional promoters were sought. However, even after extensive literature research, no suitable candidates were found. Another strategy was to utilise the 5'UTRs and place them downstream of the constitutive promoters. A study by Liang et al. (2012) had annotated *P. pastoris* transcript structures, identifying a large amount of 5'UTRs. However, their expression has not yet been characterised and therefore it would require a thorough experimental evaluation to acquire a list of 5'UTRs with varying rates of expression. A study by Qin et al. (2011) created and identified P_{GAP} mutants that not only contained a range in expression (0.6 – 20-fold) but also had higher expression than the wildtype P_{GAP} (up to ~11-fold higher) in protein concentrations. These promoters had the benefit of being constitutively expressed under glucose or glycerol and therefore were selected for the combinatorial strategy. Based on Qin et al. (2011), the promoters selected for this thesis were the wildtype P_{GAP}, the P_{G1} variant that provided the highest activity and the P_{G4} variant, with intermediate expression, between wildtype P_{GAP} and P_{G1}. However, assembling and PCR amplification as bidirectional promoters (Figure 4.12) was not possible (data not shown).

This led to the decision to create all the gene cassettes in the same (forward) orientation, using three of these promoters to vary the range of expression of genes.

The method was developed using a combination of strategies from various sources. In Figure 4.13, the three promoters contain the same overlapping regions compatible with the target plasmid, making them suitable for Gibson assembly cloning, and therefore providing a theoretical 1 in 3 possibilities of either one of the promoters to assemble randomly for the first gene and the same for the second gene via sequence homology. This in theory creates a library of fully intact plasmids with randomly assembled promoters upstream of the gene cassettes. Furthermore, unless all the promoters and genes assembled correctly, a functional plasmid would not be obtained thus providing higher accuracy for positive transformant selection. This modular construction was also reported by Casini et al. (2014) whereby Gibson assembly was used for the construction of multiple fragments. The authors also created software that identifies appropriate overlapping regions. This strategy was also used by Engler et al. (2008), where Golden Gate assembly was utilised. However, this presents the dilemma of adding unwanted restriction sites between the cassettes and plasmid regions. Here, as sequence homology is used, the original sequence can remain unmodified and assembled without additional DNA base pairs or restriction sites added.

The use of promoter combinations and assembly via cloning methods was also highlighted by Prielhofer et al. (2017) and Reider Apel et al. (2017) however it required the usage of CRISPR/Cas9 had to be used for integration. The method developed here also offers the advantage of fewer steps required and a single tube reaction. Another advantage was the addition of iLOV which allowed cPCR free selection of positive transformants as well as an automation-friendly approach. This method could then be utilised for the construction of metabolic pathways and their optimisation.

This worked resulted in the creation of two plasmids that enabled the integration of up to 4 gene clusters in *P. pastoris* in 7 days. This is a highly advantageous approach in the biotech industry, reducing time and manual labour to improve recombinant protein production. Furthermore, the expression of these gene clusters could be varied, to allow for identification of optimal expression of a metabolic pathway. This approach could be adapted for the optimisation of the metabolic pathways, such as those involved in cellulose degradation and ferulic acid degradation.

6.4 Ferulic Acid degradation

The ferulic acid degradation pathway was successfully integrated into *P. pastoris* creating the strain *P.p.* –CgCar. The pathway consisted of 3 genes: (1) CE1 (Feruoyl esterase) was bound to secretion signal Pre-Ost1-Pro to release ferulic acid from wheat arabinose extracellularly, (2) HisCar5 (carboxylic acid reductase) to convert the ferulic acid (substrate) to coniferyl aldehyde (intermediate product), and (3) CgAkr-1 (acid-keto reductase) to convert coniferyl aldehyde to coniferyl alcohol (targeted end product). This pathway has previously been reported to be expressed in *E. coli* by Tramontina et al. (2020), producing the desired coniferyl alcohol product. Expression of this pathway in *P. pastoris* would be expected to offer the advantage of higher protein production and the ability to secrete proteins at higher concentrations compared to *E. coli* (Demain and Vaishnav, 2009).

The *P.p.* –CgCar strain contained all three enzymes regulated only by the constitutive promoter *GAP*. The strain was used to determine the conversion of ferulic acid to coniferyl alcohol before metabolic pathway optimisation could occur. It was found that the strain was able to release ferulic acid from wheat arabinose successfully (Figure 5.6) via LC-MS analysis. The extracellular expression of feruloyl esterase in *P. pastoris* has also been reported by others (Juge et al., 2001; Zeng et al., 2014; Dotsenko et al., 2016). The release of ferulic acid from wheat arabinose was also observed by Tramontina et al. (2020). However, the authors carried out the heterologous expression of CE1 enzyme in a separate *E. coli* strain, purified it and then added it to the media where the *E. coli* strain containing the HisCar5 and Cg-AKR-1 was expressed. This creates a complex process whereby CE1 must be produced in large quantities and stored in appropriate buffers before usage. In the strain *P.p.* –CgCar, the feruloyl esterase would be secreted continuously for the length of the batch process allowing efficient continuous degradation of wheat arabinose to release ferulic acid as a substrate for the required conversion to coniferyl aldehyde and coniferyl alcohol.

Although ferulic acid was detected from *P.p.* –CgCar, there was no detection of coniferyl aldehyde or coniferyl alcohol even when purified 5 mM ferulic acid was

added to the culture. Tramontina et al. (2020) had added 5 mM of ferulic acid to the cultures in their study and obtained successful conversion without observable toxic effects. However, the authors expressed the pathway in bacteria whereas in this work it was expressed in yeast. Adeboye et al. (2014) reported that phenolic compounds were toxic to the growth of the yeast *S. cerevisiae.* The authors observed that 1.8mM of ferulic acid is the lower limit of toxicity where growth began to be affected negatively. Similar findings were observed by Zha et al. (2013) who reported that 0.2 g/L of ferulic acid (approximately 1.03mM) was enough to affect the growth of another yeast species (*Pichia anomala*). However, when the conversion of low ferulic acid concentration (1 mM) was investigated in this work, it was found that ferulic acid was still not converted to coniferyl aldehyde or coniferyl alcohol (Figure 5.5). The inactivation of the carboxylic acid reductase (CAR) in yeast may be a possible reason for this lack of observed conversion. Venkitasubramanian et al. (2007) identified that post-translational phosphor-pantetheinylation of CAR is required for maximal enzyme activity by phosphopantetheinyl transferase (PPTase) which attach the 4-phosphopantetheine moiety of a coenzyme A to an inactive apo-protein, converting it to a fully active holoprotein. The authors found that recombinant expression of CAR from *Nocardia sp.* to be minimal without the addition of the PPTase and acetyl-CoA. The HisCar5 used in this work may be in its inactive form. The CAR used in this study (HisCar5) originates from *Nocardia iowensis* suggesting that it may likely be an apoprotein requiring conversion to a holoprotein via PPTase to become fully active (He et al., 2004). Tramontina et al. (2020) observed activity from this CAR (HisCar5) however, it was expressed in *E. coli.* There are three PPTases present in *E. coli,* of which one is an acyl PPTase that can activate this CAR (Mootz et al., 2001).

Hansen et al. (2009) reported that *S. cerevisiae* was not able to convert carboxylic acids using CAR as a result of the lack of PPTase. Horvat et al. (2020) recognised this and identified the need for a PPTase for CAR activity and constructed a plasmid expression system combining both the CAR and PPTase in *P. pastoris*. The authors utilised bi-directional promoters to vary the range of expression of both enzymes to find an optimal flux for maximum product yield of pipernol from pipernoylic acid. These results indicate the need for the presence of a PPTase in the *P. pastoris* strains

developed in this work to allow conversion of ferulic acid to coniferyl alcohol via the activated HisCar5.

Phenolic compounds of varied molecular weights hinder the growth of yeasts in several ways, from reducing the activity and expression of ion and sugar transporters in the cell membrane (high molecular weight phenolic compounds) to being able to penetrate the cell wall and reported to be more toxic towards microorganisms (low molecular weight phenolic compounds) (Klinke et al., 2004). Sierra-Alvarez & Lettinga, (2007) found that compounds with aldehyde groups or apolar substituents were more toxic to microorganisms from granular sludge compared to compounds with carboxylic groups when observing minimum inhibitory concentrations. The study demonstrated that the compound cinnamaldehyde (belonging to the same group of cinnamaldehydes as coniferyl aldehyde [Kim et al., 2021]) was found to be inhibitory at 100-fold lower concentration than ferulic acid. Cinnamaldehydes are known to have antifungal properties and are toxic towards various microorganisms although their mechanism of inhibition is not well understood (Taherzadeh and Karimi, 2011). These molecules act by inhibiting plasma membrane ATPase, lowering intracellular pH and depletion of NADPH leading to membrane and cell wall damage (Shreaz et al., 2011). Therefore, such inhibitory activity might be a reason for the absence of conversion from the carboxylic acid group-containing compound ferulic acid, to the aldehyde group-containing compound in coniferyl aldehyde and the subsequent conversion to coniferyl alcohol.

Shreaz et al. (2011) found the minimum inhibitory concentration of coniferyl aldehyde to be between 100 and 300 ug/mL (approximately between 0.6 and 1.7 mM) in the yeast *Candida albicans*. The authors used electron microscopy to observe the toxicity effects in yeast's morphology, highlighted by the wrinkling and breakage of the cell wall and subsequent leakage of cellular content indicating the necessity for the yeast to quickly convert any coniferyl aldehyde to a less toxic compound. Electron microscopy (scanning and transmission) and minimum inhibitory concentration spot assays could be utilised in future experiments to determine whether coniferyl aldehyde and/or coniferyl alcohol displayed a similar toxic effect in *P. pastoris* as that observed by (Shreaz et al., 2011). However, coniferyl aldehyde/alcohol added to the samples in

this work were below the toxic limits reported by others (Adeboye et al., 2015), therefore it is likely that both were converted to another compound (Chapter 5.1.6).

Although the metabolism of both coniferyl aldehyde and coniferyl alcohol is not well defined in *P. pastoris,* it is well documented in the yeast *S. cerevisiae,* which may provide an insight into *P. pastoris*' metabolism of these compounds. Adeboye et al. (2015) investigated the metabolism of 1.1 mM coniferyl aldehyde over time in *S. cerevisiae*. The study found that coniferyl aldehyde was rapidly converted to ferulic acid and ferulic acid isomer within 2 hours from its addition. This is similar to this work's findings where two peaks were identified within the first 2 hours after the addition of coniferyl aldehyde (Figure S5.6), one corresponding to coniferyl aldehyde and the other to a retention time closer to that of ferulic acid. The authors also reported the complete conversion of coniferyl aldehyde after 24 hours of incubation (having been converted to a range of different reported compounds). This was also observed in the present study, where two separate peaks (Figure S5.6) were identified up to the 3-hour time point indicating different compounds. As well as that a single peak was observed after 24 hours at a different retention time compared to the original coniferyl aldehyde standards peak. This may suggest that coniferyl aldehyde underwent a conversion process over time into another, less toxic, compound. LC-MS analysis of this works' samples could assist in the identification of what compounds are present and what they have been converted to.

TLC analysis was carried out to determine whether the coniferyl aldehyde or coniferyl alcohol had been converted to ferulic acid, ferulic acid isomer, dihydroferulic acid or a carboxylic group-containing compound (Adeboye et al., 2015). Samples from the various timepoints as well as appropriate controls were spotted on TLC paper and stained with ferric chloride (Iron(III) chloride) where Iron(III) is reduced to Iron(II) when reacting with an acid, resulting in an orange/brown colouring (Hynes and O'Coinceanainn, 2004). The identification of ferulic acid using this reaction was also reported by other groups (Sharma et al., 1998; Mabinya et al., 2002, 2006), who observed an orange/brown colouring in its presence. The TLC analysis of the compounds in this work (Figure 5.13) confirmed the possible presence of an acid (possibly ferulic acid or ferulic acid derivative) among the samples due to the orange-

brown colouring (as well as the peaks identified via HPLC (Figure 5.13). The intensity of the spots increased as the time point increased which were consistent with the peak shift over time in the samples containing coniferyl aldehyde and coniferyl alcohol as starting compounds. The colouration intensity remained constant in the samples containing ferulic acid as the starting compound, which was in accordance with the HPLC data where peaks remained constant throughout (Figure 5.7) the chromatographic runs.

The process of conversion of coniferyl aldehyde/alcohol remains unknown in *P. pastoris* and is better understood in *S. cerevisiae*. Adeboye et al. (2015) hypothesised that coniferyl aldehyde may be oxidised to ferulic acid via oxidoreductases. The study identified the aldehyde dehydrogenase enzyme (*ALD5*) as being involved in the metabolism of coniferyl aldehyde in *S. cerevisiae*. Upon the knockout of *ALD5*, it was found that *S. cerevisiae* growth was affected negatively (Adeboye et al., 2017) in the presence of coniferyl aldehyde. The authors did not investigate the effect of knockout of this gene in strains created for specific coniferyl aldehyde conversion. This may be important in reducing the rate of coniferyl aldehyde to an unwanted compound. The gene *ALD5* has also been reported to be present in *P. pastoris* (Paes et al. 2021), however its role in coniferyl aldehyde metabolism has not yet been studied. A UniProt (https://www.uniprot.org/) search also revealed the presence of other aldehyde dehydrogenases present in *P. pastoris*, however, they were not annotated so their role in coniferyl aldehyde metabolism remains unclear. As highlighted by Adeboye et al. (2015 and 2017), there are a number of genes identified that are involved in the metabolism of coniferyl aldehyde (*ALD5, PAD1, ATF1* and *ATF2*) as well as others that were not yet identified. The identification of genes involved in the metabolism of coniferyl aldehyde and coniferyl alcohol similar to that described by Adeboye et al. (2017) in *S. cerevisiae*, may enable the creation of specific gene knockout strains. These strains could reduce the catabolism of coniferyl aldehyde/alcohol and so can be expected to allow for targeted production of desired compounds in *P. pastoris*. This, together with the addition of PPTase, may allow the successful conversion of ferulic acid to coniferyl alcohol. Unfortunately, due to the pandemic, less time than expected was available for me to pursue further investigation in this area.

The high-throughput method developed in this work was utilised to optimise synthesis of the enzymes responsible for the ferulic acid pathway. The high-value compounds produced via this pathway are highly desirable and valuable to the biotech industry. Despite the successful integration and expression of this metabolic pathway, the end-product (coniferyl alcohol) was not detected in samples. The absence of the expected target compounds was attributed to the presence of additional enzymes that catalyse alternative reactions. The method devised in this work paves the way for possible optimisation via specific gene knockouts and/or gene overexpression, as well as the optimisation of other pathways such as cellulose degradation that are of high-value and of interest to the biotech industry.

6.5 Optimisation of Cellulase degradation pathway

The degradation of cellulose is dependent on three cellulases; endoglucanase, exoglucanase and ß-glucosidase. Here, an optimal strain was created capable of releasing glucose from cellulose and utilising it for growth. The *P. pastoris* strain transformed with the cellulases endoglucanase (TrEGI), ß-glucosidase (PaBG1b) and exo-glucanase (ExG2) was created by utilisation of the method devised in Chapter 4. This method allowed the combination of one exoglucanase, one endoglucanase and one ß-glucosidase (Table 3.2) to be secreted extracellularly into the cellulose-containing media. This allowed the creation of twenty-seven possible combinations. The plasmids containing this range of cellulase and their expressions were transformed and integrated into *P. pastoris*, resulting in the creation of hundreds of colonies with an equal possibility of containing any of those combinations. These colonies were randomly selected to determine which may contain the optimal cellulose degradation pathway, highlighted by the range of growth rate (OD_{600}) observed. This results from the combinations of the different types of cellulase as well as the promoters regulating the expression. The combinations differ in the expression and thus the degradation of the cellulose substrate into glucose in the cultures. As *P. pastoris* utilises glucose as a carbon source, a strain capable of releasing the optimum amount of glucose will show the fastest growth. Gusakov et al. (2007) also reported the combination of cellulases to improve the degradation of various target substrates and also found that different combinations were necessary for different substrates. However, Gusakov et al. (2007)

carried out studies using different ratios of purified proteins starting from specific substrates ideal for those enzymes. Subsequently, those ratios were recreated *in vivo*.

Synergism is essential for the efficient degradation of cellulose and so lignocellulosic biomass. Endoglucanase is responsible for the degradation of internal β(1-4) glycosidic bonds releasing shorter chains that exoglucanases can cleave to cellobiose units. As cellulases are known to suffer from end-product inhibition, synergism must allow the degradation at a rate that prevents this from occurring (Lakhundi et al., 2015). From the higher OD_{600} strains (Figure 5.12), all three promoters (P_{GAP}, P_{G1} and P_{G4}) were identified regulating the expression of PaBG1b. The ß-glucosidase that was predominantly present in all faster-growing strains (PaBG1b) manifested a very high resistance towards cellobiose and glucose compared to many other ß-glucosidases that have been characterised (Y. Li et al., 2017). PaBG1b was able to retain activity at higher concentrations of cellobiose and glucose in the cell culture media. Although PaBG1b has been reported to produce maximal activity at 45 ^{o}C, it retains more than 50 % activity at 30 ^{o}C. It also retains more than 80 % activity between pH 5 and 6 (Y. Li et al., 2017). The cell culture conditions in this work were set at 30 ^{o}C and pH 6, still allowing for a relatively high activity from PaBG1b.

For the endoglucanase (TrEGI) the higher growth rate (OD_{600}) strains contained the promoter P_{G1}. The same promoter, P_{G1}, was also identified in the faster growth rate (OD_{600}) exoglucanase (ExG1) strains. However, the strains that grew slower compared to wildtype contained the promoter P_{GAP} regulating the expression of TrEGI and ExG1. The P_{GAP} promoter has the lowest transcription rate of the three promoters used, as reported by (Qin et al., 2011) resulting in reduced expression and thus the production of these cellulases. This reduces the degradation of the amorphous cellulose in the extracellular medium resulting in a lower concentration of glucose and therefore lower growth. As the promoter P_{G1} is the strongest, its expression may result in larger amounts of protein secretion. Secretion may be the rate-limiting factor for both TrEGI and ExG1. Nakazawa et al. (2008) characterized the catalytic domain of TrEGI and found it to manifest maximal productivity at 15 ^{o}C and pH 5. The cells were cultured at 30 ^{o}C and pH 5.5. This difference between the optimal activity conditions of TrEGI and the optimal conditions of the cell cultures, likely resulted in

decreased activity for the TrEGI, making it one of the rate-limiting factors. Thus, an overexpression of this cellulase may be beneficial for removing this limiting factor and increasing the cellulose degradation rate.

ExG1 has not yet been characterised (https://www.uniprot.org/) however it may have a different maximal activity parameter which may influence the activity and also make it one of the rate-limiting factors. As the cell culture conditions (pH 6, 30 ^{o}C) were more suited to PaBG1b optimal activity conditions (more than 50% activity at pH 6, 30 ^{o}C), an increase in expression was not necessary for this assay, in contrast to TrEGI and ExG1. Therefore, optimisation of the expression of TrEGI and ExG2 may be more beneficial in increasing cellulose degradation rate.

The importance of this pathway optimisation led to the identification of the need for increased secretion for two (TrEGI and ExG1) of the three enzymes. This improved strain can be further optimised by the addition of more copies of the genes of TrEGI and ExG1 using the method devised in this work. That would determine how many additional copies are required as well as what expression level (promoter combination) would be necessary. C. Li et al. (2017) successfully integrated up to 16 gene copies into *P. pastoris*. However, the authors suggested that the integration of 6 to 8 gene copies is the upper limit of integration in *P. pastoris*, as they found no significant advantage after genomic integration of 6 to 8 gene copies. Therefore, it is essential to optimise the metabolic pathway when increasing the copy number of genes in *P. pastoris*.

The advantage of using the method devised here is to allow an efficient and labour reduced approach to finding the optimal parameters for metabolic pathway optimisation. The fluorescent cassette in the plasmid allows it to be easily detected manually or via robotics as a positive transformant (with high success) without the need for cPCR. Furthermore, as cellulases function in synergism, the addition of the same type of exoglucanases or endoglucanases in a strain may improve the degradation of substrates. Igarashi et al. (2011) reported the addition of two different exoglucanases resulting in improving the hydrolysis rate of cellulose. However, this requires purification and cell culture media buffer ideal to that of the enzyme itself,

which may not be feasible. Using the method here, high-throughput screening of different types of cellulases and their different levels of expressions can be determined rapidly and efficiently. This allows the detection of novel or pre-characterised cellulase/enzymes which will be optimal for the conditions in which *P. pastoris* is cultured.

6.6 Future Perspectives

The results accomplished in this work provide a new insight into the use of industrial friendly plasmids for multi-gene metabolic pathway integration and optimisation. Using the combinatorial strategy developed here, high-throughput analysis of metabolic pathway optimisation can be carried out either at the bench or via robotics. The use of 5'UTR modifications in tuning gene expression in *P. pastoris* is not fully understood. Using this method, the study of various *P. pastoris* 5'UTRs is feasible and can allow for the identification of a range of 5'UTRs for differing expression rates from those identified by Liang et al. (2012). Another factor that remains a challenge is to increase the secretion ability of *P. pastoris*. Although multiple genes have been identified and expressed individually to improve secretion, their combined effects are yet to be investigated. A study in the combination of these gene pathways to increase secretion may result in creating a strain with excellent (or maximal) secretion ability.

The method devised here allows rapid investigation of the combinatorial utilisation of gene expression elements to optimise enzyme pathway functionality. There is still a challenge in achieving efficient degradation of lignocellulosic biomass. Pre-treatment of lignocellulosic biomass via physical or chemical disruption is a common approach to deal with toxic compounds or to release desired substrates (Baruah et al., 2018b). Exploiting the ability of *P. pastoris* to allow integration of multiple copies of genes (C. Li et al., 2017) may be a good approach to create strains capable of degradation of multiple substrates. Combining this approach with the method devised in this work, a *P. pastoris* strain that contains multiple optimal pathways is feasible. One of the pathways would be used for degrading lignocellulosic materials to release glucose as a carbon source for *P. pastoris* as well as the associated phenolic compounds. The second pathway would be used to convert the phenolic compounds to more desirable

targets such as coniferyl alcohol. This could be an efficient way to overcome both challenges via one strategy.

6.7 Conclusions

The work in this study aimed to create a high throughput combinatorial strategy to optimise metabolic pathways in yeast.

I identified two possible hosts for metabolic pathway integration (*S. cerevisiae* and *P. pastoris*) and selected *P. pastoris* based on its observed ability to secrete proteins. *P. pastoris* was then used to express novel cellulases provided by Ingenza Ltd. to obtain extracellular expression. Cellulases that were secreted could be utilised for downstream metabolic pathway optimisation. Various secretion signals were compared to obtain the signals that would result in the highest secretion of recombinant proteins.

A cre/*lox* system was identified for metabolic pathway integration from C. Li et al. (2017). Issues with unwanted *lox* site recombination were detected, however, these were rectified with modifications to the plasmid utilising different promoters and additional terminators. Combining these approaches, a method was created to allow a high-throughput one-tube reaction for random assembly of promoters and genes using Gibson assembly. The plasmid utilised iLOV recycling by the cre/*lox* system to identify positive transformants in *P. pastoris*. Assays using fluorescent proteins confirmed the ability for the plasmid to carry out high-throughput randomised assembly of components and positive selection using iLOV. The modified method allowed the integration of a metabolic pathway three times faster than previous methods.

The method devised here was used to optimise metabolic pathways. The cellulose degradation pathway was optimised using the cellulases that expressed extracellularly as highlighted earlier. Colonies were cultured on cellulose-containing media and the colonies containing optimised cellulose degradation metabolic pathway resulted in the optimal concentration of glucose and therefore growth of cultures. The pathway was

found to require a stronger promoter for cellulases endoglucanase and exoglucanase expression, whereas the promoter variation in ß-glucosidase did not affect growth.

Ferulic acid degradation was to be optimised using this pathway, however, a lack of detection of the product coniferyl alcohol through LC-MS necessitated deeper investigation. It was found that the intermediate product coniferyl alcohol and coniferyl aldehyde were readily converted to a derivative within the first few hours of incubation. Literature research revealed possible compounds that it may be converted to and gene knockouts were postulated that may result in the production of desired compounds in coniferyl alcohol. Further progress on resolving the challenges of this part of the project became a victim of the pandemic.

Appendix

Primers and gBlocks

Table S2.1: Primers and gBlock sequences for plasmids (sequences from 5' – 3')

Primer name	**Sequence**	**Use**
pGS1		
GS_7-F	GGGGACGAGGCAAGCTAAACAGATCTGGCGCGCCTCATTATCA ATACTGCCATTTCAAAGAATACGTAAATAATTAATAGTAG	*TDH3* to pnm1-tt Backbone
GS_8-R	AAACTGCAGTAAAAATTGAAGGAAATCTCATTTAATTAATTTGT TTGTTTATGTGTGTTTATTCGAAACTAAGTTC	*TDH3* to mating alpha factor
GS_9-F	AGTTTCGAATAAACACACATAAACAAACAAATTAATTAAATGA GATTTCCTTCAATTTTTACTGCAGTTTTATTCG	mating alpha factor forward
GS_10-R	TCTTCACCCTTAGAAACCATACCGGTAGCTTCAGCCTCTCTTTT ATCCAAAGATAC	mating alpha factor Reverse ymneongreen
GS_11-F	ATAAAAGAGAGGCTGAAGCTACCGGTATGGTTTCTAAGGGTGA AGAAGACAACATGGCTTCTTTGCCAGC	YmNeongreen Forward
GS_12-R	TCGCTCTTATTGACCACACCTCTACCGGGATATCTTAATGGTGG TGGTGATGGTGGTGGTGGTGATGCTTGTACAATTCGTCCATACC	YMNeongreen reverse to *ADH1*
pGS2		
gBlock		
GS_UAS-c-F	CACAAAACTCGAGAACATATTAAATGAGATTTCCTTCAATTTTT ACTGCAGTTTTATTCG	UAS-c-to mating alpha factor
GS_16-R	AAAATTGAAGGAAATCTCATTTAATATGTTCTCGAGTTTTGTGT TCTTAAGTAGATAATTGG	mating alpha factor to UAS-c
GBLOCK-UASC	GGACGAGGCAAGCTAAACAGATCTGGGGCGCGCCCCTCCTTGAAACTGAAATTTTAGCATGTGATTAATT AACTTGTAATATTCTAATCAAGCTAGCACTGTTGGGCGTGAGTGGAGGCGCCGGAAAAAAAGCATCGAAAA AAGGATCCAATTATCTACTTAAGAACACAAAACTCGAGAACATATTAAATGAGATTTCCTTCAATTTTTAC TGCAGTTTTATTCGCAGC	
pGS4		
GS_19-F	GCAAGCTAAACAGATCTGGCGCGCCAACAATTTTTCTTCACCAC ATTTTCCATTGTTCC	*PAB1* to pnm1-tt backbone
GS_22-R	TTTGTTTGTTTATGTGTGTTTATTCGAAACTAAGTTCTTGGACTGTCCTAAAAATACAAAATCAAAACTAA AGAATTTTACACAAG	
GS_21-F	CCAAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAAT GAGATTTCCTTCAATTTTTACTGC	mating alpha factor +*TDH3* forward
pGS2		
GS_20A-R	GTAAAAATTGAAGGAAATCTCATTTTATTTTTATTGGTTTTTTAG TTTTTTTTGGATTTTTTTTTTTGAG	*PAB1* to mating alpha factor
GS_25-F	AAATCCAAAAAAAAACTAAAAAACCAATAAAAATAAAATGAGA TTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGCATCCTCCGC	mating alpha factor to *PAB1*
PGS3		
GS_16-R	AAAATTGAAGGAAATCTCATTTAATATGTTCTCGAGTTTTGTGTTCTTAAGTAGATAATTGG	
gBlock	GGACGAGGCAAGCTAAACAGATCTGGGGCGCGCCTAGCATGTG ATTAATTAACTTGTAATATTCTAATCAAGCTAGCACTGTTGGGC GTGAGTGGAGGCGCCGGAAAAAAAGCATCGAAAAAAGGATCCA ATTATCTACTTAAGAACACAAAACTCGAGAACATATTAAATGA GATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGC	gBLock containing L0'UTR
PGS7		
GS_23-f	CCAAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAAT GCAACCATCTACCGCTACCG	Mating A to *TDH3*

GS_TDH3R	CCAAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAATGCAACCATCTACCGCTACCG	*TDH3* reverse
GS_24-R	CCGCCGCTGCCACCGCCACCGGGCCCCTGGAACAGAACTTCCAGTGGCCATGGCTGAGGCCTAGGCTAAGCAATAACACATGCTGGGTCC	Mating A reverse
GS_12AB-F	AGGGGCCCGGTGGCGGTGGCAGCGGCGGTGGCGGTAGCATGGTTTCTAAGGGTGAAGAAG	YmNeongreen forward MatingA
pGS8		
GS_22-R	TTTGTTTGTTTATGTGTGTTTATTCGAAACTAAGTTCTTGGACTGTCCTAAAAATACAAAATCAAAACTAAAGAATTTTACACAAG	
GS_23-F	CCAAGAACTTAGTTTCGAATAAACACACATAAACAAACAAAATGCAACCATCTACCGCTACCG	Mating A to *TDH3* UTR
pPGS_10 (*P.P.* - PABG1B)		
gBLock	GGGGTATCTCTCGAGAAAAGAGAGGCTGAAGCTGAATTCCATGAGGAGCACTCAAGGACGAAAAGGCAAGCTTATACATTTCCCGATGGATTCTTGTTGGGTGCAGCTACAGCTTCATATCAGGTGGAAGGCGCTTGGGACGAGGACGGTAAAACCTCATCTATTTGGGACACACAAACTCATGACAAAAACTACCTGATCGCTGATCACACTACAGGTGATATTGCCTGCGACTCTTACCACAAATACGATGTGGATGTGCAGATGCTGAGAGATCTGGGAGTCGATTTTTACAGATTTAGTTTTAGTTGGCCCAGAATCTTACCCGACGGCCATGGAAATAGAATAAATCAAGCCGGAATTGATTACTACAACAAGCTAATAGACTTGTTAGTCGCCAATAATATCCAGCCCGTTGCCACAATGTACCATTGGGACCTTCCCCAAAATCTTCAGGACCTAGGAGGATGGCCTAATTATGTCCTGGTGGACTACTTTGAGGACTACGCAAGAGTATTATTCAGGAATTTTGGCGATCGTGTGAAATACTGGATCACATTCAATGAACCATTGACTTTCACTGGTGGCTATGAAGGCGCATACGCCCACGCCCCTGCAATTAATGCTCCTGGCTATGGTAGATACTTAGCTACCCATACATTAATTAAAGCTCACGCAAGGGCTTATCACATATACGACGACGAATTTAGAGCAGATCAGCAAGGCAAAGTGTCTATCACTCTGAACGTAGACGCCTGCTTTAACTACCAGAATACGACGGAGTACCAAGATGCATGCGAGCGTCAGCAGCAATTTGAGATGGGACTTTTTTGCTAATCCAATATACTCTGCCGAAGGAGATTGGCCAGCTATCGTGAGAGAGAGAGTTGACGCCAATTCAAAAGCCGAGGGTCTAGCTGAGAGTAGGCTACCCGTCTTCACGCCCGACGAAATAGAGTACATCCGTGGAACATACGATTTTTTTGGTCATAATCACTACACCTCCAACTACGCAATACCATATGACGGAACTAACGATCCCGCTTCAGATCAGAAAGACCACGGTTACTACTTGACGAAAGATCCTAATTGGCCAGGCTCTGCCTCCAGTTGGTTAAAGGTGGTTCCCACAGGTTTAAGATACCAGCTTAATTCCATTGCTACCAGATACAACAACACCCCCCAATTCTGATAACTGAAAATGGCTTTAGTGACTATGGAGACTTAAATGATACAGGCAGGATAAACTACTATACGTCCTATCTGACGGAAATGCTGAGAGCCATAAATGAGGACGGCGTTAATGTTATAGGCTATACTGCATGGAGTCTAATGGACAACTTTGAATGGAATCAGGGTTATTCTGAGAAGTTTGGCTTATATCAAGTGGACTTTGAAGACCCCACCCGTCCACGTATTATGAAGGAAAGTGCAAGAGTGTTCCAGCAAATAATAGCAACGAGGCAGATCCCAGAGGCTTACCGTACAGTCGACCATCATCATCATCATTGAGTTTGTAG	PaBG1b gblock
pPGS_11 (*P.P.* - EXG1)		
gBlock	GTATCTCTCGAGAAAAGAGAGGCTGAAGCTGAATTCATGGCCCCAGTGCATGTTGACAATCCATACGCCGGTGCCGTCCAGTATGTCAATCCCACTTGGGCAGCTTCAGTCAATGCCGCCGCTGGAAGACAATCAGCCGACCCCGCTCTAGCTGCAAAGATGAGGACGGTGGCAGGCCAGCCCACGGCAGTATGGATGGATCGTATTTCCGCTATCACAGGAAATGCCGACGGTAATGGTCTTAAATTTCACCTTGACAACGCCGTAGCTCAGCAGAAAGCCGCCGGTGTACCACTTGTGTTCAACCTGGTGATATATGATCTGCCCGGTAGAGACTGTTTCGCCCTAGCCTCCAACGGTGAGCTACCCGCTACCGATGCAGGTCTGGCTAGGTATAAAAGTGAATATATTGACCCTATAGCCGATCTGCTTGATAACCCTGAGTACGAGAGTATCAGGATCGCAGCTACAATCGAACCTGACTCCCTACCTAACCTTACGACCAATATCAGTGAGCCTGCCTGCCAACAAGCTGCACCCTACTATAGGCAAGGAGTCAAATACGCATTGGATAAGCTGCATGCTATTCCTAACGTTTATAATTATATTGACATAGGTCACTCCGGATGGCTAGGATGGGATTCCAATGCCGGCCCATCTGCAACACTTTTCGCTGAAGTAGCCAAGTCAACGACTGCTGGATTTGCTTCAATTGATGGTTTTGTTAGTGACGTTGCAAATACCACGCCACTGGAGGAGCCCCTATTGTCCGACTCATCTTTAACCATTAATAACACGCCTATTAGGTCTTCCAAGTTTTATGAGTGGAATTTCGATTTTGACGAGATCGATTACACTGCTCATATGCATAGGCTATTAGTTGCTGCTGGCTTCCCATCCTCCATCGGCATGTTGGTTGACACGTCTAGAAACGGATGGGGAGGACCCAATAGACCCACCTCCATTACAGCCTCTACCGATGTTAATGCATATGTAGACGCTAATAGAGTAGACAGAAGGGTGCATAGGGGAGCATGGTGCAATCCTTTAGGTGCAGGTATAGGCAGATTCCCAGAAGCCACGCCTTCTGGATACGCCGCCTCTCACCTTGACGCATTTGTATGGATTAAGCCTCCTGGCGAGTCCGATGGAGCCAGTACAGATATTCCTAACGACCAAGGAAAAAGATTTGATAGGATGTGCGACCCAACTTTCGTCTCACCTAAACTGAACAACCAGCTGACTGGAGCAACGCCAAATGCACCTCTAGCTGGTCAATGGTTCGAGGAGCAATTCGTCACGCTAGTAAAGAACGCATATCCAGTTATTGGAGGAACGACGCCCGTCGAAGATTTAGTTGCCCCCACCGTACCAACTGGATTGACGGCCGGTACGACCACTGCAACGTCTGTGCCCCTATCCTGGACAGCTTCCACAGATAACGTGGCCGTGACAGGCTATGATGTCTATAGAGGTACTACCCTAGTCGGAACTACAGCTGCAACCAGTTACACGGTCACCGGCTTGACTCCAGCTACCGCATACTCCTTTACTGTGAGAGCAAAA	ExG1 gBlock

	GATGCTGCAGGAAATGTCTCCGCAGCTAGTGCAGCCGCCGCTG CAACTACGCAATCCGGAACTGTTACCGACACAACCGCTCCTTCT GTCCCTGCCGGACTAACAGCCGGAACCACGACGACAACTACGG TTCCTCTGTCTTGGACGGCTTCCACAGACAACGCAGGTGGTAGT GGCGTTGCTGGTTATGAGGTCTTAAGAGGAACCACTGTGGTGG GTACCACTACGGCCACTTCCTATACTGTTACAGGCTTAACGGCA GGTACTACCTATTCATTCAGTGTGCGTGCCAAGGACGTTGCCGG AAATACGTCAGCAGCCTCCGCAGCTGTAAGTGCCACGACGCAA ACAGGTACCGTCGTGGATACAACGGCCCCATCTGTCCCCACAG GTTTGACCGCTGGAACAACCACGACTTCATCCGTCCCACTAACA TGGACTGCAAGTACAGACAACGCAGGAGGTTCTGGCGTAGCCG GCTATGAAGTGTTTAATGGCACAACCAGAGTCGCAACAGTAAC GAGTACCTCATATACCGTGACGGGTTTAGCTGCAGACACCGCA TATTCTTTCACCGTCAAGGCAAAGGACGTAGCTGGCAATGTATC AGCCGCATCTGCCGCAGTCTCCGCAAGAACACAGGCAGCCACG TCAGGAGGCTGTACGGTGAAGTATTCCGCTTCATCTTGGAATAC GGGATTTACAGGTACGGTGGAAGTCAAGAACAACGGTACCGCC GCACTGAATGGATGGACGTTAGGTTTTTCTTTTGCCGACGGACA GAAAGTATCACAGGGCTGGTCTGCTGAGTGGTCTCAATCTGGA ACGGCCGTGACAGCTAAAAATGCACCCTGGAATGGAACCCTAG CCGCTGGTAGTTCAGTGTCCATCGGCTTCAATGGCACACACAAC GGCACTAACACCGCACCTACGGCCTTCACACTAAACGGTGTTG CCTGCACATTAGGTGTCGACCATCATCATCATCATTGAGTT TGTAGCCTTAGACATGACTGTTCCTCAGTTCAAGTTGGGCACTT ACGAGAAGACCGGTCTTGCTAGATTCTAA	
pPGS_12 (*P.P.* - ENG1)		
gBlock	GTATCTCTCGAGAAAAGAGAGGCTGAAGCTGAATTCATGGACA CGCAAATTTGCGAACAGTATGGTTCAACAGTGGTTGGAGGTAG ATACGTCGTCCAAAACAATCGTTGGGGAACTACCGCTCAACAG TGCATCAATGCTACGAGTAATGGTTTCGAGATAACGACTCTTAA CGGTAGTTCCCCCACGAACGGAGCCCCCACGGCTTACCCCTCA ATCTTCTTTGGCTGTCACTACACTAACTGCTCTCCTGGTACGAA TTTACCAATACAGGTTTCCCAAATATCCTCCGCAACTTCTTCAA TTTCATACAGGTACGTTAGTGGAGCTACATATAACGCTAGTTAT GACATTTGGCTAGACCCTTCACCAAAGAGGGATGGTGTTAATC AAATGGAGATTATGATATGGCTGAATCGTCAGGGACCTATTCA ACCAATTGGTTCAGTTGTGGGCACAACCAACTTGGCCGGAAGG ACGTGGGAGGTATGGCGTGGCTCTAATGGCAGTAATAACGTGA TCTCCTACGTTGCCCAATCCACGACTTCAAGTTTGAATTTCTCTG TTTTAGACTTCATCAATGACACAAGGAATCGTGGTGCTATTTCA AATTCATGGTACTTAACTTCTATACAGGCTGGCTTTGAACCTTG GCAGGGTGGTGTAGGCTTAGCTGTAACCAGTTTCTCAGCCAAC GTAAATGGTGGAGGAACGCCTCCCACTAATCCATCTACCCCTAC CCCCCCTCCTTCTGGCGGTGCCGGTTGTCGTGTTAAATATACTG CCAACGCCTGGAATAACGGTTTTACAGCCGACGTGCAGATAAC CAATACGGGTTCATCAGCCATTAACGGATGGACCTTAACGTAC GGTTTACCAGGCGGACAGCAAGTAACCTCAGCCTGGAACGCTA CTGTGTCCCAATCAGGCTCCACAGTTACGGCTCGTAACGTTTCT CATAACGGCTCCGTCGCCCCAGGTGGCACCGCTTCTTTCGGATA CCAAGGAACTTTATCCGGCTCCTACTCATCTCCCACCTCTTTTA GTTTAAACGGTACAACCTGTTCCCGTGCTGTCGACCATCATCAT CATCATCATTGAGTTTGTAGCCTTAGACATGACTGTTCCTCAGT TCAAGTTGGGCACTTACGAGAAGACCGGTCTTGCTAGATTCTAA	EnG1 gBlock
pPGS_13 (*P.P.* - BG1)		
gBlock	GTATCTCTCGAGAAAAGAGAGGCTGAAGCTGAATTCATGCCTA GTCTGAGATTTCCCGATAACTTCGTCTGGGGGACGGCCACGAG CAGCTATCAAATTGAAGGAGCCACAAACGAAGACGGGAGGAG TGAGTCAATATGGGATAGATTTGCAAAGACTCCTGGTCAGATA GTCGACGGGACGAACGGTGACGTAGCCTGCGATCATTATCACA GGTGGCCGGAAGATATCGCCCTAATGAAGTCCCTAGGAATGAT GGCATATAGATTCTCTATCGCGTGGCCGCGTATACTGCCAAACG GCCGTGGTGCTGTGAACCAGAAAGGGCTGGACTTCTATTCCAG ATTAGTCGACGGATTGTTAGCGGCTGGTATTGAGCCTTACGCGA CACTTTTCCATTGGGACCTGCCCCAAGTACTACAGGATCAAGTT GGTGGTTGGCAACATCGTTCCACCGCCGACGCTTTCGTGGAGTT TACTGACGTAATTAGTAGAAAGCTAGGTGACAGAGTGAAAAAA TGGATTACTCACAATGAACCGTGGTGTACCGCTTTCTTGTCCTA CGAGAAGGGTATACATGCACCGGGAATGAGGGACTTTAAGAAG GCATTGGCGGTATCCCACCACTTGTTGCTTAGCCATGGACTAGC GGTGCCGGTTGTCCGTCAAAATTGCCCAGGGGCGGAAGTGGGT ATTACTCTTAATATGAACTACGCAATGCCCGCGAGTCCCAGTGC AGCGGACTACGATGCATGTAGGCACTATGACGGATATTTTAAC CGTTGGTTCCTGGACCCGCTTTATGGTAGACGTTATCCCGCCGA CATGATTGAGGATTATATTGCGCTGGGCCACTTGCCCCCAGAGG GTCTGACTGTCTGTAAGCCGGGTGACCTTGACATCATCGCCACC CAGTGTGATTTTTTAGGACTTAATTATTACTCCCGTGCCGTTTTA CGTTCATCAAAAGTTCCTGAGGAGCAAAATCTTCCTAGAACAG TACACGTAGCTCCACTGTCAGAGCAAACGGAAATGCAATGGGA GGTGTACCCAGAAGGCCTTAGGAGGTTACTTTTAAGATTACATG TTGAGTACGGCGTTCCCAAATTGTACGTAACTGAGAATGGGGC AGCTTACAGCACAGAACCTGATAAAGATGGCAGGATTCCAGAT ACCCGTCGTTTGCAGTTTTTAAGAGACCATTTCGTAGCGGCTCG TCGTGCGATGGACGCGGGTGCTCCTCTAGCTGGGTACTTTGTAT GGTCCCTAATGGACAATTACGAATGGGACAGGGGTTACAGTCA GCGTTTTGGCATTGTGTGGGTGGACTACGAAACTCAAAGGAGG ATCCCGAAGGATAGTGCGCTTTGGTACAAGGGTGTTATCGCGG CTAATGCGGTCGATGACGGCAGTTCACAGGCAACTGTCGACCA TCATCATCATCATCATTGAGTTTGTAGCCTTAGACATGACTGTT CCTCAGTTCAAGTTGGGCACTTACGAGAAGACCGGTCTTGCTAG ATTCTAA	BG1 gBlock

pPGS_14 (*P.P.* - EXG2)		
gBlock	GTATCTCTCGAGAAAAGAGAGGCTGAAGCTGAATTCATGGGGA CGATCTCTGGCTCCCTTTATCGTGAGCCTGATACAAGCGTATCC AGATGGGTCGCAGCCAATCCTGCTGATAGTAGGATGCCTGTTAT CAGAGATCGTATAGCATCACAGCCCGCCAGCCATTGGTTAAGC AACTTCAATCAAAGTACGATTCGTTCAGAGGTCAGTGGGTATGT TGGGGCGGCCAACGCGGCTGGGCAAATACCTGTTTTAACGGTA TACGGCATTACGAATAGAGATTGTGGAGGCGCATCCTCCGGGG GTGCTCCCGACCTGACCCAGTACCAAACATGGATCTCTAATTTG GCTGGCGGATTAGGTAACAGAACAGTAGTCATCATCCTAGAGC CAGACTCAATAGCTTTACAAACCTGCCTATCAGGGGCGGATCTT GCCGCCCGTAACCAAGCTTTGCACACTGCTGTTCGTACGCTGAA AAGCGGCAATCCTTCCAGCAAAATTTATTTGGACGCCGGTCACT CTACATGGAATTCCGCTGCGGAACAAGCTAGGCGTCTTATTGCA GCGGGAATAGCCGACGCGGACGGCTTTTACACGAATGTCAGCA ACTTCAACCCAACGTCTGGGGAAGCCTCTTACGGTCGTTCAATT ATCTCCTCTTTGAGTGCACAGGGTGTGTCCGGTAAACGTCAAGT CATCGACACCTCAAGGAACGGCGGGGCGAGCGGCGACTGGTGT GGGGATGATAATACAGATAGAAGGATTGGCAGGTATCCTACGG TAGCCACCGGTGACGATAATATCGATGGATACCTATGGGTCAA ACCACCAGGCGAGGCGGACGGGTGTAGATATGCTGCTGGTTCC TTCCAGCCTGATCTGGCATATTCTCTAGCGTCCAGCGCGACTAA CCCTCCATCACCGTCACCATCCAACCCATCCAGCAGTCCCTCCA GTAGTCCTTCATCATCCCCGTCCGGTTCACCATCAGCGAGTCCC TCCGCGTCTCAGCCCACCGGAGCCTGTAGCGCAACCTTCAGAGT GACTGGCTCCTGGCAGGGTGGTTTTCAGGGAGAGGTGACGGTG ACGGCGGGTACCGCTGCGCTAAGTGGCTGGAGCGTCAGCTGGC CCCTGTCTGCTGGACAAACTATTACGCAAGCGTGGAATGGCAC ACTAAGTGCGACAGGGTCCACGGTCACGGTAAGGAATGTCTCT TGGAACGGCGCTTTGGCGCCCAGGGCCACCGCCCAGTTTGGTTT CCTTGCCAATGGTGCAGCTACAACGCCGACTCTTACGTGCGCTG CCGCGGTCGACCATCATCATCATCATTGAGTTTGTAGCCTT AGACATGACTGTTCCTCAGTTCAAGTTGGGCACTTACGAGAAG ACCGGTCTTGCTAGATTCTAA	ExG2 gBlock
pPGS_15 (*P.P.* - BX1)		
gBlock	GTATCTCTCGAGAAAAGAGAGGCTGAAGCTGAATTCATGGCTA CCACTTTGAAAGAAGCTGCTGATGGTGCTGGTAGAGATTTTGGT TTTGCTTTGGATCCTAACAGATTGTCCGAAGCTCAATACAAAGC TATTGCCGATTCTGAATTCAACTTGGTTGTTGCTGAAAACGCTA TGAAGTGGGATGCTACTGAACCATCCCAAAATTCTTTTTCTTTC GGTGCTGGTGATAGAGTTGCTTCTTATGCTGCTGATACCGGTAA AGAATTATACGGTCATACTTTGGTTTGGCACTCTCAATTGCCAG ATTGGGCTAAGAATTTGAATGGTTCTGCTTTCGAATCCGCTATG GTTAATCATGTTACTAAGGTTGCCGATCACTTCGAAGGTAAAGT TGCTTCATGGGATGTTGTTAATGAAGCTTTTGCTGACGGTGGTG GTAGAAGACAAGATTCTGCTTTTCAACAAAAGTTGGGTAACGG TTACATTGAAACTGCTTTTAGAGCTGCTAGAGCAGCTGATCCAA CTGCTAAATTGTGTATCAACGATTACAACGTCGAAGGTATTAAC GCCAAGTCCAACTCATTATACGATTTGGTTAAGGACTTCAAGGC TAGAGGTGTTCCATTGGATTGTGTTGGTTTTCAATCCCATTTGAT CGTTGGTCAAGTTCCAGGTGATTTCAGACAAAACTTGCAAAGA TTTGCCGATTTGGGTGTTGATGTTAGAATCACCGAATTGGACAT TAGAATGAGAACTCCATCTGATGCTACTAAGTTGGCTACTCAAG CTGCCGATTACAAAAAAGTTGTTCAAGCCTGTATGCAAGTCACC AGATGTCAAGGTGTTACTGTTTGGGGTATTACCGATAAGTATTC TTGGGTTCCAGATGTTTTTCCTGGTGAAGGTGCTGCTTTAGTTTG GGATGCTTCTTACGCTAAAAAACCAGCTTACGCTGCTGTTATGG AAGCCTTTGGTGCTTCTCCAACTCCAACACCAACTACTCCTACA CCTACTCCAACTACACCAACCCCAACTCCTACTTCTGGTCCAGC TGGTTGTCAAGTTTTATGGGGTGTTAATCAATGGAATACTGGTT TCACTGCTAACGTTACCGTTAAGAACACTTCTTCTGCTCCAGTT GATGGTTGGACTTTGACATTTTCTTTTCCATCCGGTCAACAAGT TACACAAGCTTGGTCATCTACTGTTACTCAATCTGGTTCTGCAG TTACTGTTAGAAATGCTCCATGGAATGGTTCAATTCCAGCCGGT GGTACTGCTCAATTTGGTTTTAATGGTTCTCATACCGGTACTAA TGCTGCTCCAACAGCTTTTTCATTGAATGGTACTCCATGTACCG TTGGTGTCGACCATCATCATCATCATCATTGAGTTTGTAGCCTT AGACATGACTGTTCCTCAGTTCAAGTTGGGCACTTACGAGAAG ACCGGTCTTGCTAGATTCTAA	Bx1 gBlock
pPGS_16 (*P.P.* - BG2)		
gBlock	GTATCTCTCGAGAAAAGAGAGGCTGAAGCTGAATTCATGTGCG AAAAAAAGACTGAAGCTGGTTCTAAGCCACAAGCTTTTGACAA AGAAGTTCAAGACTTGTTGAAGAACATGTCCTTGGAAGAAAAA GCTGGTCAAATGACCCAAATCGACATCAGAAATTTGTTGAACA ACGGTTACGGTAACACCGACGAAAAATTGGATACTGCCAGATT GAAAGAAGCCATCCAAACTTATCATGTCGGTTCCATTTTGAACT GCATTCAAGCTTACACTCCAGAAAAGTGGGTTGAATTGATCTCC CAAATTCAAAACGAAGCCTTGCAATCCCCAAACAAAATTCCAG TCTTGTATGGTACTGATGCTATGCATGGTGTTGGTTTCATTAAG GATGCTGTTTTGTTCCCACATAACATTGGTATGGCTGCTTCTAG AAACGATCAATTGGTTTCTCAAGCTGCTCAAGTTACTTCTACTG AAGCTAGATCTGTTGGTTTGACTTGGAATTTCGCTCCAGTTTTG GATGTTGGTAGAGAACCATATTGGTCCAGATTCGAAGAAACTT TTGGTGAAGATGTTTACATCACCACCCAAATGGGTTCTGCTGCT GTTCAAATGATGGAAGGTTCTGATTTGACTTCCAAGACCAACAT TGCTTCTTGTTTGAAGCACTTCATTGGTTACTCTGCTCCAAAGA ACGGTATTGATAGAACCCAATCTCATATCCCAGAAATCGTCTTG AGAGAATATTACTTGCCACCATTCAGAGAAGCTATTAACAAAG GTGCTTCCTCCATCATGATTAACTCTGCTGAAATCAACGGTATT	BG2 gBlock

	CCATGCCATGGTAACAAATGGTTGTTGACCGATTTGTTGAGAAC CGAATTGGGTTTTACTGGTATGGTTGTTTCTGATTGGGAAGATG TCATTAGATTGCATACCTGGCATAAGGTTGCTGCTACTCCAAAA GAAGCTGTTATGATGGCTGTTAATGCTGGTGTTGATATGTCTAT GGTTCCAAACGATTACTCATTCCCAAAGTACTTGGTCGAATTGG TCAAAGAAGGTAAGGTTTCCATGGCCAGAATTGATGAAGCAGT TGGTAGAATTTTGACCTTGAAGATCAAGTTGGGTTTGATGAAGA ATCCATTGCCATCCATTGCTGATGTTGGTGTTGTTGGTTCAGAT GCTCATCAACAAATAGCTTTGAATGCTGCCAGAGAATCTATCAC CTTGTTGAAAAACGACAAGAACATTTTGCCATTGGCCAAGGAT AAGAAGATTTTGTTGGTTGGTCCAGCTGCTAATTCTTTGTCTGC TTTACATTCTTCTTGGTCCTATACTTGGCAAGGTTCCAACGAAT CATTATACCCAGAAACTACCAAGACCATAAGAGAAGCTTTGGA AGCTTCAGGTAACAAGGCTAACATTAGAACTAATGCTACTACC GGTTTTGATGATGCTGCTAACTACGATGTTTCCTTCATTCAAAA GAATACCGCCGGTGTTGATGTTATCATAGTTTGTGTTGGTGAAG CTGCTTATGCTGAACAACCAGGTGTTATTAAGGATTTGAACTTG CCAGAAGCCCAAAAGCAATTGATTGTCGCTGCTAAAAAAACTG GTAAGCCAGTTATCGTTTGCTTGGTTGAAGGTAGACCAAGATTA TTCCCAGAAGAAGAAGCTTTAGCTGACGCTGTTATTATGTGTTA TAGACCAGGTTCTAAAGGTGCTGATGCTTTCGCTGAAATCTTGT ACGGTGATATTAACCCATCTGGTAAGTTGCCATTCACTTACCCA AGATACGATGGTGACATTACTACCTACGACTACAAGTTTAAAG AAACCGAACAACAATTGAAGCCAGGTGTTTCTGAATTCGTTGCT TTTAATCCACAATGGCCTTTTGGTCATGGTTTGTCTTATACCACT TTCGCCTACTCTAACTTGAACGTCAACAAGAGTAACTTCACCAA GAACGATTCTGTTTTGGTTACCGTTGACATTTCTAACACTGGTG CTAGAACTGGTAAAATCGCCGTTGAATTATACTCCAGAGATCAT TTCGCTTCCATTACCCCATCTGAAAGAAGATTGAGAAAGTACAC CAAGATCGAATTGAAGGCTGGTGAAAAGCAAACTGTCTCTTTC ACTATTAAGGCTGCTGACTTGCAATTCGTTAACAAGGACTTGAA AACTGTTACTGAAGCCGGTGCTTTCGATTTGATGATTGGTAACT TGCAAACCGAAATCTACTTCAACGAAGTCGACCATCATCATCAT CATCATTGAGTTTGTAGCCTTAGACATGACTGTTCCTCAGTTCA AGTTGGGCACTTACGAGAAGACCGGTCTTGCTAGATTCTAA	
pPGS36 (*P.P.* - PABG1B-POP)		
Pre-Ost1-Pro	TTGAGAAGATCAAAAAACAACTAATTATTGAAAGAATTCATGA GGCAGGTTTGGTTCTCTTGGATTGTGGGATTGTTCCTATGTTTTT TCAACGTGTCTTCTGCTGCTCCAGTCAACACTACAACAGAAGAT GAAACGGCACAAATTCCGGCTGAAGCTGTCATCGGTTACTCAG ATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAAC AGCACAAATAACGGGTTATTGTTTATAAATACTACTATTGCCAG CATTGCTGCTAAAGAAGAAGGGGTATCTCTCGAGAAAAGAGAG GCTGAGGCGCATGAGGAGCACTCAAGGACGAAAAGGCAAGCT TA	Pre-Ost1-Pro gBlock
Pr-GS-POPBGL-1F	CTCTCGAGAAAAGAGAGGCTGAGGCGCATGAGGAGCACTCAA GGACG	Pabg1b to Pre-Ost1-Pro
Pr-GS-BGL-1R	GCAAATGGCATTCTGACATCCTCTTGAGCGGCCGCCCTAGTCAA TGATGATGATGATGATGGTCGACTGTACG	PaBG1b to Pre-Ost1-Pro
pPGS38		
gBLock	GAATTCCGAAACGATGAGATTCCCATCTATTTTCACCGCTGTCT TGTTCGCTGCCTCCTCTGCATTGGCTGCCCCTGTTAACACTACC ACTGAAGACGAGCTTGAGGGTGATTTCGACGTCGCTGTTTTGCC TTTCTCTAACTCCACTAACAACGGTTTGTTGTTCATTAACACCA CTATCGCTTCCATTGCTGCTAAGGAAGAGGGTGTCTCTCTCGAG AAGAGAGAG	Mating Factor alpha Del 30-43
pPGS39		
gBLock	ATGAGATTCCCATCTATTTTCACCGCTGTCTTGTTCGCTGCCTCC TCTGCATTGGCTGCCCCTGTTAACACTACCACTGAAGACGAGAC TGCTCAAATTCCAGCTGAAGCAGTTATCGGTTACTCTGACCTTG AGGGTGATTTCGACGTCGCTGTTTTGCCTTTCTCTGCTTCCATTG CTGCTAAGGAAGAGGGTGTCTCTCTCGAGAAGAGA	Mating Factor alpha Del 57-70
pPGS40		
Pr-GS-1-F	TGGCCTTTTGCTGGCCTTTTGCTCAATAACTTCGTATAGCATAC ATTATACGAACGGTAAATTCCCGCTTTGACTGCCTG	PAOx1+lox66 to PucOri
Pr-GS-1-R	GCAAGGGAAATGTCATGCACAAACGAAGGTCTCACTTAATCTT C	AOX1TT to HIS4
Pr-GS-2-F	GACCTTCGTTTGTGCATGACATTTCCCTTGCTACCTGC	HIS4+lox71 to pAOX1
Pr-GS-2-R	ATAACTTCGTATAATGTATGCTATACGAACGGTATTAAATAAGT CCCAGTTTCTCCATACGAACC	HIS4+lox71
Pr-GS-3-F	TACCGTTCGTATAGCATACATTATACGAAGTTATAACATCCAAA GACGAAAGGTTGAATGAAACC	PAOx1+lox66
Pr-GS-3-R	TTTCAATAATTAGTTGTTTTTTGATCTTCTCAAGTTGTCG	PAOx1
gBlock	CGACAACTTGAGAAGATCAAAAAACAACTAATTATTGAAAATG TCCAATTTACTGACCGTACACCAAAATTTGCCTGCATTACCGGT CGATGCAACGAGTGATGAGGTTCGCAAGAACCTGATGGACATG TTCAGGGATCGCCAGGCGTTTTCTGAGCATACCTGGAAAATGCT TCTGTCCGTTTGCCGGTCGTGGGCGGCATGGTGCAAGTTGAATA	Cre-recombinase gBlock

	ACCGGAAATGGTTTCCCGCAGAACCTGAAGATGTTCGCGATTA TCTTCTATATCTTCAGGCGCGCGGTCTGGCAGTAAAAACTATCC AGCAACATTTGGGCCAGCTAAACATGCTTCATCGTCGGTCCGG GCTGCCACGACCAAGTGACAGCAATGCTGTTTCACTGGTTATGC GGCGGATCCGAAAAGAAAACGTTGATGCCGGTGAACGTGCAAA ACAGGCTCTAGCGTTCGAACGCACTGATTTCGACCAGGTTCGTT CACTCATGGAAAATAGCGATCGCTGCCAGGATATACGTAATCT GGCATTTCTGGGGATTGCTTATAACACCCTGTTACGTATAGCCG AAATTGCCAGGATCAGGGTTAAAGATATCTCACGTACTGACGG TGGGAGAATGTTAATCCATATTGGCAGAACGAAAACGCTGGTT AGCACCGCAGGTGTAGAGAAGGCACTTAGCCTGGGGGTAACTA AACTGGTCGAGCGATGGATTTCCGTCTCTGGTGTAGCTGATGAT CCGAATAACTACCTGTTTTGCCGGGTCAGAAAAAATGGTGTTGC CGCGCCATCTGCCACCAGCCAGCTATCAACTCGCGCCCTGGAA GGGATTTTTGAAGCAACTCATCGATTGATTTACGGCGCTAAGGA TGACTCTGGTCAGAGATACCTGGCCTGGTCTGGACACAGTGCCC GTGTCGGAGCCGCGCGAGATATGGCCCGCGCTGGAGTTTCAAT ACCGGAGATCATGCAAGCTGGTGGCTGGACCAATGTAAATATT GTCATGAACTATATCCGTACCCTGGATAGTGAAACAGGGGCAA TGGTGCGCCTGCTGGAAGATGGCGATTAGGGCCGCTCAAGAGG ATGTCAGAAT	

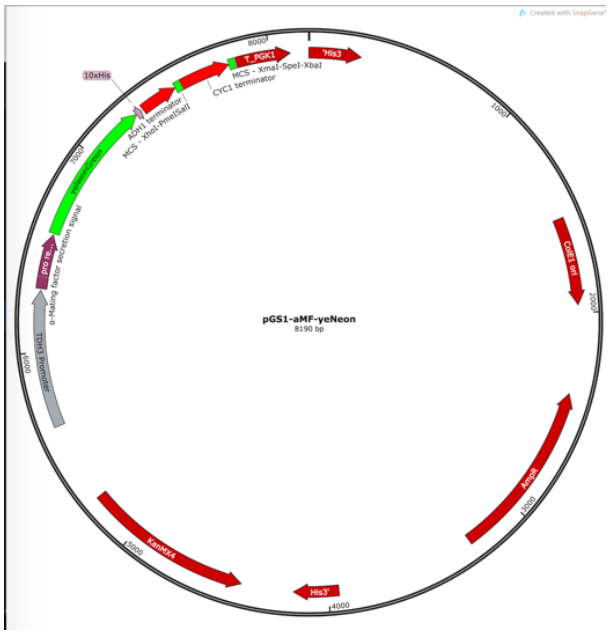

Figure S2.1: Plasmid map of pGS1 (S.c. - YmNG). Map of plasmid transformed into S. cerevisiae to create S.c. – YmNG.

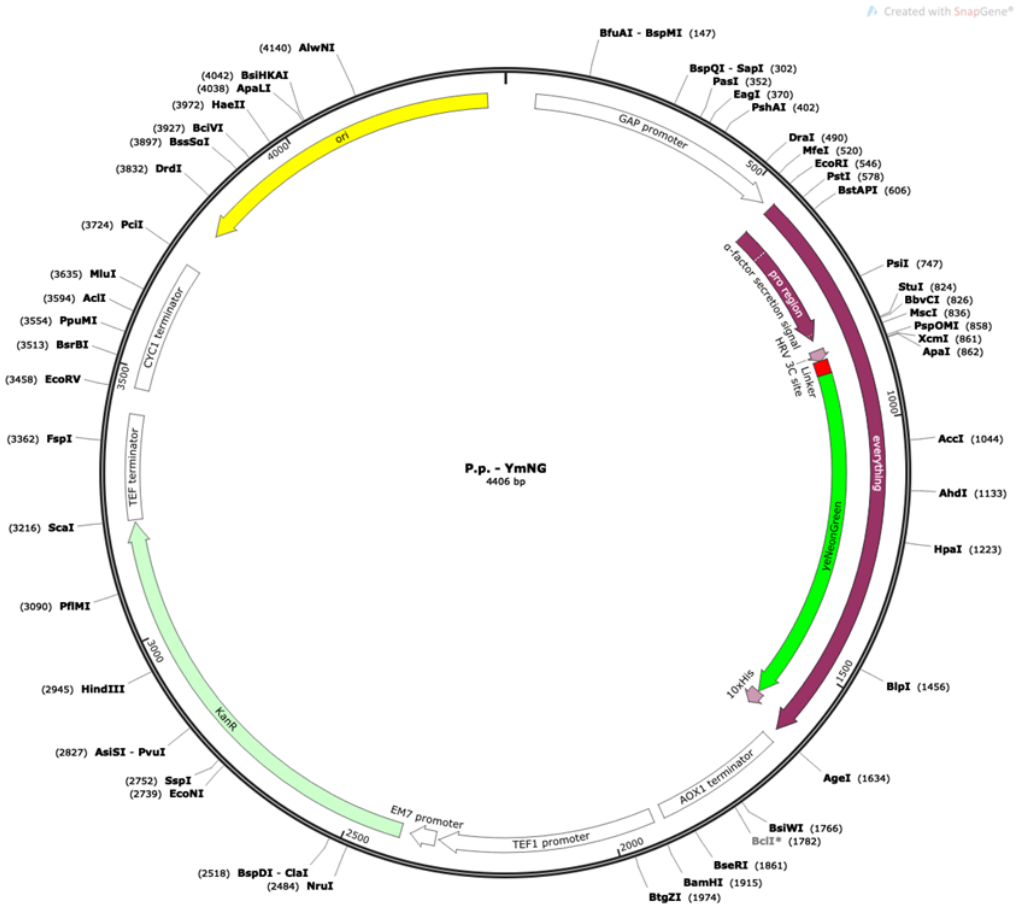

Figure S2.2: Plasmid map of pPGS1 (P.p. - YmNG). Map of plasmid transformed into P. pastoris to create P.p. – YmNG.

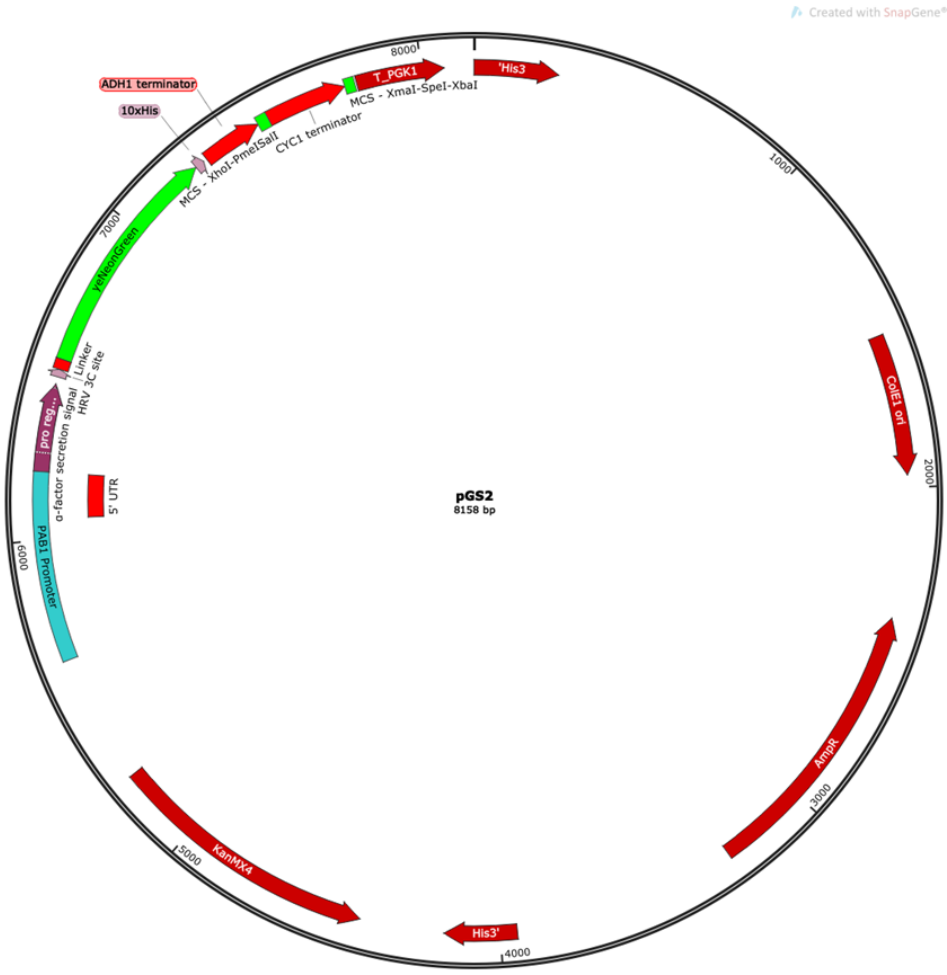

Figure S2.3: Plasmid map of pGS2. Map of plasmid transformed into S. cerevisiae to create S.c. –pGS2.

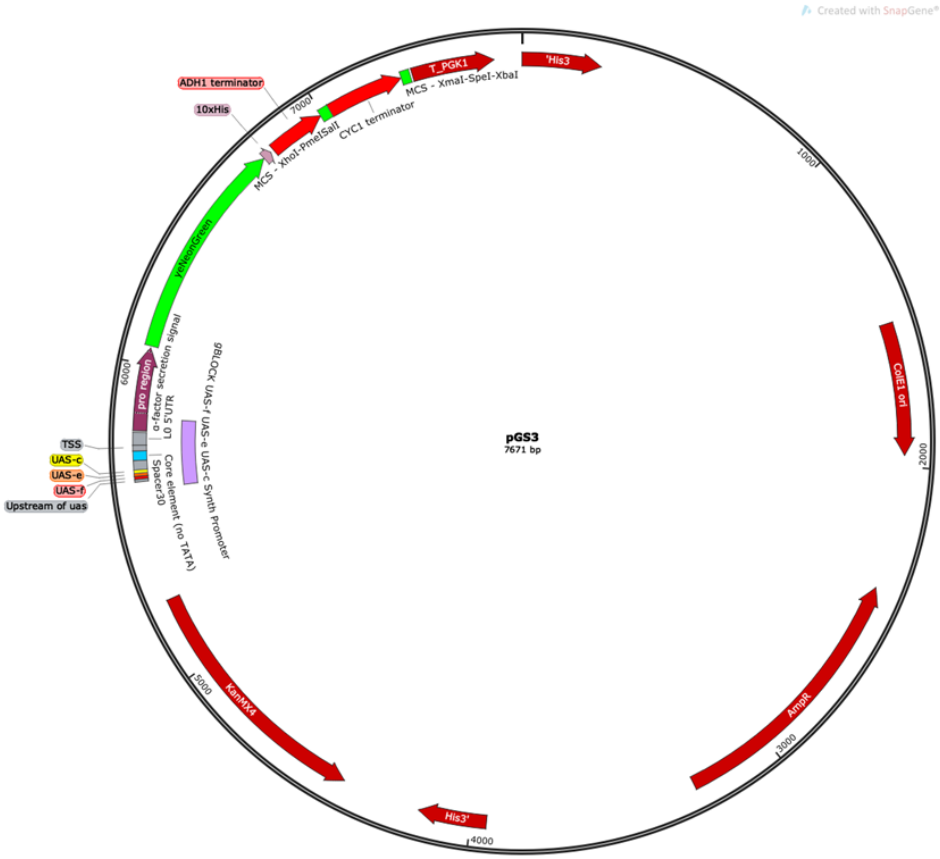

Figure S2.4: Plasmid map of pGS3. Map of plasmid transformed into S. cerevisiae to create S.c. –pGS3.

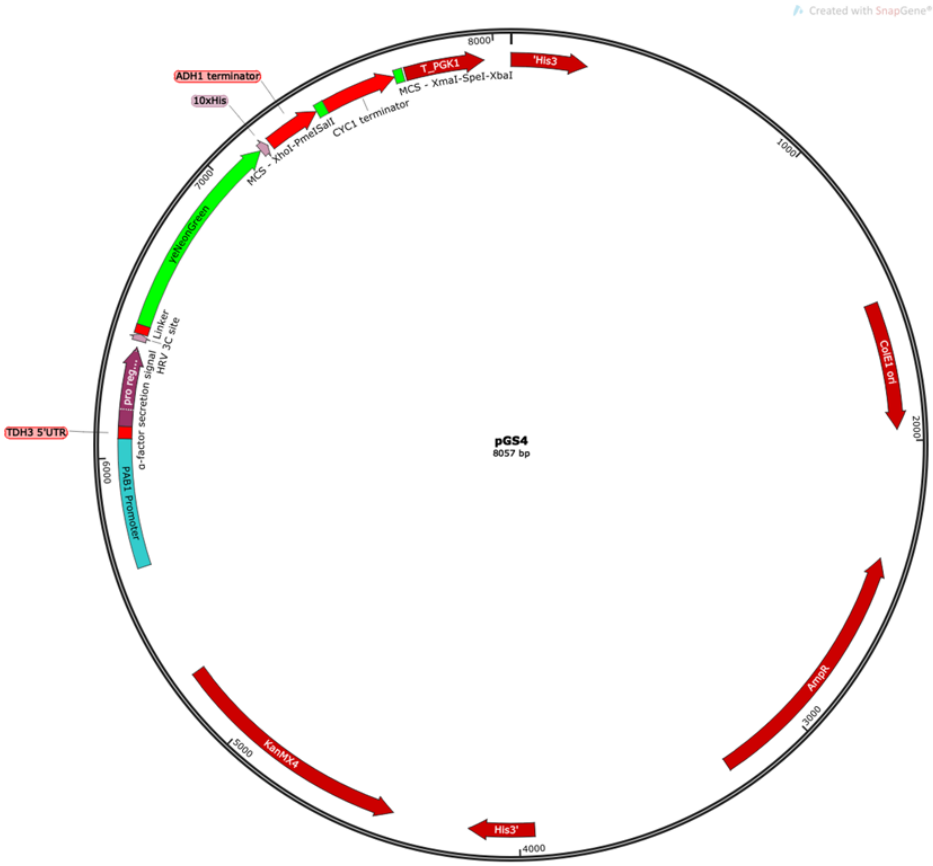

Figure S2.5: Plasmid map of pGS4. Map of plasmid transformed into S. cerevisiae to create S.c. –pGS4.

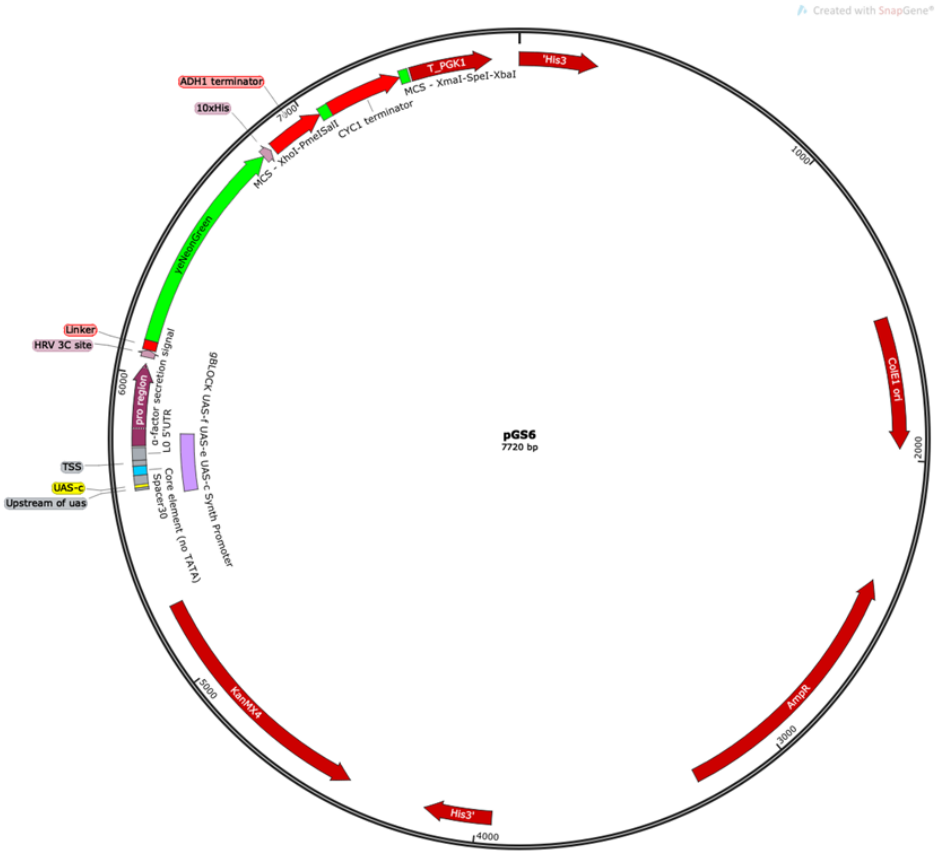

Figure S2.6: Plasmid map of pGS6. Map of plasmid transformed into S. cerevisiae to create S.c. –pGS6.

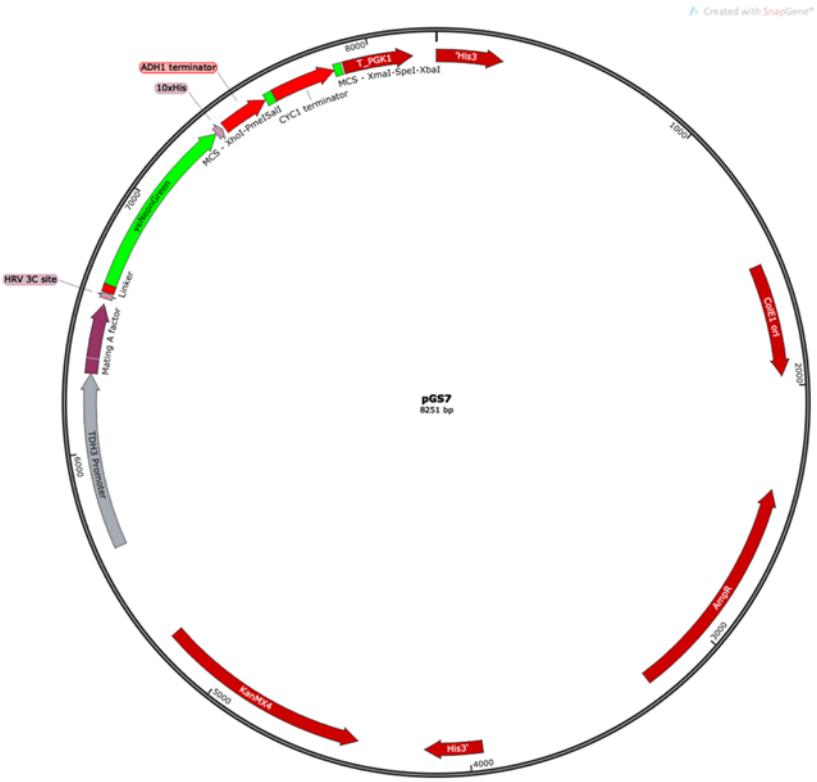

Figure S2.7: Plasmid map of pGS7. Map of plasmid transformed into S. cerevisiae to create S.c. –pGS7.

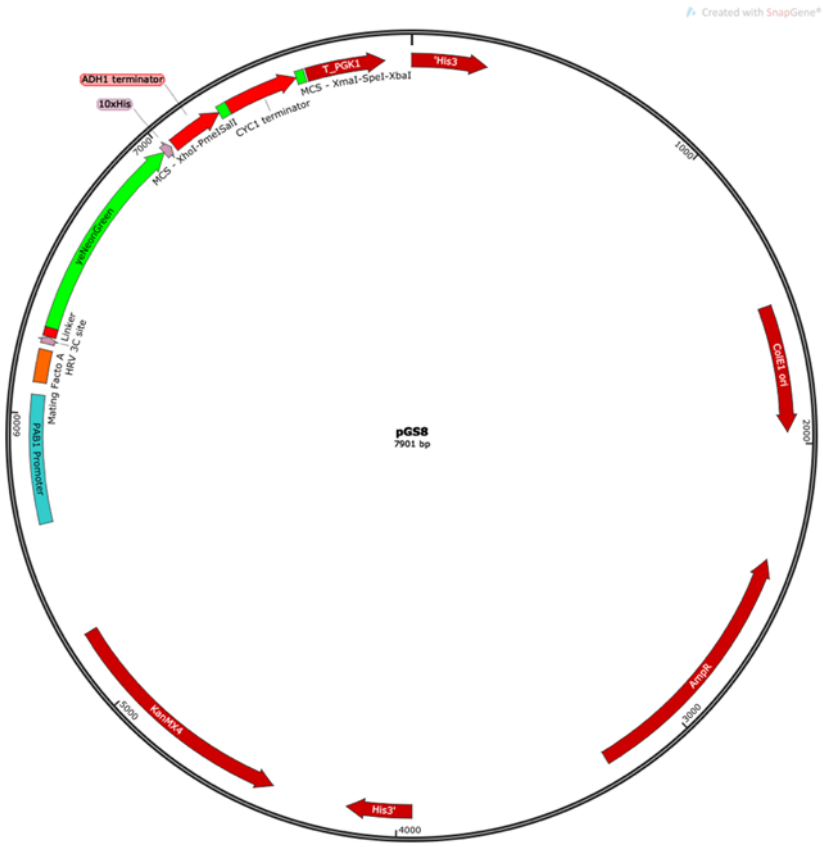

Figure S2.8: Plasmid map of pGS8. Map of plasmid transformed into S. cerevisiae to create S.c. –pGS8.

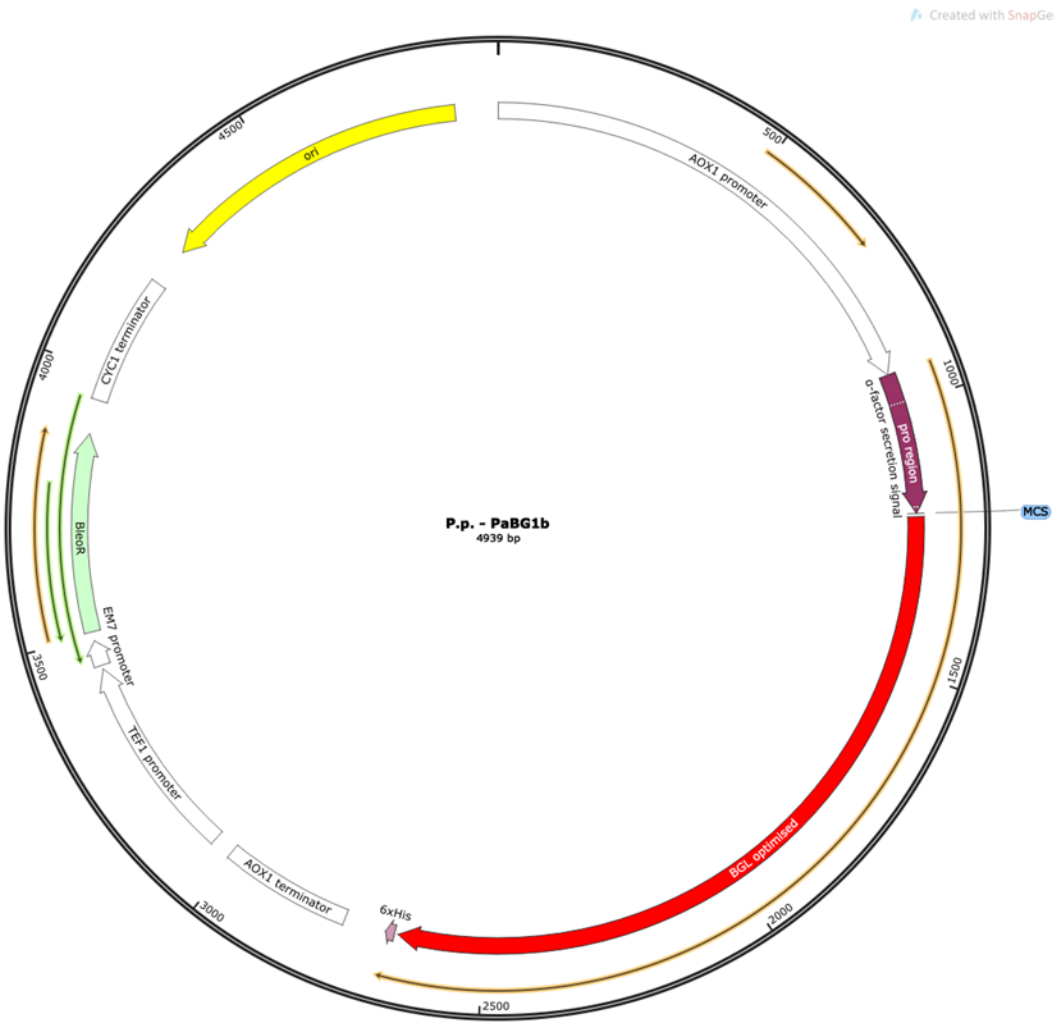

Figure S2.9: Plasmid map of plasmid containing PaBG1b. Map of plasmid transformed into P. pastoris to create P.p. –PaBG1b.

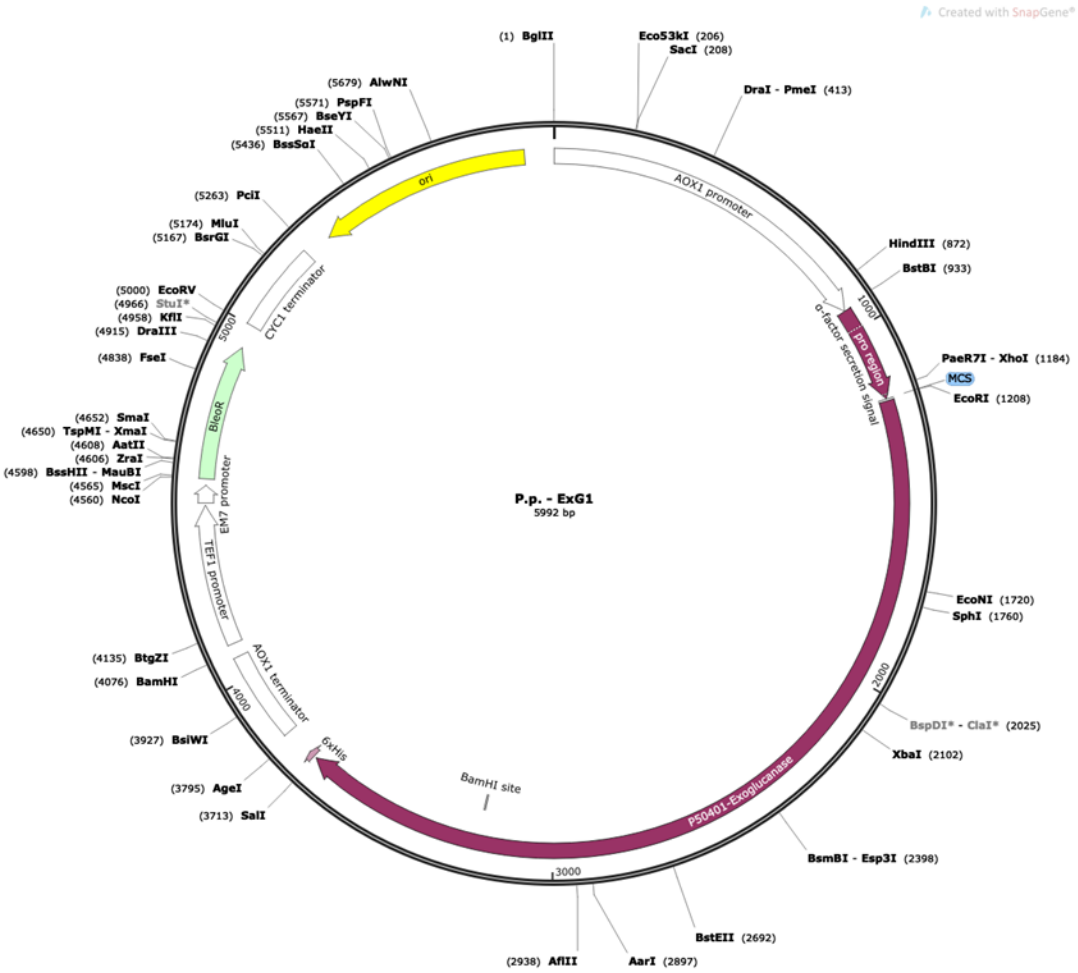

Figure S2.10: Plasmid map of plasmid containing ExG1. Map of plasmid transformed into P. pastoris to create P.p. –ExG1.

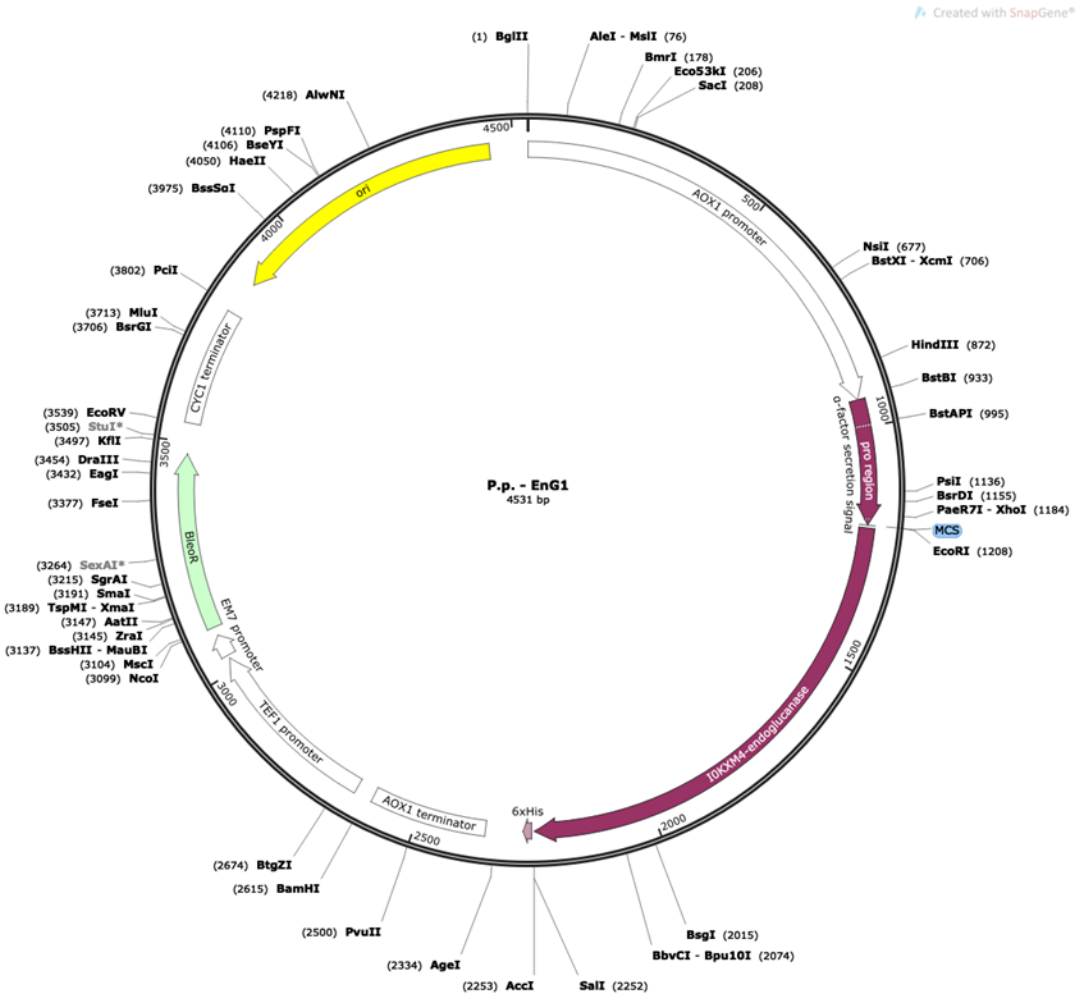

Figure S2.11: Plasmid map of plasmid containing EnG1. Map of plasmid transformed into P. pastoris to create P.p. –EnG1.

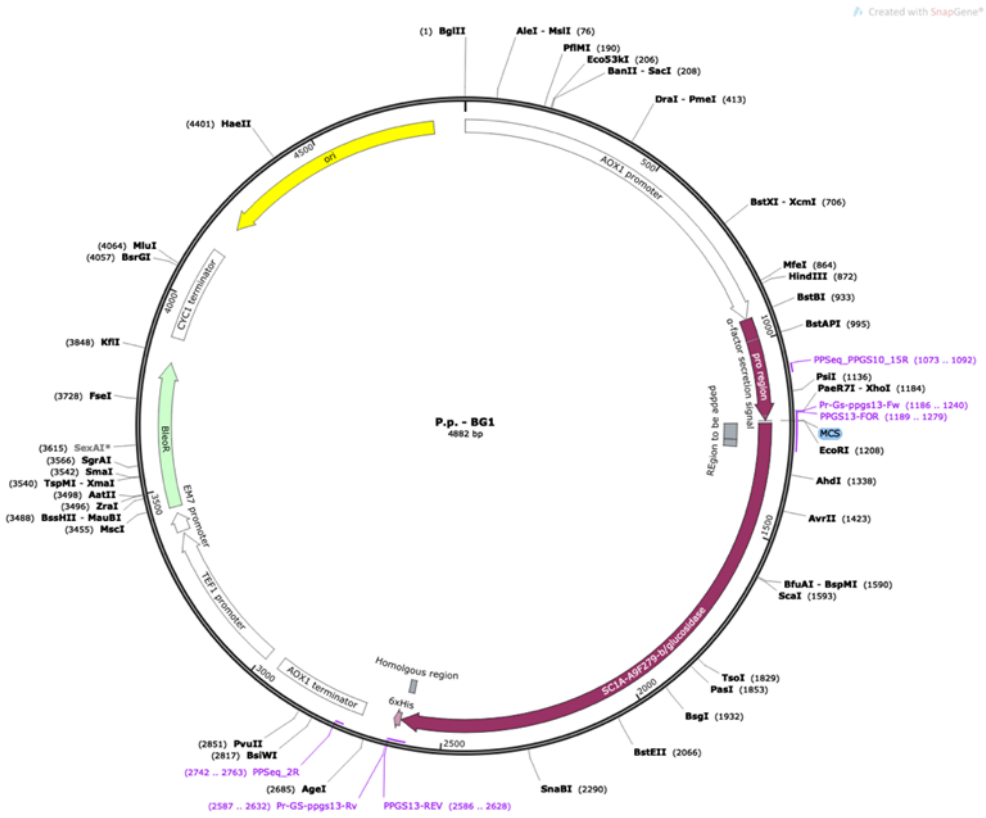

Figure S2.12: Plasmid map of plasmid containing BG1. Map of plasmid transformed into P. pastoris to create P.p. –BG1.

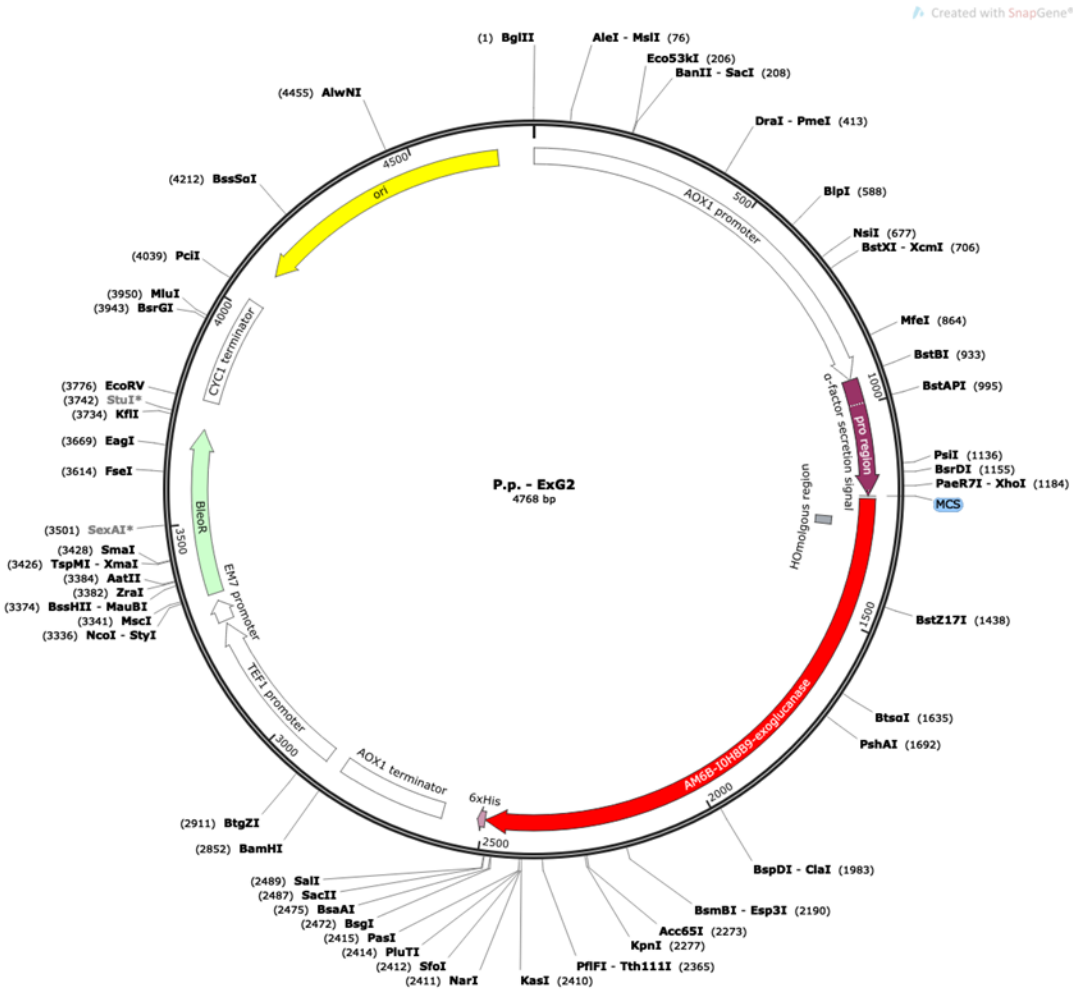

Figure S2.13: Plasmid map of plasmid containing ExG2. Map of plasmid transformed into P. pastoris to create P.p. –ExG2.

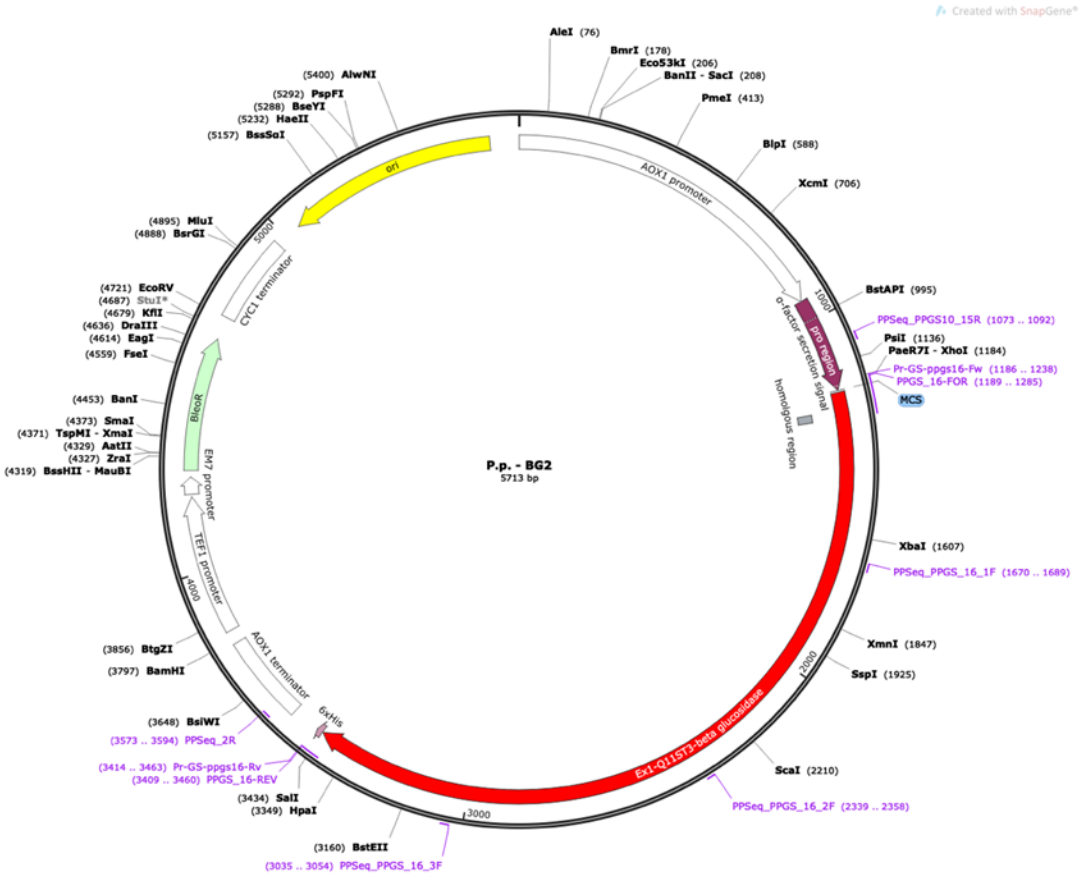

Figure S2.14: Plasmid map of plasmid containing BG2. Map of plasmid transformed into P. pastoris to create P.p. –BG2.

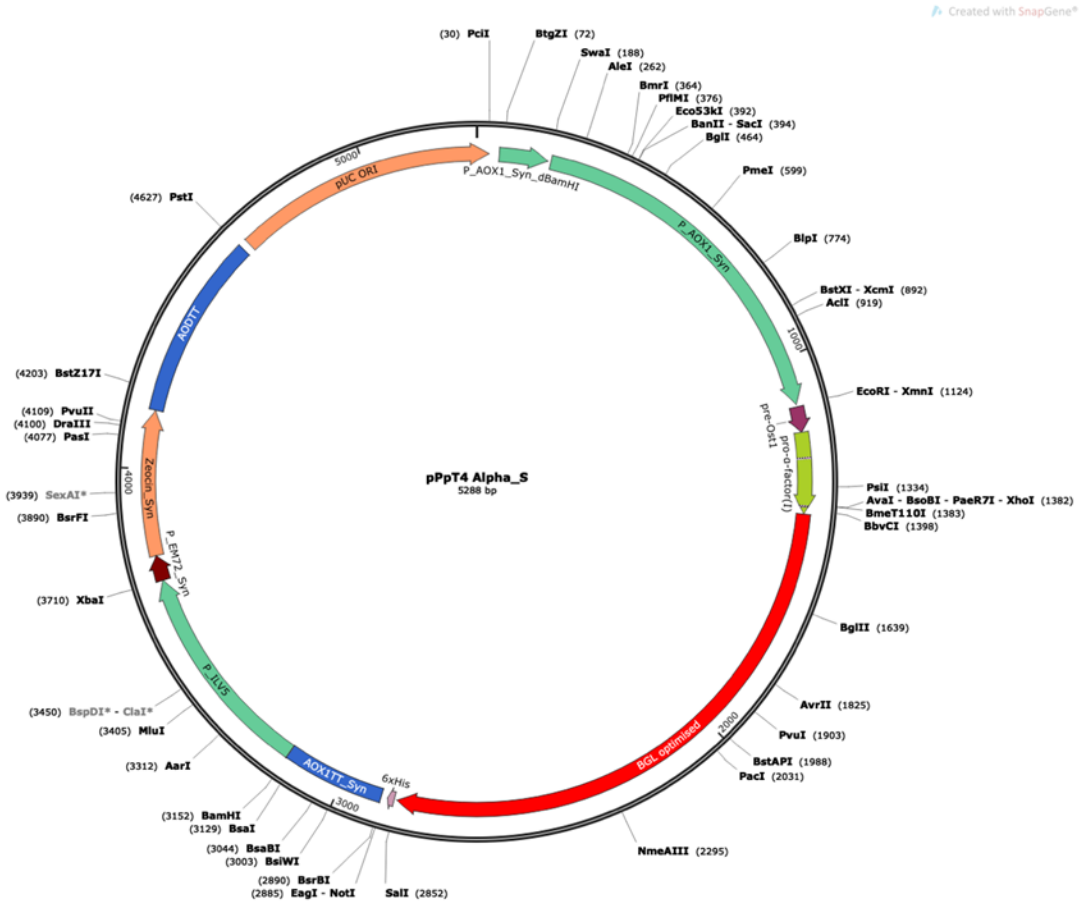

Figure S2.15: Plasmid map of plasmid containing PaBG1b+S_{POP}. Map of plasmid transformed into P. pastoris to create P.p. –PaBG1b-S_{POP}. pPpT4 Alpha_S was provided by Ingenza, Edinburgh.

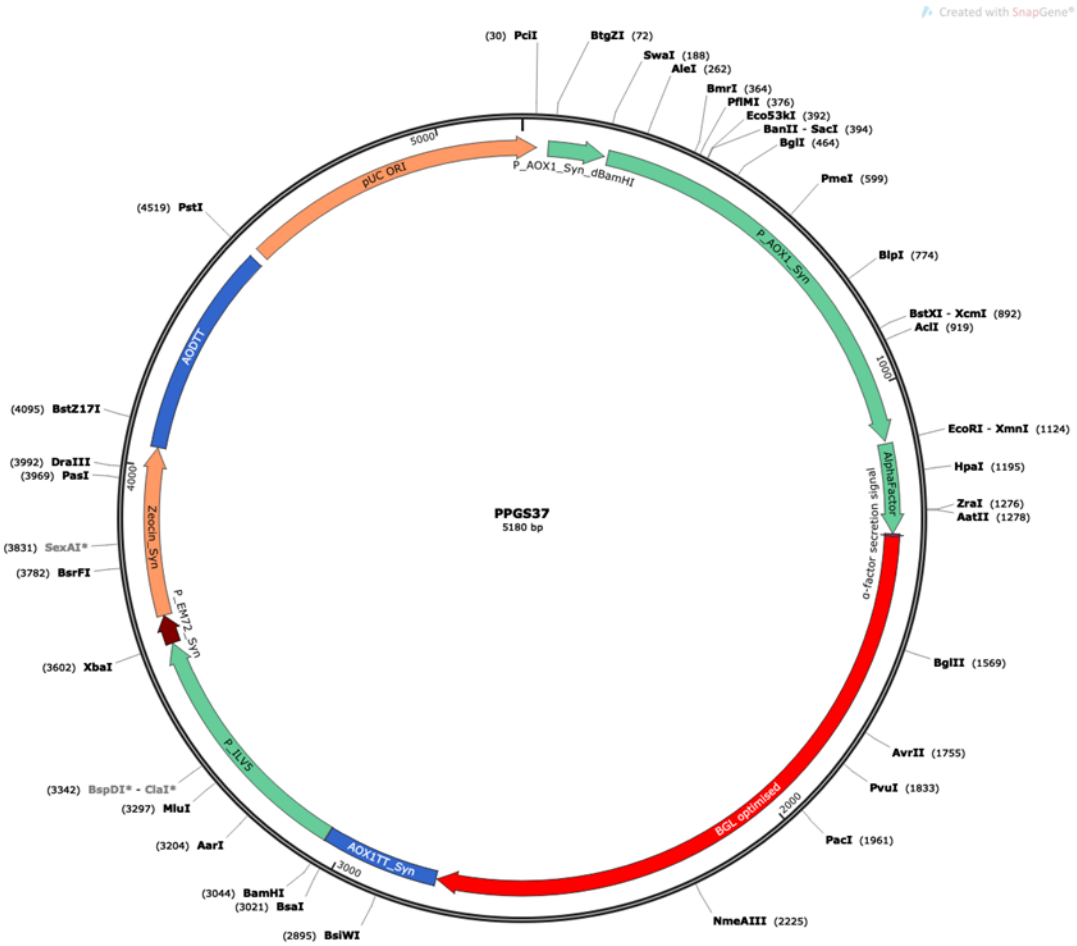

Figure S2.16: Plasmid map of plasmid containing PaBG1b+S$_{MAT57-70}$. Map of plasmid transformed into P. pastoris to create P.p. –PaBG1b-S $_{MAT57-70}$.

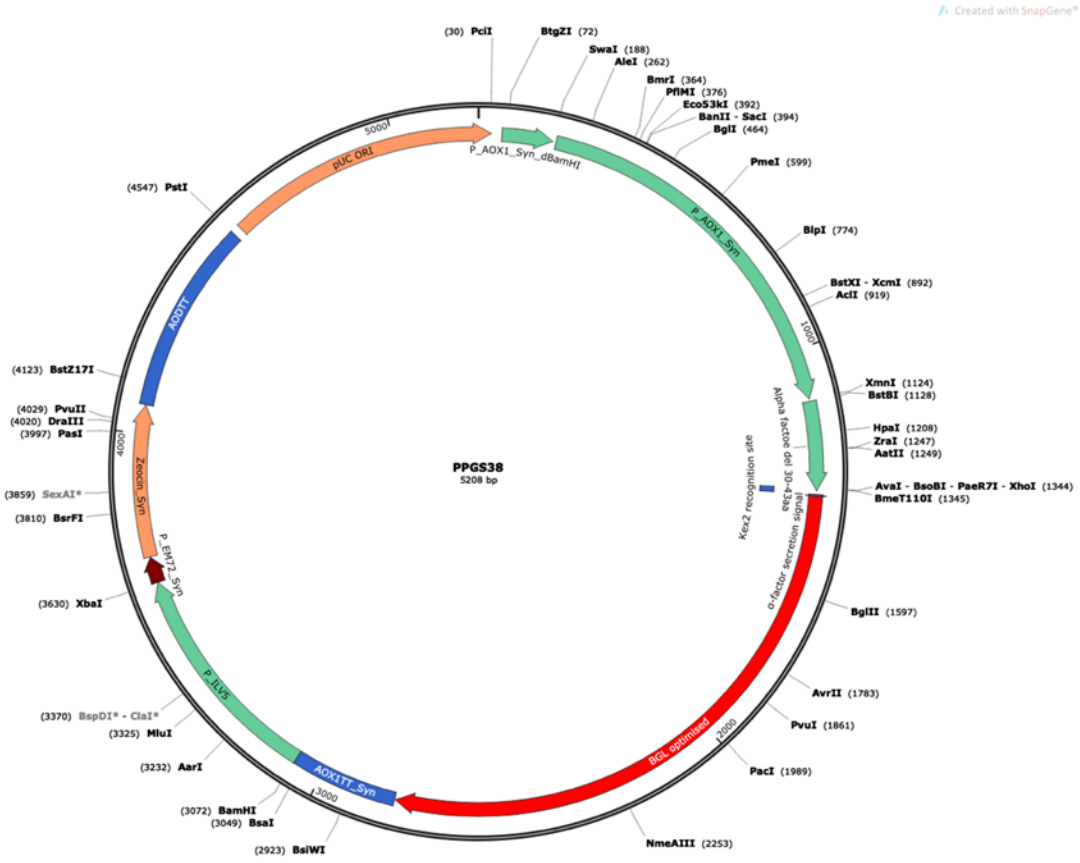

Figure S2.17: Plasmid map of plasmid containing PaBG1b+S$_{MAT30-43}$. Map of plasmid transformed into P. pastoris to create P.p. –PaBG1b-S $_{MAT30-43}$.

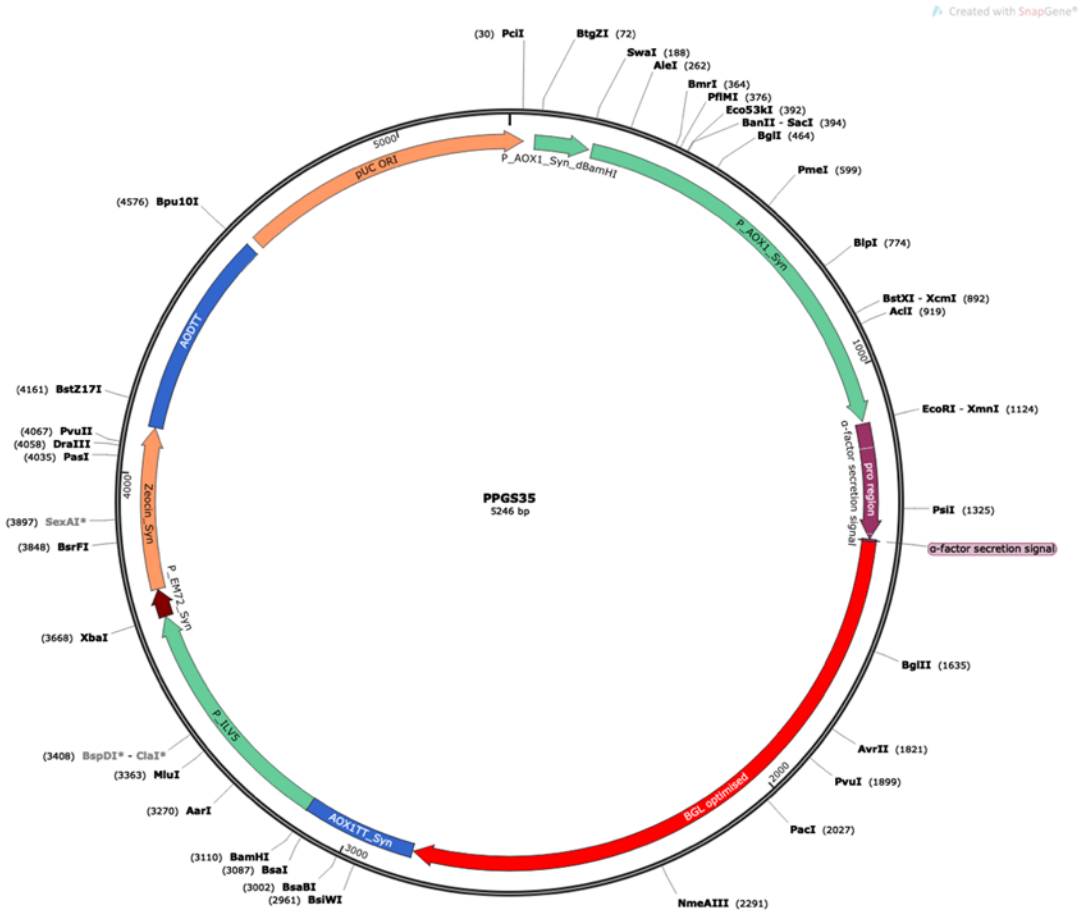

Figure S2.18: Plasmid map of plasmid containing PaBG1b+SMAT. Map of plasmid transformed into P. pastoris to create P.p. –PaBG1b-S MAT.

Additional Data

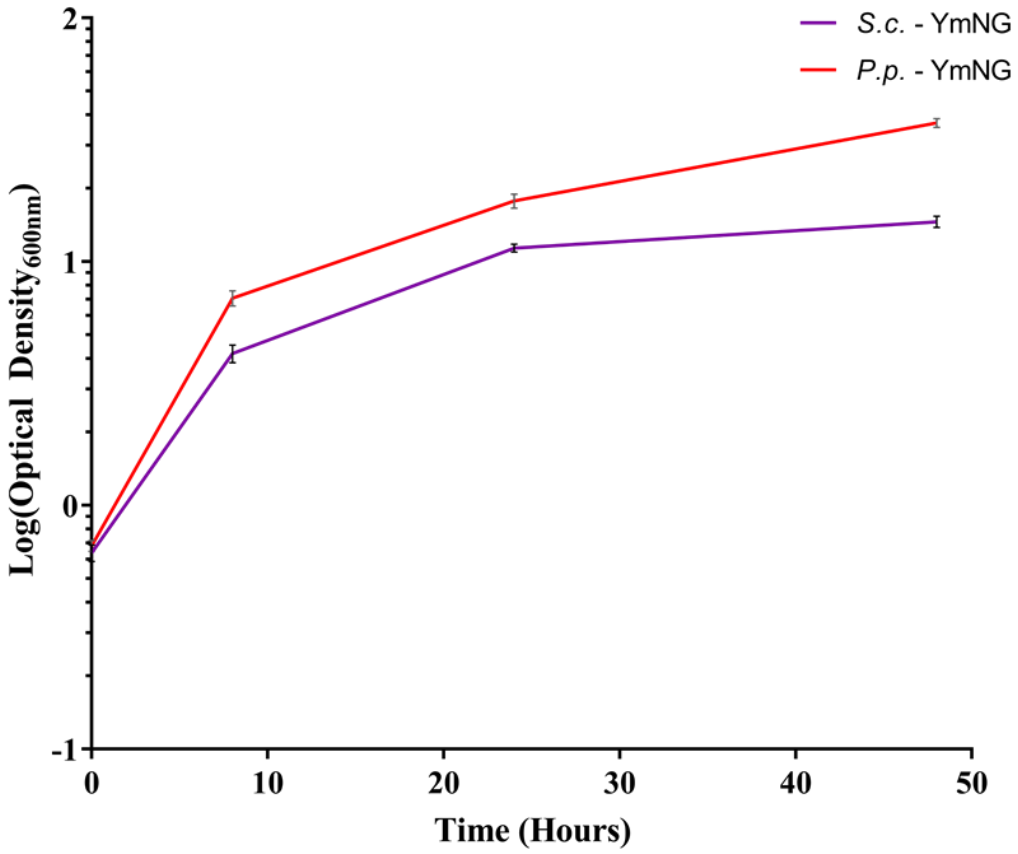

Figure S3.1: Growth comparison between S. cerevisiae and P. pastoris strains expressing ymNeonGreen. Growth of P. pastoris (P.p. – ymNG) and S. cerevisiae (S.c. – ymNG) expressing *mating alpha factor secretion signal fused to ymNeongreen when cultivated in BMGY and YEPD resectively at 30°C measured as the absorbance (optical density) 600 nm for 48 hours.*

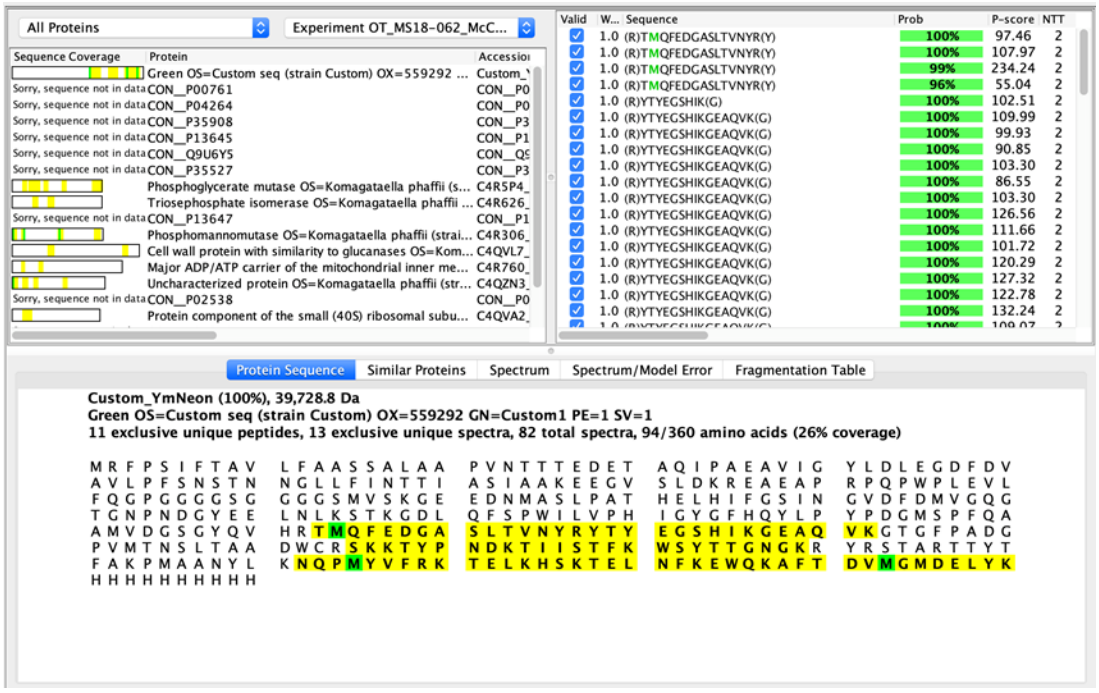

Figure S3.2: Protein mass spectrometry analysis of P.p. –YmNG bands.

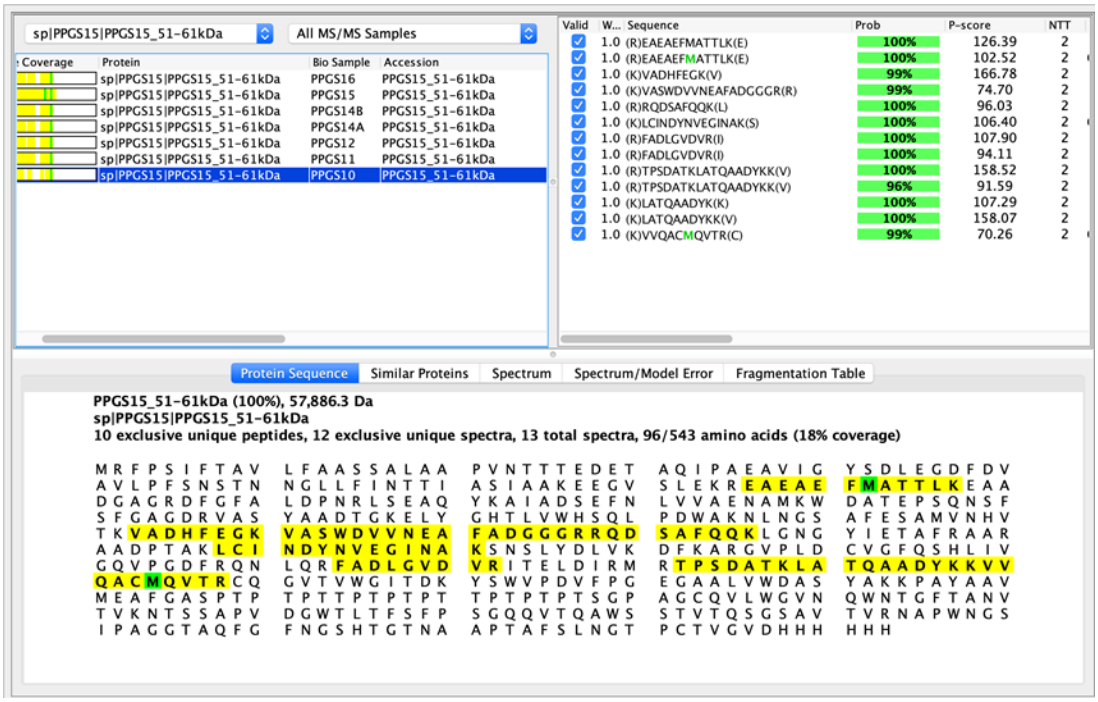

Figure S3.3: Protein mass spectrometry identification of PaBG1b.

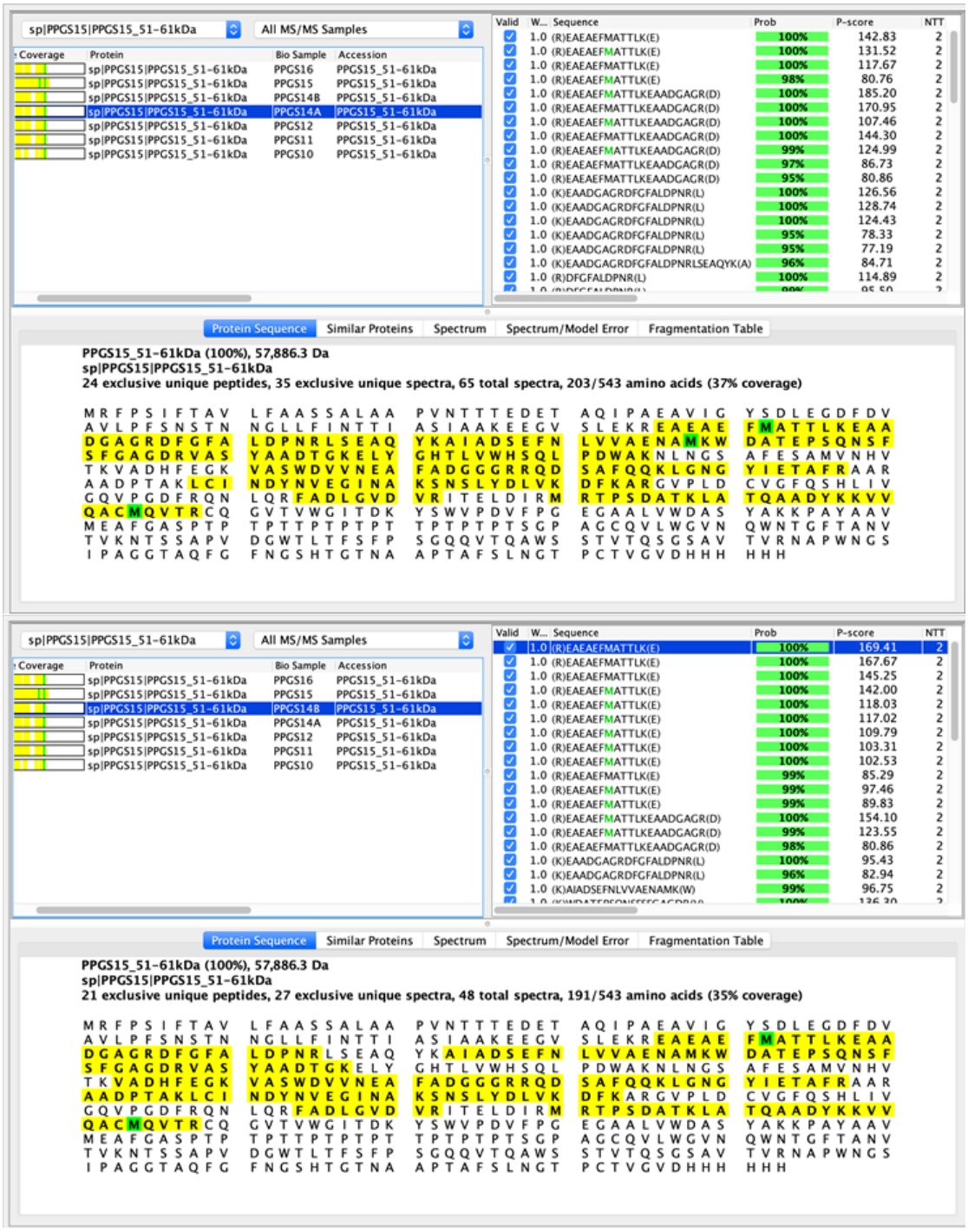

Figure S3.4: Protein mass spectrometry identification of ExG2.

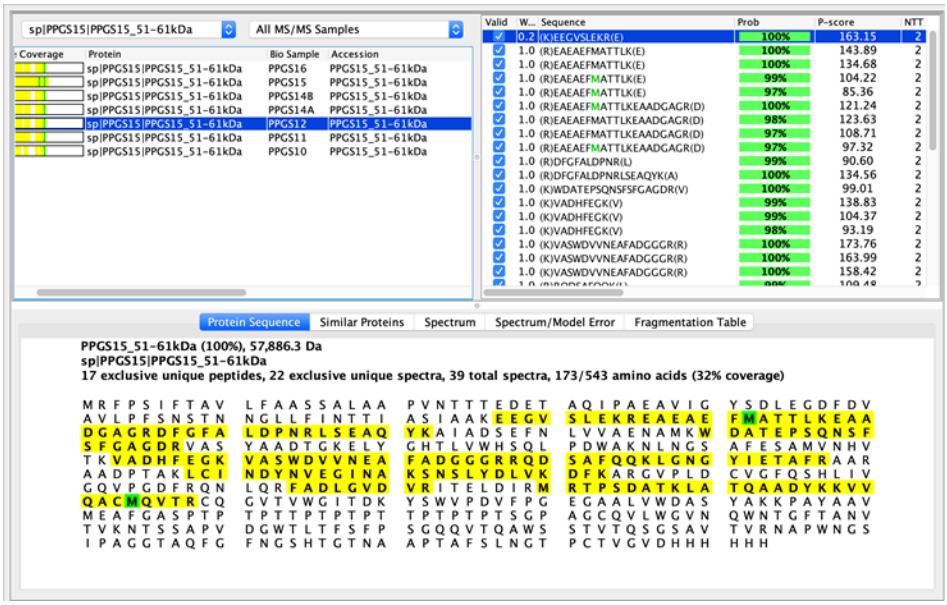

Figure S3.5: Protein mass spectrometry analysis of EnG1.

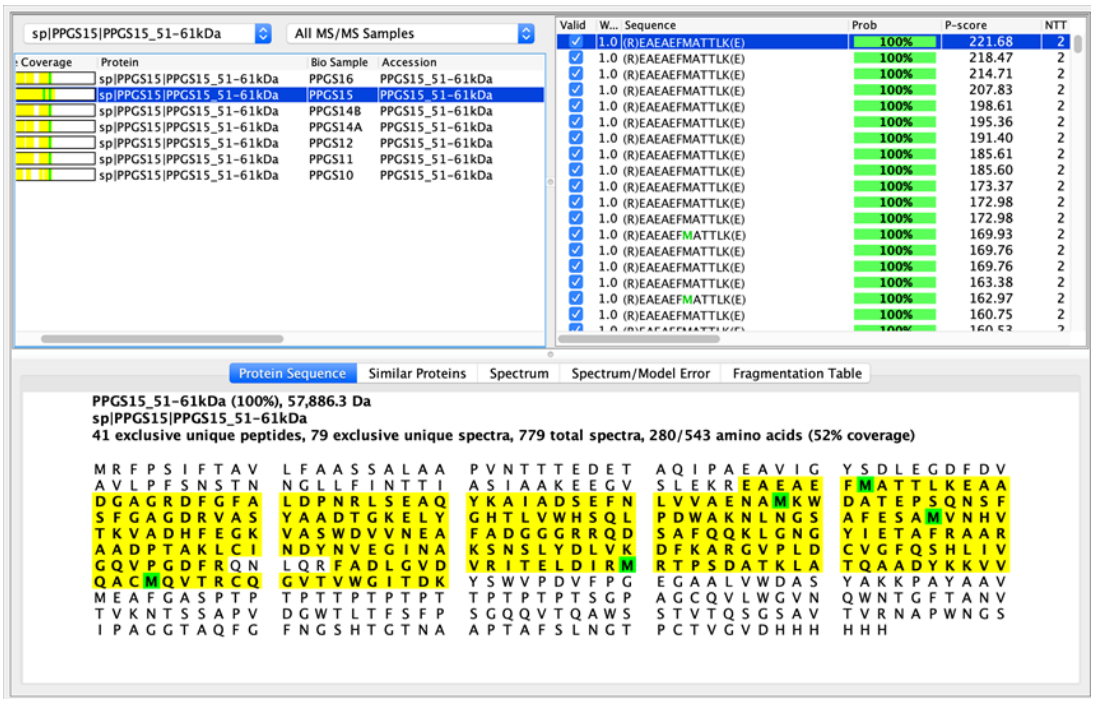

Figure S3.6: Protein mass spectrometry identification of BG1.

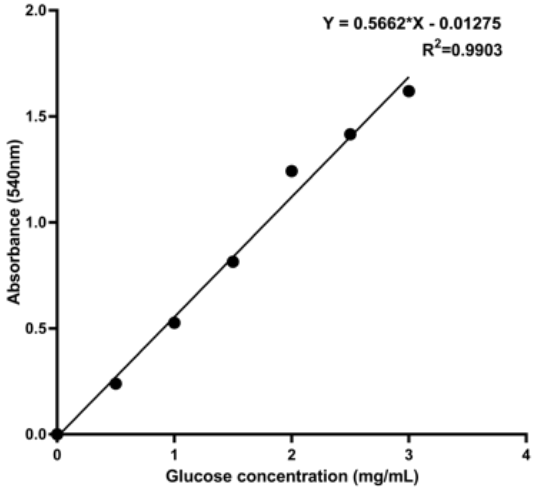

Figure S3.7: Standard curve for HK glucose assay kit.

Table S5. 1: Robotics script for automating plate reader assays

```
Heated Incubator1 SetTemperature(0,29)
1
Set Variable carousel = 1
2
Set Variable slot = 1
3
Begin Loop 4 times "transfer"
4
PresentPlate(1,~carousel~)
5
Transfer Labware Source: Grid '69,' Site '1'; Destination: Grid '50', Site 'slot'; Narrow
(ROMA 1)
6
Heated Incubator1 CloseDoor()
7
Set Variable slot = slot+1
8
Set Variable carousel = carousel+1
9
End Loop "transfer"
10
Comment How many times do you want it to cycle?
11
Begin Loop 200 times "plate reading"
12
Set Variable slot = 1
13
Begin Loop 4 times "transfer"
14
Heated Incubator1 StartShaker(4)
15
Start Timer 1
16
Wait for Timer Timer 1 : 1000 sec
17
Heated Incubator1 StopShaker()
18
Magellan Open()
19
Transfer Labware Source: Grid '50,' Site 'slot'; Destination: Grid '47', Site '1'; User defined (Narrow)
(ROMA 1)
20
Heated Incubator1 CloseDoor()
21
Magellan Close()
22
```

```
Heated Incubator1 StartShaker(4)
23
Comment At what plate number (slot) do you want a different reading?
24
If - Then slot > 7
25
Magellan Measure( C:\ProgramData\Tecan\EVOware\Output\<YYMMDD-NNN>, 0, 0)
26
Else
27
Magellan Measure(C:\ProgramData\Tecan\EVOware\Output\<YYMMDD-NNN>, 0,0)
28
End If
29
Magellan Open()
30
Heated Incubator1 StopShaker()
31
Transfer Labware Source: Grid '47,' Site '1'; Destination: Grid '50', Site 'slot'; User defined (Narrow)
(ROMA 1)
32
Heated Incubator1 CloseDoor()
33
Magellan Close()
34
Set Variable slot = slot+1
35
End Loop "transfer"
36
End Loop "plate reading"
37
Set Variable slot = 1
38
Set Variable carousel = 1
39
Begin Loop 4 times "Return plates to carousel"
40
Transfer Labware Source: Grid '50,' Site 'slot'; Destination: Grid '69', Site '1'; Narrow
(ROMA 1)
41
ReturnPlate(1,~carousel~)
42
Heated Incubator1 CloseDoor()
43
Set Variable slot = slot+1
44
Set Variable carousel = carousel+1
45
End Loop "Return plates to carousel"
46
```

Table S5.2: Overlapping regions for P_{GAP}, P_{G1} and P_{G4} and sequences in chapters 4 and 5 (sequences from 5' – 3',).

Fragment	**Sequence**	**Use**
5'overlap promoter cassette 1 P_{GAP}, P_{G1}, P_{G4}	ACTTCGTATAGCATACATTATACGAACGGTAAGC	Promoter to pPGS60 and pPGS80 backbone (*lox*66)
5'overlap promoter cassette 2 P_{GAP}, P_{G1}, P_{G4}	AGATTAAGTGAGACCTTCGTTTGTGC	Promoter for cassette 2 to cassette 1 *AOX1*TT
	CCTAGGGCGGCCGCTCAAGA	
3'overlap to gene		
3'overlap promoter cassette 1 P_{GAP}, P_{G1}, P_{G4} for mCherry	ATGGTAAGTAAAGGCGAAGAGGACAATATG	3' promoter region to mCherry
3'overlap promoter cassette 1 P_{GAP}, P_{G1}, P_{G4} for mCiterine	ATGTCAAAAGGAGAGGAAATGTTCACTG	3' promoter region to mCiterine
3'overlap promoter cassette 1 P_{GAP}, P_{G1}, P_{G4} for mTFP1	ATGGTCAGTAAAGGCGAAGAGACAAC	3' promoter region to mTFP1
3'overlap promoter cassette 1 P_{GAP}, P_{G1}, P_{G4} for TrEGI	ATGGCACCATCCGTTACATTGC	3' promoter region to TrEGI
3'overlap promoter cassette 1 P_{GAP}, P_{G1}, P_{G4} for ExG1	ATGGGGACGATCTCTGGCTCCCTTTAT	3' promoter region to ExG1

3'overlap promoter cassette 1 P_{GAP}, P_{G1}, P_{G4} for PaBG1b	CATGAGGAGCACTCAAGGACGA	3' promoter region to PaBG1b
3'overlap promoter cassette 1 P_{GAP}, P_{G1}, P_{G4} for CE1	ATGAGGCAGGTTTGGTTCTCTTGG	3' promoter region to CE1
3'overlap promoter cassette 1 P_{GAP}, P_{G1}, P_{G4} for HisCar5	ATGGCCGTTGATTCACCAGATGAACGATT	3' promoter region to HisCar5
3'overlap promoter cassette 1 P_{GAP}, P_{G1}, P_{G4} for CgAKR-1	ATGCCTAAGCAGCTGTCATCTGTTACCTTCC	3' promoter region to CgAKr-1

Table S5.3: Sanger sequences for promoter identification

Colony	Sequence	Promoter identification	Gene Regulated
S38	*CGTATAGCATACATTATACGAACGGTA AGCGCTTAAG*TTTTTGTAGAAATGTC TTGGTGTCCTCGACCAATCAGGTAG CCATCCCTGAAATACCTGGCTCCGT GGCAACACCGAACGACCTGCTGGCA ACGTTAAATTCTCCGGGGTAAAACT TAAATGTGGAGTAATAGAACCAGAA ACGTCTCTTCCCTTCTCTCCTTCC ACCGCCCGTTACCGTCCCTAGGAAA TTTTACTCTGCTGGAGAGCTTCTTCT ACGGCCCCCTTGCAGCAATGCTCTT CCCAGCATTACGTTGCGGGTAAAAC GGAGGTCGTGTACCCGACCTAGCAG CCCAGGGATGGAAAGTCCCGGCCGT CGCTGGCAATAACTGCGGGCGGACG	P_{G1} (underlined letters indicate unique bases for identified promoter)	TrEGI

	CATGTCTTGAGATTATTGGAAACCA CCAGAATCGAATATAAAAGGCGAA CACCTTTCCCAATTTTGGTTTCTCCT GACCCAAAGACTTTAAATTTAATTT ATTTGTCCCTATTTCAATCAATTGAA CAACTA		
S38	*AAGATTAAGTGAGACCTTCGTTTGTGC* TTTTTGTAGAAATGTCTTGGTGTCCT CGACCAATCAGGTAGCCATCCCTGA AATACCTGGCTCCGTGGCAACACCG AACGACCTGCTGGCAACGTTAAATT CTCCGGGGTAAAACTTAAATGTGGA GTAATAGAACCAGAAACGTCTCTTC CCTTCTCTCTCCTTCCACCGCCCGTT ACCGTCCCTAGGAAATTTTACTCTG CTGGAGAGCTTCTTCTACGGCCCCC TTGCAGCAATGCTCTTCCCAGCATT ACGTTGCGGGTAAAACGGAGGTCGT GTACCCGACCTAGCAGCCCAGGGAT GGAAAGTCCCGGCCGTCGCTGGCAA TAACTGCGGGCGGACGCATGTCTTG AGATTATTGGAAACCACCAGAATCG AATATAAAAGGCGAACACCTTTCCC AATTTTGGTTTCTCCTGACCCAAAG ACTTTAAATTTAATTTATTTGTCCCT ATTTCAATCAATTGAACAACTATCG A	P_{G1} (underlined letters indicate unique bases for identified promoter)	ExG2
S38	*CGTATAGCATACATTATACGAACGGTA* *AGCGCTTAAG*TTTTTGTAGAAATGTC TTGGTGTCCTCGACCAATCAGGTAG CCATCCCTGAAATACCTGGCTCCGT GGCAACACCGAACGACCTGCTGGCA ACGTTAAATTCTCCGGGGTAAAACT TAAATGTGGAGTAATAGAACCAGAA ACGTCTCTTCCCTTCTCTCCTTCC ACCGCCCGTTACCGTCCCTAGGAAA TTTTACTCTGCTGGAGAGCTTCTTCT ACGGCCCCCTTGCAGCAATGCTCTT CCCAGCATTACGTTGCGGGTAAAAC GGAGGTCGTGTACCCGACCTAGCAG CCCAGGGATGGAAAGTCCCGGCCGT CGCTGGCAATAACTGCGGGCGGACG CATGTCTTGAGATTATTGGAAACCA CCAGAATCGAATATAAAAGGCGAA CACCTTTCCCAATTTTGGTTTCTCCT GACCCAAAGACTTTAAATTTAATTT ATTTGTCCCTATTTCAATCAATTGAA CAACTATCGAA	P_{G1} (underlined letters indicate unique bases for identified promoter)	PaBG1b
S39	*CGTATAGCATACATTATACGAACGGTA* *AGCGCTTAAG*TTTTTGTAGAAATGTC	P_{G1} (underlined	TrEGI

	TTGGTGTCCTCGACCAATCAGGTAG CCATCCCTGAAATACCTGGCTCCGT GGCAACACCGAACGACCTGCTGGCA ACGTTAAATTCTCCGGGGTAAAACT TAAATGTGGAGTAATAGAACCAGAA ACGTCTCTTCCCTTCTCTCTCCTTCC ACCGCCCGTTACCGTCCCTAGGAAA TTTTACTCTGCTGGAGAGCTTCTTCT ACGGCCCCCTTGCAGCAATGCTCTT CCCAGCATTACGTTGCGGGTAAAAC GGAGGTCGTGTACCCGACCTAGCAG CCCAGGGATGGAAAGTCCCGGCCGT CGCTGGCAATAACTGCGGGCGGACG CATGTCTTGAGATTATTGGAAACCA CCAGAATCGAATATAAAAGGCGAA CACCTTTCCCAATTTTGGTTTCTCCT GACCCAAAGACTTTAAATTTAATTT ATTTGTCCCTATTTCAATCAATTGAA CA	letters indicate unique bases for identified promoter)	
S39	*AAGATTAAGTGAGACCTTCGTTTGTGC* TTTTTGTAGAAATGTCTTGGTGTCCT CGACCAATCAGGTAGCCATCCCTGA AATACCTGGCTCCGTGGCAACACCG AACGACCTGCTGGCAACGTTAAATT CTCCGGGGTAAAACTTAAATGTGGA GTAATAGAACCAGAAACGTCTCTTC CCTTCTCTCTCCTTCCACCGCCCGTT ACCGTCCCTAGGAAATTTTACTCTG CTGGAGAGCTTCTTCTACGGCCCCC TTGCAGCAATGCTCTTCCCAGCATT ACGTTGCGGGTAAAACGGAGGTCGT GTACCCGACCTAGCAGCCCAGGGAT GGAAAGTCCCGGCCGTCGCTGGCAA TAACTGCGGGCGGACGCATGTCTTG AGATTATTGGAAACCACCAGAATCG AATATAAAAGGCGAACACCTTTCCC AATTTTGGTTTCTCCTGACCCAAAG ACTTTAAATTTAATTTATTTGTCCCT ATTTCAATCAATTGAACAACTATCG AA	P_{G1} (underlined letters indicate unique bases for identified promoter)	ExG2
S39	*CGTATAGCATACATTATACGAACGGTA AGCGCTTAAG*TTTTTGTAGAAATGTC TTGGTGTCCTCGTCCAATCAGGTAG CCATCTCTGAAATATCTGGCTCCGTT GCAACTCCGAACGACCTGCTGGCAA CGTAAAATTCTCCGGGGTAAAACTT AAATGTGGAGTAATGGAACCAGAA ACGTCTCTTCCCTTCTCTCTCCTTCC ACCGCCCGTTACCGTCCCTAGGAAA TTTTACTCTGCTGGAGAGCTTCTTCT	P_{GAP} (underlined letters indicate unique bases for identified promoter)	PaBG1b

	ACGGCCCCCTTGCAGCAATGCTCTT CCCAGCATTACGTTGCGGGTAAAAC GGAGGTCGTGTACCCGACCTAGCAG CCCAGGGATGGAAAAGTCCCGGCCG TCGCTGGCAATAATAGCGGGCGGAC GCATGTCATGAGATTATTGGAAACC ACCAGAATCGAATATAAAAGGCGA ACACCTTTCCCAATTTTGGTTTCTCC TGACCCAAAGACTTTAAATTTAATT TATTTGTCCCTATTTCAATCAATTGA ACAACT		
S40	*CGTATAGCATACATTATACGAACGGTA AGCGCTTAAG*TTTTTGTAGAAATGTC TTGGTGTCCTCGACCAATCAGGTAG CCATCCCTGAAATACCTGGCTCCGT GGCAACACCGAACGACCTGCTGGCA ACGTTAAATTCTCCGGGGTAAAACT TAAATGTGGAGTAATAGAACCAGAA ACGTCTCTTCCCTTCTCTCCTTCC ACCGCCCGTTACCGTCCCTAGGAAA TTTTACTCTGCTGGAGAGCTTCTTCT ACGGCCCCCTTGCAGCAATGCTCTT CCCAGCATTACGTTGCGGGTAAAAC GGAGGTCGTGTACCCGACCTAGCAG CCCAGGGATGGAAAGTCCCGGCCGT CGCTGGCAATAACTGCGGGCGGACG CATGTCTTGAGATTATTGGAAACCA CCAGAATCGAATATAAAAGGCGAA CACCTTTCCCAATTTTGGTTTCTCCT GACCCAAAGACTTTAAATTTAATTT ATTTGTCCCTATTTCAATCAATTGAA CA	P_{G1} (underlined letters indicate unique bases for identified promoter)	TrEGI
S40	*AAGATTAAGTGAGACCTTCGTTTGTGC* TTTTTGTAGAAATGTCTTGGTGTCCT CGACCAATCAGGTAGCCATCCCTGA AATACCTGGCTCCGTGGCAACACCG AACGACCTGCTGGCAACGTTAAATT CTCCGGGGTAAAACTTAAATGTGGA GTAATAGAACCAGAAACGTCTCTTC CCTTCTCTCTCCTTCCACCGCCCGTT ACCGTCCCTAGGAAATTTTACTCTG CTGGAGAGCTTCTTCTACGGCCCCC TTGCAGCAATGCTCTTCCCAGCATT ACGTTGCGGGTAAAACGGAGGTCGT GTACCCGACCTAGCAGCCCAGGGAT GGAAAGTCCCGGCCGTCGCTGGCAA TAACTGCGGGCGGACGCATGTCTTG AGATTATTGGAAACCACCAGAATCG AATATAAAAGGCGAACACCTTTCCC AATTTTGGTTTCTCCTGACCCAAAG	P_{G1} (underlined letters indicate unique bases for identified promoter)	ExG2

	ACTTTAAATTTAATTTATTTGTCCCT ATTTCAATCAATTGAACAACTATCG AA		
S40	*CGTATAGCATACATTATACGAACGGTA AGCGCTTAAG*TTTTTGTAGAAATGTC TTGGTGTCCTCGACCAATCAGGTAG CCATCCCTGAAATACCTGGCTCCGT GGCAACACCGAACGACCTGCTGGCA ACGTTAAATTCTCCGGGGTAAAACT TAAATGTGGAGTAATAGAACCAGAA ACGTCTCTTCCCTTCTCTCCTTCC ACCGCCCGTTACCGTCCCTAGGAAA TTTTACTCTGCTGGAGAGCTTCTTCT ACGGCCCCCTTGCAGCAATGCTCTT CCCAGCATTACGTTGCGGGTAAAAC GGAGGTCGTGTACCCGACCTAGCAG CCCAGGGATGGAAAGTCCCGGCCGT CGCTGGCAATAACTGCGGGCGGACG CATGTCTTGAGATTATTGGAAACCA CCAGAATCGAATATAAAAGGCGAA CACCTTTCCCAATTTTGGTTTCTCCT GACCCAAAGACTTTAAATTTAATTT ATTTGTCCCTATTTCAATCAATTGAA CAACTATCGAA	P_{GAP} (underlined letters indicate unique bases for identified promoter)	PaBG1b
S7, S8, S9	*CGTATAGCATACATTATACGAACGGTA AGCGCTTAAG*TTTTTGTAGAAATGTC TTGGTGTCCTCGACCAATCAGGTAG CCATCCCTGAAATACCTGGCTCCGT GGCAACACCGAACGACCTGCTGGCA ACGTTAAATTCTCCGGGGTAAAACT TAAATGTGGAGTAATAGAACCAGAA ACGTCTCTTCCCTTCTCTCCTTCC ACCGCCCGTTACCGTCCCTAGGAAA TTTTACTCTGCTGGAGAGCTTCTTCT ACGGCCCCCTTGCAGCAATGCTCTT CCCAGCATTACGTTGCGGGTAAAAC GGAGGTCGTGTACCCGACCTAGCAG CCCAGGGATGGAAAGTCCCGGCCGT CGCTGGCAATAACTGCGGGCGGACG CATGTCTTGAGATTATTGGAAACCA CCAGAATCGAATATAAAAGGCGAA CACCTTTCCCAATTTTGGTTTCTCCT GACCCAAAGACTTTAAATTTAATTT ATTTGTCCCTATTTCAATCAATTGAA CAACTAT	P_{GAP} (underlined letters indicate unique bases for identified promoter)	TrEGI
S7, S8, S9	*AAGATTAAGTGAGACCTTCGTTTGTGC* TTTTTGTAGAAATGTCTTGGTGTCCT CGACCAATCAGGTAGCCATCCCTGA AATACCTGGCTCCGTGGCAACACCG AACGACCTGCTGGCAACGTTAAATT	P_{GAP} (underlined letters indicate unique bases	

	CTCCGGGGTAAAACTTAAATGTGGA GTAATAGAACCAGAAACGTCTCTTC CCTTCTCTCTCCTTCCACCGCCCGTT ACCGTCCCTAGGAAATTTTACTCTG CTGGAGAGCTTCTTCTACGGCCCCC TTGCAGCAATGCTCTTCCCAGCATT ACGTTGCGGGTAAAACGGAGGTCGT GTACCCGACCTAGCAGCCCAGGGAT GGAAAGTCCCGGCCGTCGCTGGCAA TAACTGCGGGCGGACGCATGTCTTG AGATTATTGGAAACCACCAGAATCG AATATAAAAGGCGAACACCTTTCCC AATTTTGGTTTCTCCTGACCCAAAG ACTTTAAATTTAATTTATTTGTCCCT ATTTCAATCAATTGAACAACTAT	for identified promoter)	
S7, S8, S9	*CGTATAGCATACATTATACGAACGGTA* *AGCGCTTAAG*TTTTTGTAGAAATGTC TTGGTGTCCTCGACCAATCAGGTAG CCATCCCTGAAATACCTGGCTCCGT GGCAACACCGAACGACCTGCTGGCA ACGTTAAATTCTCCGGGGTAAAACT TAAATGTGGAGTAATAGAACCAGAA ACGTCTCTTCCCTTCTCTCTCCTTCC ACCGCCCGTTACCGTCCCTAGGAAA TTTTACTCTGCTGGAGAGCTTCTTCT ACGGCCCCCTTGCAGCAATGCTCTT CCCAGCATTACGTTGCGGGTAAAAC GGAGGTCGTGTACCCGACCTAGCAG CCCAGGGATGGAAAGTCCCGGCCGT CGCTGGCAATAACTGCGGGCGGACG CATGTCTTGAGATTATTGGAAACCA CCAGAATCGAATATAAAAGGCGAA CACCTTTCCCAATTTTGGTTTCTCCT GACCCAAAGACTTTAAATTTAATTT ATTTGTCCCTATTTCAATCAATTGAA CAACTAT	P_{GAP} (underlined letters indicate unique bases for identified promoter)	PaBG1b

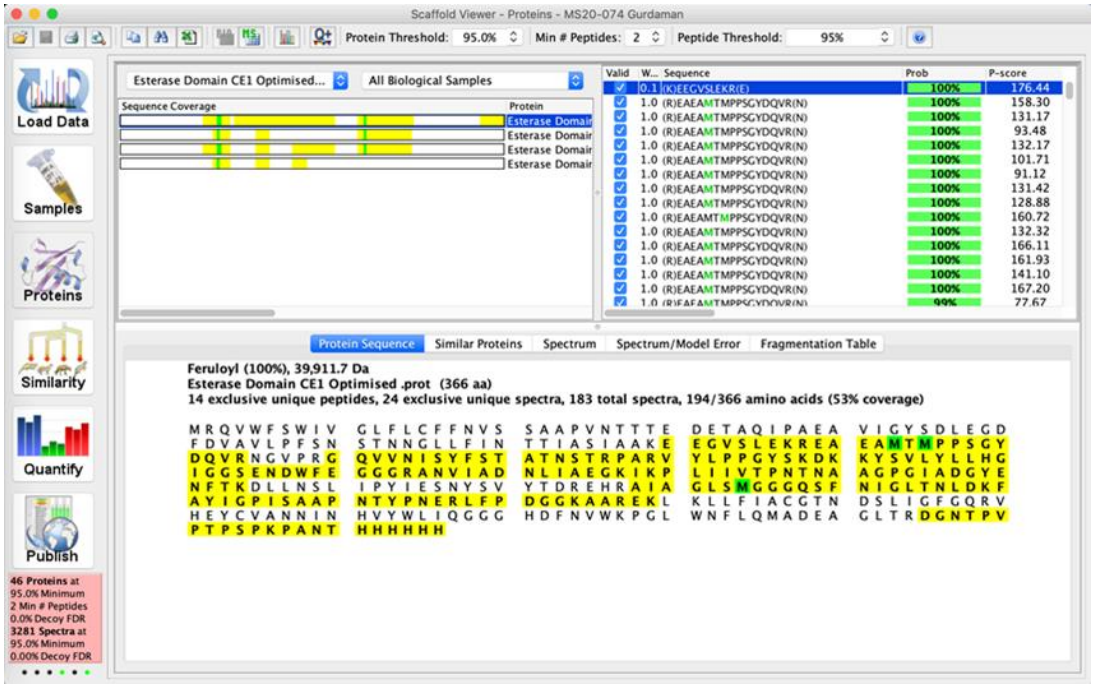

Figure S5.1: Protein mass spectrometry identification of HisCar5.

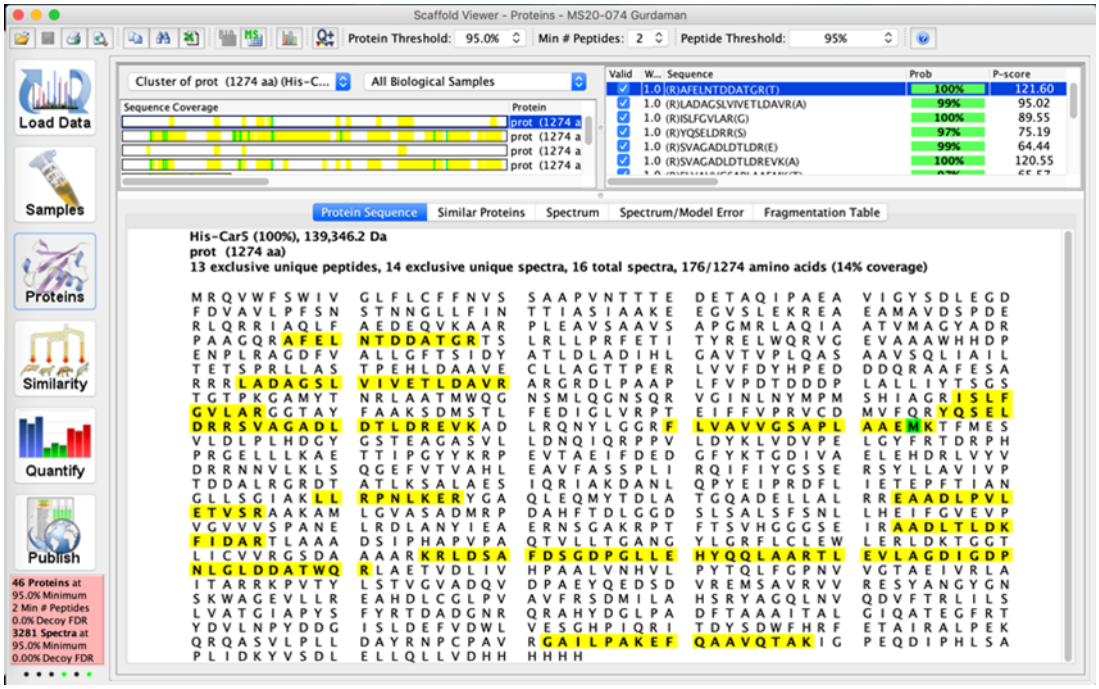

Figure S5.2: Protein mass spectrometry identification of CE1.

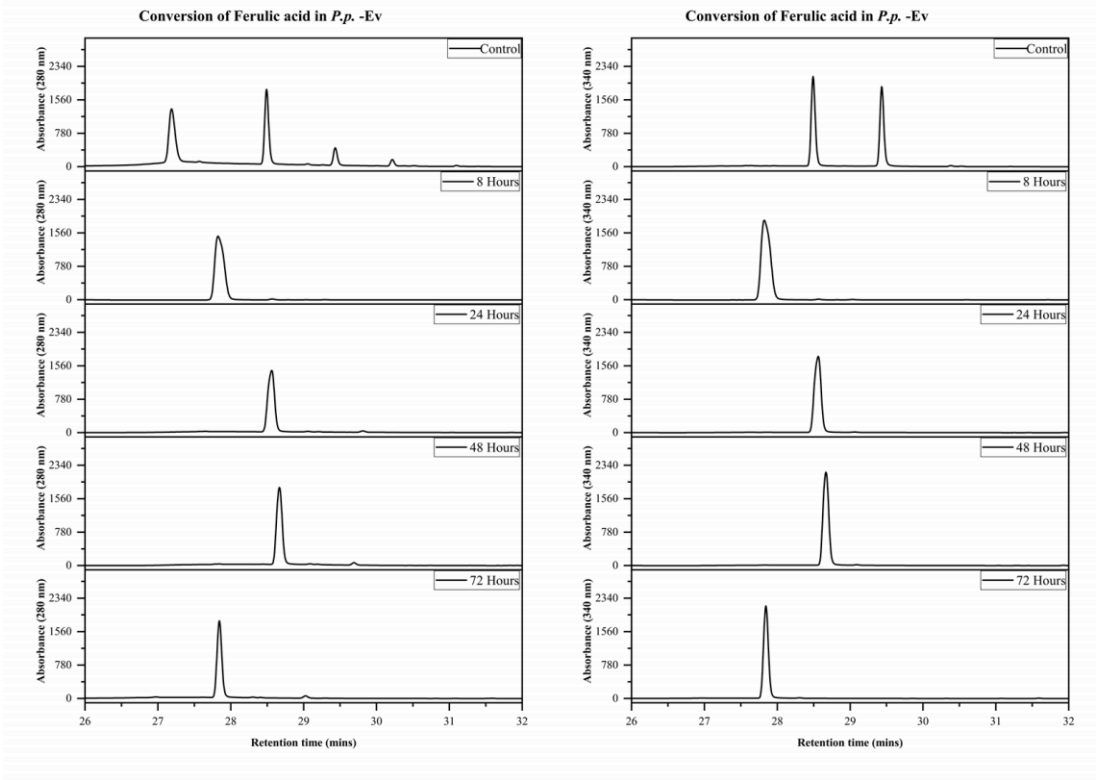

Figure S5.3: HPLC chromatogram of conversion of ferulic acid in P.p. – Ev at absorption 280 and 340 nm. The strain P.p. -Ev was grown for 24 hours in BMGY. The compound ferulic acid was then added at a final concentration of 1mM. Samples were taken and metabolites extraction carried out using acetonitrile at time points of 8 hours (B) 24 hours (C) 48 hours (D) and 72 hours (E). The absorbance was measured at 280 nm (left) and 340 nm (right). The controls (A) were 0.33mM final concentration of coniferyl alcohol, ferulic acid and coniferyl aldehyde in 50 % BGGY and 50 % acetonitrile.

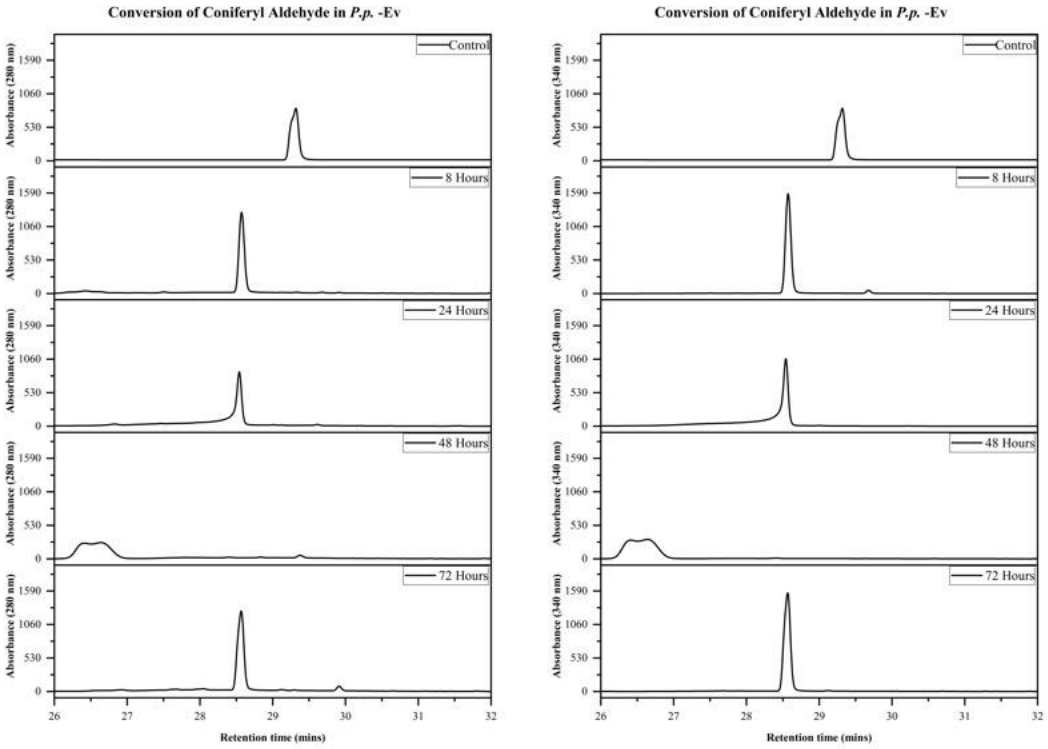

Figure S5.4: HPLC chromatogram of conversion of coniferyl aldehyde in P.p. – Ev. The strain P.p. -Ev were grown for 24 hours in BMGY. Coniferyl aldehyde was then added at a final concentration of 1mM. Metabolite extraction was carried out using acetonitrile at time points of 24, 48 and 72 hours. The absorbance was measured at 280 (left) and 340 (right) nm. The controls (A) were 1mM final concentration of coniferyl aldehyde in 50 % BMGY and 50 % acetonitrile.

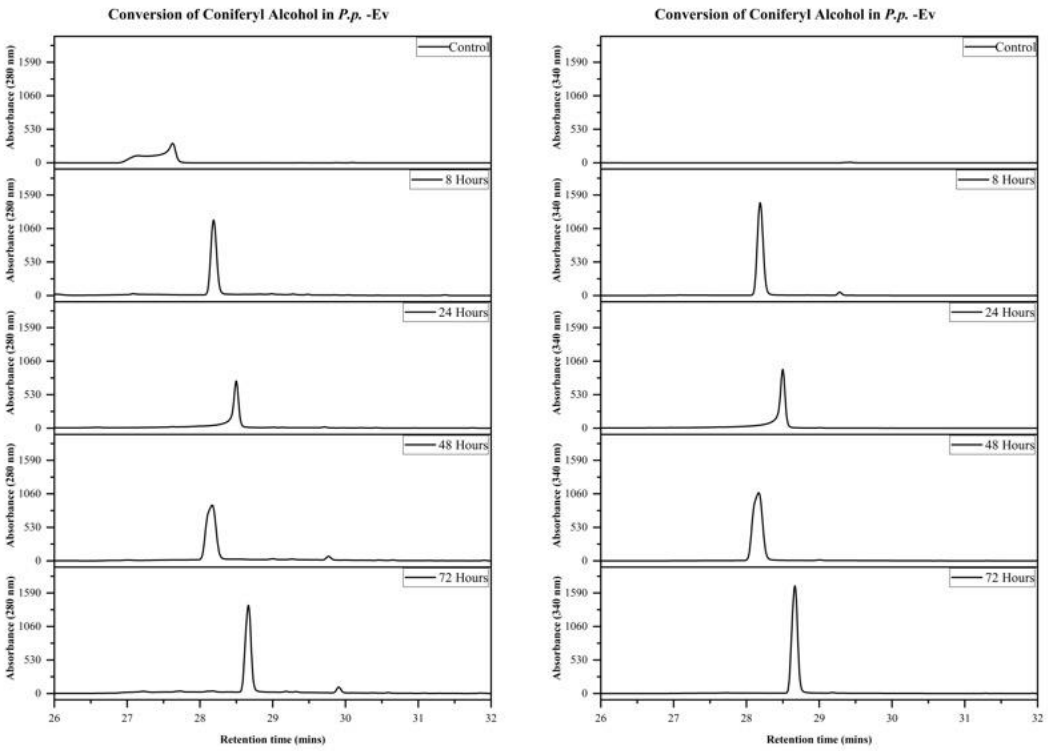

Figure S5.5: HPLC chromatogram of degradation of coniferyl alcohol in P.p. – Ev 0-72 hours. The strain P.p. - Ev was grown for 24 hours in BMGY. Coniferyl alcohol was then added at a final concentration of 1mM. Metabolite extraction was carried out using acetonitrile at time points of 24, 48 and 72 hours. The absorbance was measured at 280 (left) and 340 (right) nm. The controls (A) were 1mM final concentration of coniferyl aldehyde in 50 % BMGY and 50 % acetonitrile.

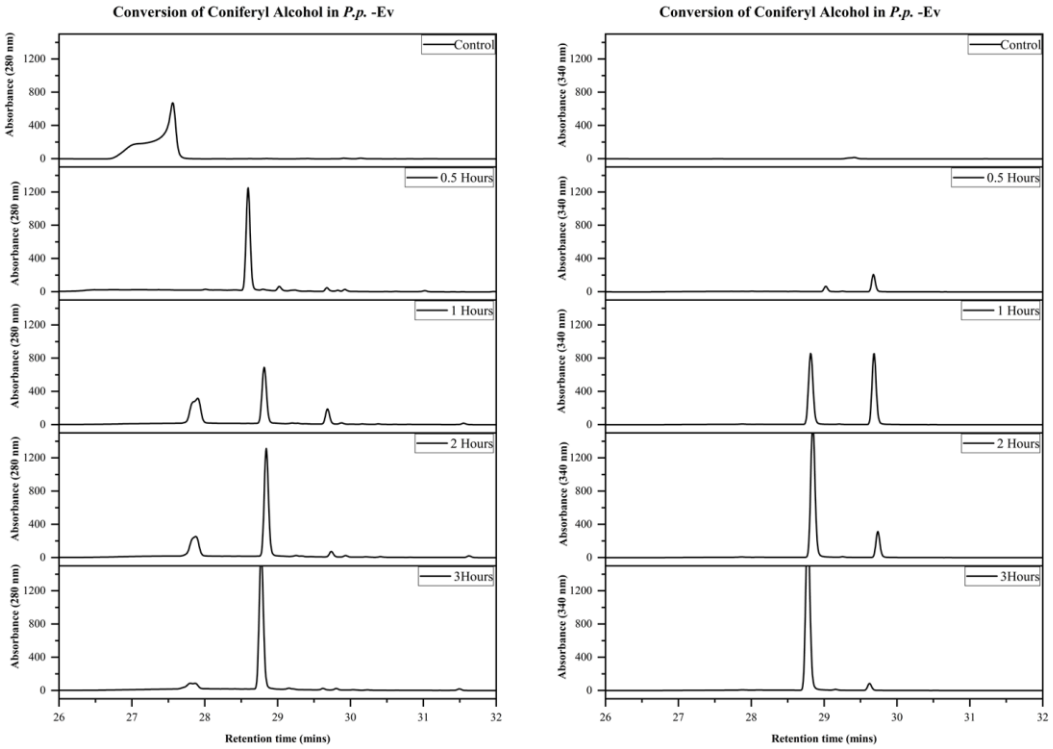

Figure S5.6: HPLC chromatogram of degradation of coniferyl alcohol in P.p. – Ev 0-3 hours. The strain P.p. -Ev were grown for 24 hours in BMGY. Coniferyl alcohol was then added at a final concentration of 1mM. Metabolite extraction was carried out using acetonitrile at time points of 0.5, 1, 2 and 3 hours. The absorbance was measured at 280 (left) and 340 (right) nm. The controls (A) were 1mM final concentration of coniferyl alcohol in 50 % BMGY and 50 % acetonitrile.

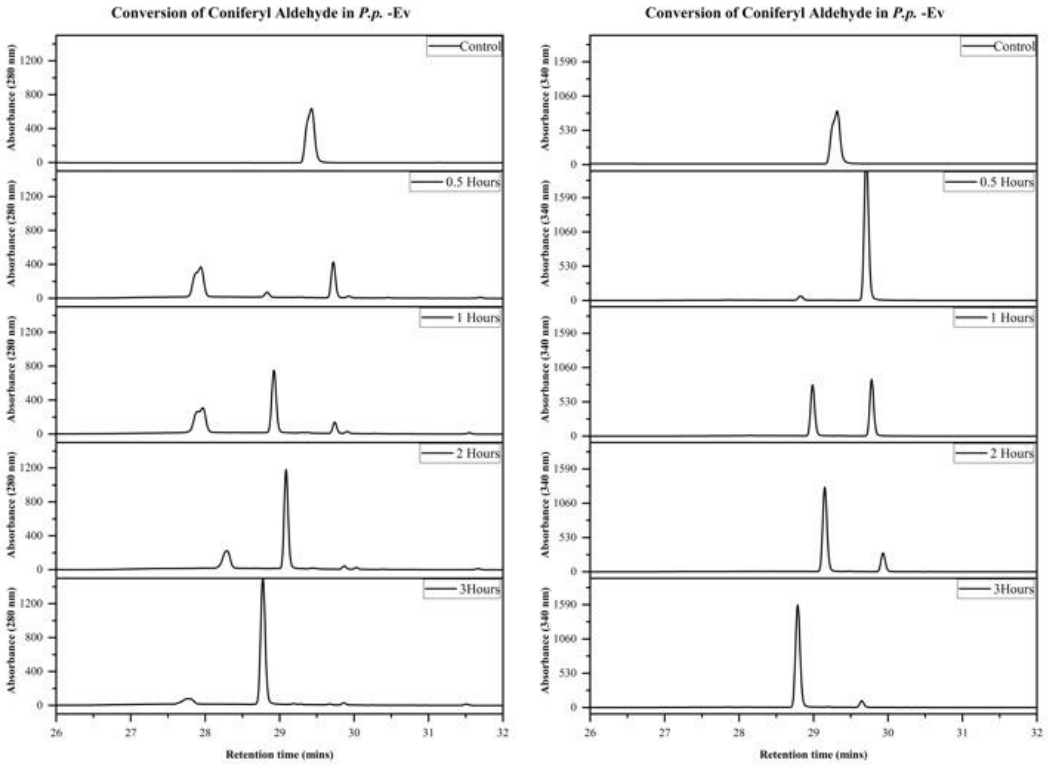

Figure S5.7: HPLC chromatogram of degradation of coniferyl aldehyde in P.p. – Ev 0-3 hours. The strain P.p. - Ev were grown for 24 hours in BMGY. Coniferyl aldehyde was then added at a final concentration of 1mM. Metabolite extraction was carried out using acetonitrile at time points of 0.5, 1, 2 and 3 hours. The absorbance was measured at 280 (left) and 340 (right) nm. The controls (A) were 1mM final concentration of coniferyl aldehyde in 50 % BMGY and 50 % acetonitrile.

Protein Sequences (5'-3') from chapter 5:

>TrEGI

MAPSVTLPLTTAILAIARLVAAQQPGTSTPEVHPKLTTYKCTKSGGCVAQDT
SVVLDWNYRWMHDANYNSCTVNGGVNTTLCPDEATCGKNCFIEGVDYAA
SGVTTSGSSLTMNQYMPSSSGGYSSVSPRLYLLDSDGEYVMLKLNGQELSF
DVDLSALPCGENGSLYLSQMDENGGANQYNTAGANYGSGYCDAQCPVQT
WRNGTLNTSHQGFCCNEMDILEGNSRANALTPHSCTATACDSAGCGFNPYG
SGYKSYYGPGDTVDTSKTFTIITQFNTDNGSPSGNLVSITRKYQQNGVDIPSA
QPGGDTISSCPSASAYGGLATMGKALSSGMVLVFSIWNDNSQYMNWLDSG
NAGPCSSTEGNPSNILANNPNTHVVFSNIRWGDIGSTTNSTAPPPPPASSTTFS
TTRRSSTTSSSPSCTQTHWGQCGGIGYSGCKTCTSGTTCQYSNDYYSQCL*

>mCherry:
MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKL
KVTKGGPLPFAWDILSPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVM
NFEDGGVVTVTQDSSLQDGEFIYKVKLRGTNFPSDGPVMQKKTMGWEASS
ERMYPEDGALKGEIKQRLKLKDGGHYDAEVKTTYKAKKPVQLPGAYNVNI
KLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK*

>mCitrine:
MSKGEEMFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTG
KLPVPWPTLVTTFGYGLMCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKD
DGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIM
ADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSYQS
KLSKDPNEKRDHMVLLEFVTAAGITLGMDELYK*

>mTFP1:
MVSKGEETTMGVIKPDMKIKLKMEGNVNGHAFVIEGEGEGKPYDGTNTINL
EVKEGAPLPFSYDILTTAFAYGNRAFTKYPDDIPNYFKQSFPEGYSWERTMT
FEDKGIVKVKSDISMEEDSFIYEIHLKGENFPPNGPVMQKKTTGWDASTERM
YVRDGVLKGDVKHKLLLEGGGHHRVDFKTIYRAKKAVKLPDYHFVDHRIEI
LNHDKDYNKVTVYESAVARNSTDGMDELYK*

HisCar5:
MAVDSPDERLQRRIAQLFAEDEQVKAARPLEAVSAAVSAPGMRLAQIAATV
MAGYADRPAAGQRAFELNTDDATGRTSLRLLPRFETITYRELWQRVGEVAA
AWHHDPENPLRAGDFVALLGFTSIDYATLDLADIHLGAVTVPLQASAAVSQ
LIAILTETSPRLLASTPEHLDAAVECLLAGTTPERLVVFDYHPEDDDQRAAFE
SARRRLADAGSLVIVETLDAVRARGRDLPAAPLFVPDTDDDPLALLIYTSGS
TGTPKGAMYTNRLAATMWQGNSMLQGNSQRVGINLNYMPMSHIAGRISLF
GVLARGGTAYFAAKSDMSTLFEDIGLVRPTEIFFVPRVCDMVFQRYQSELDR
RSVAGADLDTLDREVKADLRQNYLGGRFLVAVVGSAPLAAEMKTFMESVL
DLPLHDGYGSTEAGASVLLDNQIQRPPVLDYKLVDVPELGYFRTDRPHPRGE
LLLKAETTIPGYYKRPEVTAEIFDEDGFYKTGDIVAELEHDRLVYVDRRNNV

LKLSQGEFVTVAHLEAVFASSPLIRQIFIYGSSERSYLLAVIVPTDDALRGRDT
ATLKSALAESIQRIAKDANLQPYEIPRDFLIETEPFTIANGLLSGIAKLLRPNLK
ERYGAQLEQMYTDLATGQADELLALRREAADLPVLETVSRAAKAMLGVAS
ADMRPDAHFTDLGGDSLSALSFSNLLHEIFGVEVPVGVVVSPANELRDLAN
YIEAERNSGAKRPTFTSVHGGGSEIRAADLTLDKFIDARTLAAADSIPHAPVP
AQTVLLTGANGYLGRFLCLEWLERLDKTGGTLICVVRGSDAAAARKRLDSA
FDSGDPGLLEHYQQLAARTLEVLAGDIGDPNLGLDDATWQRLAETVDLIVH
PAALVNHVLPYTQLFGPNVVGTAEIVRLAITARRKPVTYLSTVGVADQVDP
AEYQEDSDVREMSAVRVVRESYANGYGNSKWAGEVLLREAHDLCGLPVA
VFRSDMILAHSRYAGQLNVQDVFTRLILSLVATGIAPYSFYRTDADGNRQRA
HYDGLPADFTAAAITALGIQATEGFRTYDVLNPYDDGISLDEFVDWLVESGH
PIQRITDYSDWFHRFETAIRALPEKQRQASVLPLLDAYRNPCPAVRGAILPAK
EFQAAVQTAKIGPEQDIPHLSAPLIDKYVSDLELLQLLVD*

>CE1
MTMPPSGYDQVRNGVPRGQVVNISYFSTATNSTRPARVYLPPGYSKDKKYS
VLYLLHGIGGSENDWFEGGGRANVIADNLIAEGKIKPLIIVTPNTNAAGPGIA
DGYENFTKDLLNSLIPYIESNYSVYTDREHRAIAGLSMGGGQSFNIGLTNLDK
FAYIGPISAAPNTYPNERLFPDGGKAAREKLKLLFIACGTNDSLIGFGQRVHE
YCVANNINHVYWLIQGGGHDFNVWKPGLWNFLQMADEAGLTRDGNTPVP
TPSPKPANT

>CgAKr-1
MPKQLSSVTFHNGRKMPVVGLGTWQSPPEEVTAAIDVALEVGYRHIDTAFM
YQNEAAIGKTLKKWFDSGKLKREDVFIVTKLPPIGNRAESVEKFLTKSLEAL
QLDYVDLYLIHLPVGFQYKGDDNLWPRGEAGEFLIDTSTDLISLWKAMEAQ
VDAGRTRSVGLSNFNSRQIARIVKSARIRPANLQVELNVYFQQRELVAFCRA
LDITVCAYAPIGSPGLANVIKARGAEVPESAKFDPLTDPVVLKIAEHHKKTPA
QVLLRHCMQRDIVVIPKSTNAGRIKENFQVFDFELSKAEVDELDSLDKRAAG
RRFRMDQYKGLREHPEHPYDEPY

Bibliography

Adeboye, P. T., Bettiga, M., Aldaeus, F., Larsson, P. T. and Olsson, L. (2015) 'Catabolism of coniferyl aldehyde, ferulic acid and p-coumaric acid by *Saccharomyces cerevisiae* yields less toxic products.' *Microbial Cell Factories*, 14(1) p. 149.

Adeboye, P. T., Bettiga, M. and Olsson, L. (2014) 'The chemical nature of phenolic compounds determines their toxicity and induces distinct physiological responses in *Saccharomyces cerevisiae* in lignocellulose hydrolysates.' *AMB Express*, 4, May, p. 46.

Adeboye, P. T., Bettiga, M. and Olsson, L. (2017) 'ALD5, PAD1, ATF1 and ATF2 facilitate the catabolism of coniferyl aldehyde, ferulic acid and p-coumaric acid in *Saccharomyces cerevisiae*.' *Scientific Reports*. Nature Publishing Group, 7(1) p. 42635.

Aden, A. and Foust, T. (2009) 'Technoeconomic analysis of the dilute sulfuric acid and enzymatic hydrolysis process for the conversion of corn stover to ethanol.' *Cellulose*, 16(4) pp. 535–545.

Adi Santoso, Neng Herawati, and Yana Rubiana (2012) 'Effect of Methanol Induction and Incubation TIME on Expression of Human Erythropoietin in Methylotropic Yeast *Pichia Pastoris*.' *Makara Journal of Technology*, 16(1).

Agaphonov, M. and Alexandrov, A. (2014) 'Self-excising integrative yeast plasmid vectors containing an intronated recombinase gene.' *FEMS yeast research*, 14(7) pp. 1048–1054.

Aggarwal, S. and Mishra, S. (2020) 'Differential role of segments of α-mating factor secretion signal in *Pichia pastoris* towards granulocyte colony-stimulating factor emerging from a wild type or codon optimized copy of the gene.' *Microbial Cell Factories*, 19(1) p. 199.

Aharonowitz, Y. and Cohen, G. (1981) 'The microbiological production of pharmaceuticals.' *Scientific American*, 245(3) pp. 140–152.

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(12) pp. 5301–5317.

Ahmad, M., Winkler, C. M., Kolmbauer, M., Pichler, H., Schwab, H. and Emmerstorfer-Augustin, A. (2019) '*Pichia pastoris* protease-deficient and auxotrophic strains generated by a novel, user-friendly vector toolbox for gene deletion.' *Yeast*, 36(9) pp. 557–570.

Arai, T., Araki, R., Tanaka, A., Karita, S., Kimura, T., Sakka, K. and Ohmiya, K. (2003) 'Characterization of a Cellulase Containing a Family 30 Carbohydrate-Binding Module (CBM) Derived from Clostridium thermocellum CelJ: Importance of the CBM to Cellulose Hydrolysis.' *Journal of Bacteriology*, 185(2) pp. 504–512.

Araujo, P. R., Yoon, K., Ko, D., Smith, A. D., Qiao, M., Suresh, U., Burns, S. C. and Penalva, L. O. F. (2012) 'Before It Gets Started: Regulating Translation at the 5' UTR.' *Comparative and Functional Genomics,* 2012, p. 1-8.

Arevalo-Gallegos, A., Ahmad, Z., Asgher, M., Parra-Saldivar, R. and Iqbal, H. M. N. (2017) 'Lignocellulose: A sustainable material to produce value-added products with a zero waste approach—A review.' *International Journal of Biological Macromolecules*, 99, June, pp. 308–318.

Bajpai, P. (2018) 'Wood and Fiber Fundamentals.' *In Biermann's Handbook of Pulp and Paper*. Elsevier, pp. 19–74.

Barnett, J. A., Payne, R. W. and Yarrow, D. (2000) *Yeasts: characteristics and identification*. 3rd ed, Cambridge, U.K. ; New York, NY, USA: Cambridge University Press.

Barrero, J. J., Casler, J. C., Valero, F., Ferrer, P. and Glick, B. S. (2018) 'An improved secretion signal enhances the secretion of model proteins from *Pichia pastoris*.' *Microbial Cell Factories,* 17(1) p. 161.

Barski, O. A., Tipparaju, S. M. and Bhatnagar, A. (2008) 'The Aldo-Keto Reductase Superfamily and its Role in Drug Metabolism and Detoxification.' *Drug Metabolism Reviews*, 40(4) pp. 553–624.

Baruah, J., Nath, B. K., Sharma, R., Kumar, S., Deka, R. C., Baruah, D. C. and Kalita, E. (2018b) 'Recent Trends in the Pretreatment of Lignocellulosic Biomass for Value-Added Products.' *Frontiers in Energy Research,* 6, December, p. 141-160.

Berlemont, R. and Martiny, A. C. (2015) 'Genomic Potential for Polysaccharide Deconstruction in Bacteria.' Kivisaar, M. (ed.) *Applied and Environmental Microbiology*, 81(4) pp. 1513–1519.

Bill, R. M. (ed.) (2012) *Recombinant protein production in yeast: methods and protocols*. New York: Humana Press (Methods in molecular biology, 866).

Binod, P. and Pandey, A. (2015) 'Introduction.' *In Pretreatment of Biomass*. Elsevier, pp. 3–6.

Bischof, R. H., Ramoni, J. and Seiboth, B. (2016) 'Cellulases and beyond: the first 70 years of the enzyme producer Trichoderma reesei.' *Microbial Cell Factories*, 15(1) p. 106.

Blum, D. L., Kataeva, I. A., Li, X.-L. and Ljungdahl, L. G. (2000) 'Feruloyl Esterase Activity of the Clostridium thermocellum Cellulosome Can Be Attributed to Previously Unknown Domains of XynY and XynZ.' *Journal of Bacteriology,* 182(5) pp. 1346–1351.

Blum, D. L., Schubot, F. D., Ljungdahl, L. G., Rose, J. P. and Wang, B.-C. (2000) 'Crystallization and preliminary X-ray analysis of the *Clostridium thermocellum* cellulosome xylanase Z feruloyl esterase domain.' *Acta Crystallographica Section D Biological Crystallography*, 56(8) pp. 1027–1029.

Boer, W. de, Folman, L. B., Summerbell, R. C. and Boddy, L. (2005) 'Living in a fungal world: impact of fungi on soil bacterial niche development.' *FEMS Microbiology Reviews*, 29(4) pp. 795–811.

Boraston, A. B., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2004) 'Carbohydrate-binding modules: fine-tuning polysaccharide recognition.' *Biochemical Journal*, 382(3) pp. 769–781.

Broeker, J., Mechelke, M., Baudrexl, M., Mennerich, D., Hornburg, D., Mann, M., Schwarz, W. H., Liebl, W. and Zverlov, V. V. (2018) 'The hemicellulose-degrading enzyme system of the thermophilic bacterium Clostridium stercorarium: comparative characterisation and addition of new hemicellulolytic glycoside hydrolases.' *Biotechnology for Biofuels*, 11(1) p. 229.

Bugg, T. D. and Rahmanpour, R. (2015) 'Enzymatic conversion of lignin into renewable chemicals.' *Current Opinion in Chemical Biology*, 29, December, pp. 10–17.

Cameron, D. E., Bashor, C. J. and Collins, J. J. (2014) 'A brief history of synthetic biology.' *Nature Reviews Microbiology*, 12(5) pp. 381–390.

Caruthers, M. (1985) 'Gene synthesis machines: DNA chemistry and its uses.' *Science*, 230(4723) pp. 281–285.

Casini, A., MacDonald, J. T., Jonghe, J. D., Christodoulou, G., Freemont, P. S., Baldwin, G. S. and Ellis, T. (2014) 'One-pot DNA construction for synthetic biology: the Modular Overlap-Directed Assembly with Linkers (MODAL) strategy.' *Nucleic Acids Research*, 42(1) p. e7.

Cereghino, J. L. and Cregg, J. M. (2000) 'Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*.' *FEMS Microbiology Reviews*, 24(1) pp. 45–66.

Chakrabarty, A. M. (2010) 'Bioengineered bugs, drugs and contentious issues in patenting.' *Bioengineered Bugs*, 1(1) pp. 2–8.

Chang, S.-W., Shieh, C.-J., Lee, G.-C., Akoh, C. C. and Shaw, J.-F. (2006) 'Optimized growth kinetics of *Pichia pastoris* and recombinant Candida rugosa LIP1 production by RSM.' *Journal of Molecular Microbiology and Biotechnology*, 11(1–2) pp. 28–40.

Chaudhuri, B., Steube, K. and Stephan, C. (1992) 'The pro-region of the yeast prepro-alpha-factor is essential for membrane translocation of human insulin-like growth factor 1 in vivo.' *European Journal of Biochemistry*, 206(3) pp. 793–800.

Chen, W.-D. and Zhang, Y. (2012) 'Regulation of Aldo–Keto Reductases in Human Diseases.' *Frontiers in Pharmacology*, 3.

Chen, Y.-J., Liu, P., Nielsen, A. A. K., Brophy, J. A. N., Clancy, K., Peterson, T. and Voigt, C. A. (2013) 'Characterization of 582 natural and synthetic terminators and quantification of their design constraints.' *Nature Methods*, 10(7) pp. 659–664.

Church, G. (2014) *Regenesis - how synthetic biology will reinvent nature and ourselves.*

Coussement, P., Maertens, J., Beauprez, J., Van Bellegem, W. and De Mey, M. (2014) 'One step DNA assembly for combinatorial metabolic engineering.' *Metabolic Engineering*, 23, May, pp. 70–77.

Crane, D. I. and Gould, S. J. (1994) 'The *Pichia pastoris* HIS4 gene: nucleotide sequence, creation of a non-reverting his4 deletion mutant, and development of HIS4-based replicating and integrating plasmids.' *Current Genetics*, 26(5–6) pp. 443–450.

Cregg, J. M. (2004) 'Promoter for the *Pichia pastoris* formaldehyde dehydrogenase gene FLD1.'

Daly, R. and Hearn, M. T. W. (2005) 'Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production.' *Journal of Molecular Recognition*, 18(2) pp. 119–138.

Dahmen, N., Lewandowski, I., Zibek, S. and Weidtmann, A. (2019). Integrated lignocellulosic value chains in a growing bioeconomy: Status quo and perspectives. *GCB Bioenergy*, 11(1), pp.107-117.

Darby, R. A. J., Cartwright, S. P., Dilworth, M. V. and Bill, R. M. (2012) 'Which Yeast Species Shall I Choose? *Saccharomyces cerevisiae* Versus *Pichia pastoris* (Review).' *In* Bill, R. M. (ed.) *Recombinant Protein Production in Yeast*. Totowa, NJ: Humana Press (Methods in Molecular Biology), 2012(866), pp. 11–23.

Delic, M., Valli, M., Graf, A. B., Pfeffer, M., Mattanovich, D. and Gasser, B. (2013) 'The secretory pathway: exploring yeast diversity.' *FEMS Microbiology Reviews*, 37(6) pp. 872–914.

Demain, A. L. and Vaishnav, P. (2009) 'Production of recombinant proteins by microbes and higher organisms.' *Biotechnology Advances*, 27(3) pp. 297–306.

Dotsenko, G., Tong, X., Pilgaard, B., Busk, P. K. and Lange, L. (2016) 'Acidic–alkaline ferulic acid esterase from Chaetomium thermophilum var. dissitum: Molecular cloning and characterization of recombinant enzyme expressed in *Pichia pastoris*.' *Biocatalysis and Agricultural Biotechnology*, 5, January, pp. 48–55.

Duellman, T., Burnett, J. and Yang, J. (2015) 'Quantitation of secreted proteins using mCherry fusion constructs and a fluorescent microplate reader.' *Analytical biochemistry*, 473, March, pp. 34–40.

Ebringerová, A. and Heinze, T. (2000) 'Xylan and xylan derivatives – biopolymers with valuable properties, 1. Naturally occurring xylans structures, isolation procedures and properties.' Macromolecular Rapid Communications, 21(9) pp. 542–556.

Elena, C., Ravasi, P., Castelli, M. E., Peirú, S. and Menzella, H. G. (2014) 'Expression of codon optimized genes in microbial systems: current industrial applications and perspectives.' Frontiers in Microbiology, 5, February, p. 21.

Ellis, D. (1921) 'The Yeasts.' *Nature*, 107(2691) pp. 387–388.

Engler, C., Kandzia, R. and Marillonnet, S. (2008) ‘A one pot, one step, precision cloning method with high throughput capability.’ *PloS One*, 3(11) p. e3647.

Eugene Raj, A., Sathish Kumar, H. S., Umesh Kumar, S., Misra, M. C., Ghildyal, N. P. and Karanth, N. G. (2002) ‘High-Cell-Density Fermentation of Recombinant *Saccharomyces cerevisiae* Using Glycerol.’ *Biotechnology Progress*, 18(5) pp. 1130–1132.

Eveleigh, D., Mandels, M., Andreotti, R. and Roche, C. (2009) ‘Measurement of saccharifing cellulase.’ *Biotechnology for biofuels*, 2, September, p. 21.

Fargione, J., Hill, J., Tilman, D., Polasky, S. and Hawthorne, P. (2008) ‘Land Clearing and the Biofuel Carbon Debt.’ *Science*, 319(5867) pp. 1235–1238.

Faulds, C. B., Aliwan, F. O., de Vries, R. P., Pickersgill, R. W., Visser, J. and Williamson, G. (1998) ‘Chemical and thermal stability of ferulic acid (feruloyl) esterases from Aspergillus.’ *In Progress in Biotechnology*. Elsevier, pp. 41–46.

Finnigan, W., Thomas, A., Cromar, H., Gough, B., Snajdrova, R., Adams, J. P., Littlechild, J. A. and Harmer, N. J. (2017) ‘Characterization of Carboxylic Acid Reductases as Enzymes in the Toolbox for Synthetic Chemistry.’ *ChemCatChem*, 9(6) pp. 1005–1017.

Fitzgerald, I. and Glick, B. S. (2014) ‘Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting.’ *Microbial Cell Factories*, 13(1) pp. 125.

Foyle, T., Jennings, L. and Mulcahy, P. (2007) ‘Compositional analysis of lignocellulosic materials: Evaluation of methods used for sugar analysis of waste paper and straw.’ *Bioresource Technology*, 98(16) pp. 3026–3036.

Freemont, P. S. (2019) ‘Synthetic biology industry: data-driven design is creating new opportunities in biotechnology.’ Bayley, H. (ed.) *Emerging Topics in Life Sciences*, 3(5) pp. 651–657.

Gahloth, D., Aleku, G. A. and Leys, D. (2020) ‘Carboxylic acid reductase: Structure and mechanism.’ *Journal of Biotechnology*, 307, January, pp. 107–113.

Gao, D., Chundawat, S. P. S., Sethi, A., Balan, V., Gnanakaran, S. and Dale, B. E. (2013) ‘Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis.’ *Proceedings of the National Academy of Sciences*, 110(27) pp. 10922–10927.

García-Fruitós, E. (ed.) (2015) *Insoluble proteins: methods and protocols*. New York: Humana Press (Methods in molecular biology, (1258), pp 10.

Gasser, B. and Mattanovich, D. (2018) ‘A yeast for all seasons – Is *Pichia pastoris* a suitable chassis organism for future bioproduction?’ *FEMS Microbiology Letters*, 365(17).

Gent, S., Twedt, M., Gerometta, C. and Almberg, E. (2017) 'Introduction to Feedstocks.' *In Theoretical and Applied Aspects of Biomass Torrefaction*. Elsevier, pp. 17–39.

Glazer, A. N. and Nikaido, H. (2007) *Microbial biotechnology: fundamentals of applied microbiology*. 2nd ed, Cambridge ; New York: Cambridge University Press. Pp. 432.

de Gonzalo, G., Colpa, D. I., Habib, M. H. M. and Fraaije, M. W. (2016) 'Bacterial enzymes involved in lignin degradation.' *Journal of Biotechnology*, 236, October, pp. 110–119.

Gray, K. A., Zhao, L. and Emptage, M. (2006) 'Bioethanol.' *Current Opinion in Chemical Biology*, 10(2) pp. 141–146.

Guerriero, G., Hausman, J.-F., Strauss, J., Ertan, H. and Siddiqui, K. S. (2016) 'Lignocellulosic biomass: Biosynthesis, degradation, and industrial utilization.' *Engineering in Life Sciences*, 16(1) pp. 1–16.

Gusakov, A. V., Salanovich, T. N., Antonov, A. I., Ustinov, B. B., Okunev, O. N., Burlingame, R., Emalfarb, M., Baez, M. and Sinitsyn, A. P. (2007) 'Design of highly efficient cellulase mixtures for enzymatic hydrolysis of cellulose.' *Biotechnology and Bioengineering*, 97(5) pp. 1028–1038.

Halachmi, N. and Lev, Z. (1996) 'The Sec1 family: a novel family of proteins involved in synaptic transmission and general secretion.' *Journal of Neurochemistry*, 66(3) pp. 889–897.

Hansen, E. H., Møller, B. L., Kock, G. R., Bünner, C. M., Kristensen, C., Jensen, O. R., Okkels, F. T., Olsen, C. E., Motawia, M. S. and Hansen, J. (2009) 'De novo biosynthesis of vanillin in fission yeast (Schizosaccharomyces pombe) and baker's yeast (*Saccharomyces cerevisiae*).' *Applied and Environmental Microbiology*, 75(9) pp. 2765–2774.

Hanson, G. and Coller, J. (2018) 'Codon optimality, bias and usage in translation and mRNA decay.' *Nature Reviews Molecular Cell Biology*, 19(1) pp. 20–30.

Harris, P. J. and Hartley, R. D. (1980) 'Phenolic constituents of the cell walls of monocotyledons.' *Biochemical Systematics and Ecology*, 8(2) pp. 153–160.

Harris, P. J. and Stone, B. A. (2008) 'Chemistry and Molecular Organization of Plant Cell Walls.' *In* Himmel, M. E. (ed.) *Biomass Recalcitrance*. Oxford, UK: Blackwell Publishing Ltd., pp. 61–93.

Hasslacher, M., Schall, M., Hayn, M., Bona, R., Rumbold, K., Lückl, J., Griengl, H., Kohlwein, S. D. and Schwab, H. (1997) 'High-Level Intracellular Expression of Hydroxynitrile Lyase from the Tropical Rubber TreeHevea brasiliensisin Microbial Hosts.' *Protein Expression and Purification*, 11(1) pp. 61–71.

Hatfield, R. D., Rancour, D. M. and Marita, J. M. (2016) 'Grass Cell Walls: A Story of Cross-Linking.' *Frontiers in Plant Science*, 7 p. 2056.

He, A., Li, T., Daniels, L., Fotheringham, I. and Rosazza, J. P. N. (2004) 'Nocardia sp. carboxylic acid reductase: cloning, expression, and characterization of a new aldehyde oxidoreductase family.' *Applied and Environmental Microbiology*, 70(3) pp. 1874–1881.

Hirose, Y. and Ohkuma, Y. (2007) 'Phosphorylation of the C-terminal domain of RNA polymerase II plays central roles in the integrated events of eucaryotic gene expression.' *Journal of Biochemistry*, 141(5) pp. 601–608.

Horvat, M., Larch, T.-S., Rudroff, F. and Winkler, M. (2020) 'Amino Benzamidoxime (ABAO)-Based Assay to Identify Efficient Aldehyde-Producing *Pichia pastoris* Clones.' *Advanced Synthesis & Catalysis*, 362(21) pp. 4673–4679.

Hori, R. and Carey, M. (1994) 'The role of activators in assembly of RNA polymerase II transcription complexes.' *Current opinion in genetics & development*. 4(2), pp.236-244.

Hoshida, H., Kondo, M., Kobayashi, T., Yarimizu, T. and Akada, R. (2017) '5′-UTR introns enhance protein expression in the yeast *Saccharomyces cerevisiae*.' *Applied Microbiology and Biotechnology*, 101(1) pp. 241–251.

Hou, C. T. and Shaw, J.-F. (2008) *Biocatalysis and Bioenergy*. John Wiley & Sons.

Hou, J., Tyo, K. E. J., Liu, Z., Petranovic, D. and Nielsen, J. (2012) 'Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*.' *FEMS Yeast Research*, 12(5) pp. 491–510.

Hou, J., Tyo, K., Liu, Z., Petranovic, D. and Nielsen, J. (2012) 'Engineering of vesicle trafficking improves heterologous protein secretion in *Saccharomyces cerevisiae*.' *Metabolic Engineering*, 14(2) pp. 120–127.

Huang, D. and Shusta, E. V. (2008) 'Secretion and Surface Display of Green Fluorescent Protein Using the Yeast *Saccharomyces cerevisiae*.' *Biotechnology Progress*, 21(2) pp. 349–357.

Hunt, C. J., Antonopoulou, I., Tanksale, A., Rova, U., Christakopoulos, P. and Haritos, V. S. (2017) 'Insights into substrate binding of ferulic acid esterases by arabinose and methyl hydroxycinnamate esters and molecular docking.' *Scientific Reports*, 7(1) p. 17315.

Hynes, M. J. and O'Coinceanainn, M. (2004) 'The kinetics and mechanisms of reactions of iron(III) with caffeic acid, chlorogenic acid, sinapic acid, ferulic acid and naringin☆.' *Journal of Inorganic Biochemistry*, 98(8) pp. 1457–1464.

Idiris, A., Tohda, H., Kumagai, H. and Takegawa, K. (2010) 'Engineering of protein secretion in yeast: strategies and impact on protein production.' *Applied Microbiology and Biotechnology*, 86(2) pp. 403–417.

Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttilä, M., Ando, T. and Samejima, M. (2011) 'Traffic Jams Reduce Hydrolytic Efficiency of Cellulase on Cellulose Surface.' *Science*, 333(6047) pp. 1279–1282.

Jackson, R. J., Hellen, C. U. T. and Pestova, T. V. (2012) 'Termination and post-termination events in eukaryotic translation.' *Advances in Protein Chemistry and Structural Biology*, 86 pp. 45–93.

Jalak, J., Kurašin, M., Teugjas, H. and Väljamäe, P. (2012) 'Endo-exo Synergism in Cellulose Hydrolysis Revisited.' *Journal of Biological Chemistry*, 287(34) pp. 28802–28815.

Jensen, D. and Schekman, R. (2011) 'COPII-mediated vesicle formation at a glance.' *Journal of Cell Science*, 124(1) pp. 1–4.

Jozala, A. F., Geraldes, D. C., Tundisi, L. L., Feitosa, V. de A., Breyer, C. A., Cardoso, S. L., Mazzola, P. G., Oliveira-Nascimento, L. de, Rangel-Yagui, C. de O., Magalhães, P. de O., Oliveira, M. A. de and Pessoa, A. (2016) 'Biopharmaceuticals from microorganisms: from production to purification.' *Brazilian Journal of Microbiology*, 47(Suppl 1) pp. 51–63.

Juge, N., Williamson, G., Puigserver, A., Cummings, N. J., Connerton, I. F. and Faulds, C. B. (2001) 'High-level production of recombinant Aspergillus niger cinnamoyl esterase (FAEA) in the methylotrophic yeast *Pichia pastoris*.' *FEMS Yeast Research*, 1(2) pp. 127–132.

Jung, H. G. and Phillips, R. L. (2010) 'Putative Seedling Ferulate Ester (*sfe*) Maize Mutant: Morphology, Biomass Yield, and Stover Cell Wall Composition and Rumen Degradability.' *Crop Science*, 50(1) pp. 403–418.

Kapp, L. D. and Lorsch, J. R. (2004) 'The molecular mechanics of eukaryotic translation.' *Annual Review of Biochemistry*, 73 pp. 657–704.

Karbalaei, M., Rezaee, S. A. and Farsiani, H. (2020) 'Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins.' *Journal of Cellular Physiology*, 235(9) pp. 5867–5881.

Kaushik, N., Lamminmäki, U., Khanna, N. and Batra, G. (2020) 'Enhanced cell density cultivation and rapid expression-screening of recombinant *Pichia pastoris* clones in microscale.' *Scientific Reports*, 10(1) p. 7458.

Keshavarz, B. and Khalesi, M. (2016) '*Trichoderma reesei* , a superior cellulase source for industrial applications.' *Biofuels*, 7(6) pp. 713–721.

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J. and Bolton, E. E. (2021) 'PubChem in 2021: new data content and improved web interfaces.' *Nucleic Acids Research*, 49(D1) pp. D1388–D1395.

Klinke, H. B., Thomsen, A. B. and Ahring, B. K. (2004) 'Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass.' *Applied Microbiology and Biotechnology*, 66(1) pp. 10–26.

Kondo, T. and Yumura, S. (2019) ‘An improved molecular tool for screening bacterial colonies using GFP expression enhanced by a Dictyostelium sequence.’ *BioTechniques*. Future Science, 68(2) pp. 91–95.

Kont, R., Kari, J., Borch, K., Westh, P. and Väljamäe, P. (2016) ‘Inter-domain Synergism Is Required for Efficient Feeding of Cellulose Chain into Active Site of Cellobiohydrolase Cel7A.’ *Journal of Biological Chemistry*, 291(50) pp. 26013–26023.

Krainer, F. W., Dietzsch, C., Hajek, T., Herwig, C., Spadiut, O. and Glieder, A. (2012) ‘Recombinant protein expression in Pichia pastoris strains with an engineered methanol utilization pathway.’ *Microbial Cell Factories*, 11(1) article no: 22

Kumar, R. (2019) ‘Simplified protocol for faster transformation of (a large number of) *Pichia pastoris* strains.’ *Yeast*, 36(6) pp. 399–410.

Kung, S. H., Retchless, A. C., Kwan, J. Y. and Almeida, R. P. P. (2013) ‘Effects of DNA Size on Transformation and Recombination Efficiencies in *Xylella fastidiosa*.’ *Applied and Environmental Microbiology*, 79(5) pp. 1712–1717.

Ladisch, M. R. and Kohlmann, K. L. (1992) ‘Recombinant human insulin.’ *Biotechnology Progress*, 8(6) pp. 469–478.

Lakhundi, S., Siddiqui, R. and Khan, N. (2015) ‘Cellulose degradation: a therapeutic strategy in the improved treatment of Acanthamoeba infections.’ *Parasites & Vectors*, 8(1) p. 23.

Laluce, C., Schenberg, A. C. G., Gallardo, J. C. M., Coradello, L. F. C. and Pombeiro-Sponchiado, S. R. (2012) ‘Advances and Developments in Strategies to Improve Strains of *Saccharomyces cerevisiae* and Processes to Obtain the Lignocellulosic Ethanol−A Review.’ *Applied Biochemistry and Biotechnology*, 166(8) pp. 1908–1926.

Li, C., Lin, Y., Zheng, X., Yuan, Q., Pang, N., Liao, X., Huang, Y., Zhang, X. and Liang, S. (2017) ‘Recycling of a selectable marker with a self-excisable plasmid in *Pichia pastoris*.’ *Scientific Reports*, 7(1) p. 11113.

Li, D., Zhang, B., Li, S., Zhou, J., Cao, H., Huang, Y. and Cui, Z. (2017) ‘A Novel Vector for Construction of Markerless Multicopy Overexpression Transformants in *Pichia pastoris*.’ *Frontiers in Microbiology*, 8 p. 1698.

Li, J., Du, L. and Wang, L. (2010) ‘Glycosidic-Bond Hydrolysis Mechanism Catalyzed by Cellulase Cel7A from Trichoderma reesei: A Comprehensive Theoretical Study by Performing MD, QM, and QM/MM Calculations.’ *The Journal of Physical Chemistry B*, 114(46) pp. 15261–15268.

Li, J., Xu, H., Bentley, W. E. and Rao, G. (2002) ‘Impediments to Secretion of Green Fluorescent Protein and Its Fusion from *Saccharomyces cerevisiae*.’ *Biotechnology Progress*, 18(4) pp. 831–838.

Li, X., He, X., Li, Z. and Wang, F. (2013) 'Combined Strategies for Improving the Production of Recombinant Rhizopus oryzae Lipase in *Pichia pastoris*.' *BioResources*, 8(2) pp. 2867–2880.

Li, Y., Arakawa, G., Tokuda, G., Watanabe, H. and Arioka, M. (2017) 'Heterologous expression in *Pichia pastoris* and characterization of a β-glucosidase from the xylophagous cockroach Panesthia angustipennis spadica displaying high specific activity for cellobiose.' *Enzyme and Microbial Technology*, 97, February, pp. 104–113.

Liang, S., Wang, B., Pan, L., Ye, Y., He, M., Han, S., Zheng, S., Wang, X. and Lin, Y. (2012) 'Comprehensive structural annotation of *Pichia pastoris* transcriptome and the response to various carbon sources using deep paired-end RNA sequencing.' *BMC Genomics*, 13(1) p. 738.

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. (2013a) 'The effect of α-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*.' *Gene*, 519(2) pp. 311–317.

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. (2013b) 'The effect of α-mating factor secretion signal mutations on recombinant protein expression in *Pichia pastoris*.' *Gene*, 519(2) pp. 311–317.

Lin-Cereghino, J., Wong, W. W., Xiong, S., Giang, W., Luong, L. T., Vu, J., Johnson, S. D. and Lin-Cereghino, G. P. (2005) 'Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast *Pichia pastoris*.' *BioTechniques*, 38(1) pp. 44–48.

Liu, Q., Shi, X., Song, L., Liu, H., Zhou, X., Wang, Q., Zhang, Y. and Cai, M. (2019) 'CRISPR–Cas9-mediated genomic multiloci integration in *Pichia pastoris*.' *Microbial Cell Factories*, 18(1) p. 144.

Liu, W.-C., Gong, T., Wang, Q.-H., Liang, X., Chen, J.-J. and Zhu, P. (2016) 'Scaling-up Fermentation of *Pichia pastoris* to demonstration-scale using new methanol-feeding strategy and increased air pressure instead of pure oxygen supplement.' *Scientific Reports*, 6(1) p. 18439.

Lynd, L. R., Weimer, P. J., van Zyl, W. H. and Pretorius, I. S. (2002) 'Microbial Cellulose Utilization: Fundamentals and Biotechnology.' *Microbiology and Molecular Biology Reviews*, 66(3) pp. 506–577.

Mabinya, L., Mafunga, T. and Brand, J. (2006) 'Determination of ferulic acid and related compounds by thin layer chromatography.' *AFRICAN JOURNAL OF BIOTECHNOLOGY*, 5(13), pp. 1271-1273.

Mabinya, L. V., Brand, J. M., Raats, J. G. and Trollope, W. S. W. (2002) 'Estimation of grazing by herbivores from analysis of dung.' *African Journal of Range & Forage Science*, 19(3) pp. 175–176.

Mandelli, F., Brenelli, L. B., Almeida, R. F., Goldbeck, R., Wolf, L. D., Hoffmam, Z. B., Ruller, R., Rocha, G. J. M., Mercadante, A. Z. and Squina, F. M. (2014) 'Simultaneous production of xylooligosaccharides and antioxidant compounds from sugarcane bagasse via enzymatic hydrolysis.' *Industrial Crops and Products*, 52, January, pp. 770–775.

Mapemba, L. and Epplin, F. M. (2004). Economics of a coordinated biorefinery feedstock harvest system: lignocellulosic biomass harvest cost. *Biomass and Bioenergy*, 27(4), pp. 327-337.

Marriott, P. E., Gómez, L. D. and McQueen-Mason, S. J. (2016) 'Unlocking the potential of lignocellulosic biomass through plant science.' *New Phytologist*, 209(4) pp. 1366–1381.

Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., Chapman, J., Chertkov, O., Coutinho, P. M., Cullen, D., Danchin, E. G. J., Grigoriev, I. V., Harris, P., Jackson, M., Kubicek, C. P., Han, C. S., Ho, I., Larrondo, L. F., de Leon, A. L., Magnuson, J. K., Merino, S., Misra, M., Nelson, B., Putnam, N., Robbertse, B., Salamov, A. A., Schmoll, M., Terry, A., Thayer, N., Westerholm-Parvinen, A., Schoch, C. L., Yao, J., Barabote, R., Nelson, M. A., Detter, C., Bruce, D., Kuske, C. R., Xie, G., Richardson, P., Rokhsar, D. S., Lucas, S. M., Rubin, E. M., Dunn-Coleman, N., Ward, M. and Brettin, T. S. (2008) 'Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina).' *Nature Biotechnology*, 26(5) pp. 553–560.

Martínez, J. L., Liu, L., Petranovic, D. and Nielsen, J. (2012) 'Pharmaceutical protein production by yeast: towards production of human blood proteins by microbial fermentation.' *Current Opinion in Biotechnology*, 23(6) pp. 965–971.

Mayr, C. (2019) 'What Are 3' UTRs Doing?' *Cold Spring Harbor Perspectives in Biology*, 11(10) p. 1-16.

Miyauchi, S., Te'o, V. S., Bergquist, P. L. and Nevalainen, K. M. H. (2013) 'Expression of a bacterial xylanase in Trichoderma reesei under the egl2 and cbh2 glycosyl hydrolase gene promoters.' *New Biotechnology*, 30(5) pp. 523–530.

Monod, J. and Jacob, F. (1961) 'Teleonomic mechanisms in cellular metabolism, growth, and differentiation.' *Cold Spring Harbor Symposia on Quantitative Biology*, 26 pp. 389–401.

Moon, C.-S. (2017) 'Estimations of the lethal and exposure doses for representative methanol symptoms in humans.' *Annals of Occupational and Environmental Medicine*, 29(1) p. 44-50.

Moore, K., & Jung, H. (2001). Lignin and fiber digestion. *Journal of Range Management, 54*, pp. 420-430.

Mootz, H. D., Finking, R. and Marahiel, M. A. (2001) '4′-Phosphopantetheine Transfer in Primary and Secondary Metabolism of Bacillus subtilis*.' *Journal of Biological Chemistry*. Elsevier, 276(40) pp. 37289–37298.

Morrow, J. F., Cohen, S. N., Chang, A. C. Y., Boyer, H. W., Goodman, H. M. and Helling, R. B. (1974) 'Replication and Transcription of Eukaryotic DNA in Esherichia coli.' *Proceedings of the National Academy of Sciences*, 71(5) pp. 1743–1747.

Mortimer, R. K. (2000) 'Evolution and Variation of the Yeast (Saccharomyces) Genome.' *Genome Research*, 10(4) pp. 403–409.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. (1986) 'Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction.' *Cold Spring Harbor Symposia on Quantitative Biology*, 51(0) pp. 263–273.

Murashima, K., Kosugi, A. and Doi, R. H. (2002) 'Synergistic Effects on Crystalline Cellulose Degradation between Cellulosomal Cellulases from Clostridium cellulovorans.' *Journal of Bacteriology*, 184(18) pp. 5088–5095.

Näätsaari, L., Mistlberger, B., Ruth, C., Hajek, T., Hartner, F. S. and Glieder, A. (2012) 'Deletion of the *Pichia pastoris* KU70 Homologue Facilitates Platform Strain Generation for Gene Expression and Synthetic Biology.' *PLOS ONE*. Public Library of Science, 7(6) p. e39720.

Nakamura, Y., Nishi, T., Noguchi, R., Ito, Y., Watanabe, T., Nishiyama, T., Aikawa, S., Hasunuma, T., Ishii, J., Okubo, Y. and Kondo, A. (2018) 'A Stable, Autonomously Replicating Plasmid Vector Containing *Pichia pastoris* Centromeric DNA.' Löffler, F. E. (ed.) *Applied and Environmental Microbiology*, 84(15) pp. e02882-17, /aem/84/15/e02882-17.atom.

Nakazawa, H., Okada, K., Kobayashi, R., Kubota, T., Onodera, T., Ochiai, N., Omata, N., Ogasawara, W., Okada, H. and Morikawa, Y. (2008) 'Characterization of the catalytic domains of Trichoderma reesei endoglucanase I, II, and III, expressed in Escherichia coli.' *Applied Microbiology and Biotechnology*, 81(4) pp. 681–689.

Nanda, S., A. Kozinski, J. and K. Dalai, A. (2015) 'Lignocellulosic Biomass: A Review of Conversion Technologies and Fuel Products.' *Current Biochemical Engineering*, 3(1) pp. 24–36.

Napora-Wijata, K., Strohmeier, G. A. and Winkler, M. (2014) 'Biocatalytic reduction of carboxylic acids.' *Biotechnology Journal*, 9(6) pp. 822–843.

Nett, J. H. and Gerngross, T. U. (2003) 'Cloning and disruption of the PpURA5 gene and construction of a set of integration vectors for the stable genetic modification of *Pichia pastoris*.' *Yeast (Chichester, England)*, 20(15) pp. 1279–1290.

Niazi, S. K. (2002) *Handbook of Biogeneric Therapeutic Proteins: Regulatory, Manufacturing, Testing, and Patent Issues*. CRC Press. Pp. 49-50.

Niño-Medina, G., Carvajal-Millán, E., Rascon-Chu, A., Marquez-Escalante, J. A., Guerrero, V. and Salas-Muñoz, E. (2010) 'Feruloylated arabinoxylans and

arabinoxylan gels: structure, sources and applications.' *Phytochemistry Reviews*, 9(1) pp. 111–120.

de O. Buanafina, M. M. (2009) 'Feruloylation in Grasses: Current and Future Perspectives.' *Molecular Plant*, 2(5) pp. 861–872.

Oberdoerffer, P. (2003) 'Unidirectional Cre-mediated genetic inversion in mice using the mutant loxP pair lox66/lox71.' *Nucleic Acids Research*, 31(22) pp. 140e–1140.

Ogata, K., Nishikawa, H. and Ohsugi, M. (1969) 'A Yeast Capable of Utilizing Methanol.' *Agricultural and Biological Chemistry*, 33(10) pp. 1519–1520.

Oliveira, C., Carvalho, V., Domingues, L. and Gama, F. M. (2015) 'Recombinant CBM-fusion technology — Applications overview.' *Biotechnology Advances*, 33(3–4) pp. 358–369.

Paes, B. G., Steindorff, A. S., Formighieri, E. F., Pereira, I. S. and Almeida, J. R. M. (2021) 'Physiological characterization and transcriptome analysis of *Pichia pastoris* reveals its response to lignocellulose-derived inhibitors.' *AMB Express*, 11(1) p. 2.

Pan, R., Zhang, J., Shen, W.-L., Tao, Z.-Q., Li, S.-P. and Yan, X. (2011) 'Sequential deletion of *Pichia pastoris* genes by a self-excisable cassette: Rapid and unmarked gene deletions in *Pichia pastoris*.' *FEMS Yeast Research*, 11(3) pp. 292–298.

Pell, G., Taylor, E. J., Gloster, T. M., Turkenburg, J. P., Fontes, C. M. G. A., Ferreira, L. M. A., Nagy, T., Clark, S. J., Davies, G. J. and Gilbert, H. J. (2004) 'The Mechanisms by Which Family 10 Glycoside Hydrolases Bind Decorated Substrates.' *Journal of Biological Chemistry*, 279(10) pp. 9597–9605.

Peng, B., Williams, T. C., Henry, M., Nielsen, L. K. and Vickers, C. E. (2015) 'Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities.' *Microbial Cell Factories*, 14(1) p. 91-100.

Pérez, M. L. and González, G. V. (2017) 'DIFFERENTIAL TOXICITY CAUSED BY METHANOL ON THE GROWTH OF *Pichia pastoris* CULTURED IN SOLID-STATE AND IN SUBMERGED FERMENTATION.' *Revista Mexicana de Ingeniería Química*, 16(3) pp. 735–743.

Pérez, S. and Mazeau, K. (2004) 'Conformations, Structures, and Morphologies of Celluloses.' *In* Dumitriu, S. (ed.) *Polysaccharides*. CRC Press. Pp. 44-68.

Pham, P. V. (2018) 'Chapter 19 - Medical Biotechnology: Techniques and Applications.' *In* Barh, D. and Azevedo, V. (eds) *Omics Technologies and Bio-Engineering*. Academic Press, pp. 449–469.

Pilar Rauter, A., Lindhorst, T. and Queneau, Y. (eds) (2017) *Carbohydrate Chemistry: Chemical and Biological Approaches*. *Cambridge: Royal Society of Chemistry*, 43, pp. 1-70.

Pope, B. and Kent, H. M. (1996) 'High Efficiency 5 Min Transformation of Escherichia Coli.' *Nucleic Acids Research*, 24(3) pp. 536–537.

Prade, R. A. (1996) 'Xylanases: from Biology to BioTechnology.' *Biotechnology and Genetic Engineering Reviews*, 13(1) pp. 101–132.

Prielhofer, R., Barrero, J. J., Steuer, S., Gassler, T., Zahrl, R., Baumann, K., Sauer, M., Mattanovich, D., Gasser, B. and Marx, H. (2017) 'GoldenPiCS: a Golden Gate-derived modular cloning system for applied synthetic biology in the yeast *Pichia pastoris*.' *BMC Systems Biology*, 11(1) p. 123.

Prielhofer, R., Cartwright, S. P., Graf, A. B., Valli, M., Bill, R. M., Mattanovich, D. and Gasser, B. (2015) '*Pichia pastoris* regulates its gene-specific response to different carbon sources at the transcriptional, rather than the translational, level.' *BMC Genomics*, 16(1) p. 167.

Puetz, J. and Wurm, F. M. (2019) 'Recombinant Proteins for Industrial versus Pharmaceutical Purposes: A Review of Process and Pricing.' *Processes*. Multidisciplinary Digital Publishing Institute, 7(8) p. 476-485.

Qin, X., Qian, J., Yao, G., Zhuang, Y., Zhang, S. and Chu, J. (2011) '*GAP* Promoter Library for Fine-Tuning of Gene Expression in *Pichia pastoris*.' *Applied and Environmental Microbiology*, 77(11) pp. 3600–3608.

Reider Apel, A., d'Espaux, L., Wehrs, M., Sachs, D., Li, R. A., Tong, G. J., Garber, M., Nnadi, O., Zhuang, W., Hillson, N. J., Keasling, J. D. and Mukhopadhyay, A. (2017) 'A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*.' *Nucleic Acids Research*, 45(1) pp. 496–508.

Rennie, E. A. and Scheller, H. V. (2014) 'Xylan biosynthesis.' *Current Opinion in Biotechnology*, 26, April, pp. 100–107.

Richards, H., Baker, P. and Iwuoha, E. (2012) 'Metal Nanoparticle Modified Polysulfone Membranes for Use in Wastewater Treatment: A Critical Review.' *Journal of Surface Engineered Materials and Advanced Technology*, 2, July, p. 183-193.

Roberts, M., Cranenburgh, R., Stevens, M. and Oyston, P. (2013) 'Synthetic Biology. Biology by design.' *Microbiology (Reading, England)*, 159, June, pp. 1219-1220.

Robinson, A. S., Hines, V. and Wittrup, K. D. (1994) 'Protein disulfide isomerase overexpression increases secretion of foreign proteins in *Saccharomyces cerevisiae*.' *Bio/Technology (Nature Publishing Company)*, 12(4) pp. 381–384.

Robinson, A. S., Bockhaus, J. A., Voegler, A. C. and Wittrup, K. D. (1996) 'Reduction of BiP levels decreases heterologous protein secretion in Saccharomyces cerevisiae.' *The Journal of Biological Chemistry*, 271(17) pp. 1017–1022.

Romanos, M. A., Scorer, C. A. and Clare, J. J. (1992) 'Foreign gene expression in yeast: a review.' *Yeast*, 8(6) pp. 423–488.

Rosales-Calderon, O. and Arantes, V. (2019) 'A review on commercial-scale high-value products that can be produced alongside cellulosic ethanol.' *Biotechnology for Biofuels*, 12(1) article no:240.

Rosano, G. L. and Ceccarelli, E. A. (2014) ‘Recombinant protein expression in Escherichia coli: advances and challenges.’ *Frontiers in Microbiology*, 5, April.

Sanderson, K. (2011) ‘Lignocellulose: A chewy problem.’ *Nature*. Nature Publishing Group, 474(7352) pp. S12–S14.

Schellekens, J. (2013) *The use of molecular chemistry (pyrolysis-GC/MS) in the environmental interpretation of peat*.

Schoenherr, S., Ebrahimi, M. and Czermak, P. (2018) ‘Lignin Degradation Processes and the Purification of Valuable Products.’ *In* Poletto, M. (ed.) *Lignin - Trends and Applications*. InTech.

Schuster, A. and Schmoll, M. (2010) ‘Biology and biotechnology of Trichoderma.’ *Applied Microbiology and Biotechnology*, 87(3) pp. 787–799.

Serrano, L. (2007) ‘Synthetic biology: promises and challenges.’ Molecular Systems Biology, 3(1) article no: 158.

Sezonov, G., Joseleau-Petit, D. and D’Ari, R. (2007) ‘Escherichia coli Physiology in Luria-Bertani Broth.’ *Journal of Bacteriology*, 189(23) pp. 8746–8749.

Shapira, P., Kwon, S. and Youtie, J. (2017) ‘Tracking the emergence of synthetic biology.’ *Scientometrics*, 112(3) pp. 1439–1469.

Sharma, O. P., Bhat, T. K. and Singh, B. (1998) ‘Thin-layer chromatography of gallic acid, methyl gallate, pyrogallol, phloroglucinol, catechol, resorcinol, hydroquinone, catechin, epicatechin, cinnamic acid, p-coumaric acid, ferulic acid and tannic acid.’ *Journal of Chromatography A*, 822(1) pp. 167–171.

Shen, S., Sulter, G., Jeffries, T. W. and Cregg, J. M. (1998) ‘A strong nitrogen source-regulated promoter for controlled expression of foreign genes in the yeast *Pichia pastoris*.’ *Gene*, 216(1) pp. 93–102.

Sheng, J., Flick, H. and Feng, X. (2017) ‘Systematic Optimization of Protein Secretory Pathways in *Saccharomyces cerevisiae* to Increase Expression of Hepatitis B Small Antigen.’ *Frontiers in Microbiology*, 8, May, article no: 875.

Shiloach, J. and Fass, R. (2005) ‘Growing E. coli to high cell density—A historical perspective on method development.’ *Biotechnology Advances*, 23(5) pp. 345–357.

Shoseyov, O., Shani, Z. and Levy, I. (2006) ‘Carbohydrate Binding Modules: Biochemical Properties and Novel Applications.’ *Microbiology and Molecular Biology Reviews*, 70(2) pp. 283–295.

Shreaz, S., Maurya, I., Bhatia, R., Khan, N., Muralidhar, S. and Khan, L. (2011) ‘Influences of cinnamic aldehydes on plasma membrane H+ ATPase activity and ultrastructure of Candida.’ *Journal of medical microbiology*, 62, October, pp. 232-240.

Sierra-Alvarez, R. and Lettinga, G. (2007) 'The methanogenic toxicity of wastewater lignins and lignin related compounds.' *Journal of Chemical Technology & Biotechnology*, 50(4) pp. 443–455.

Silva, E. A. B. da, Zabkova, M., Araújo, J. D., Cateto, C. A., Barreiro, M. F., Belgacem, M. N. and Rodrigues, A. E. (2009) 'An integrated process to produce vanillin and lignin-based polyurethanes from Kraft lignin.' *Chemical Engineering Research and Design*, 87(9) pp. 1276–1292.

Smith, H. O. and Welcox, K. W. (1970) 'A Restriction enzyme from Hemophilus influenzae.' *Journal of Molecular Biology*, 51(2) pp. 379–391.

Souza, W. R. de (2013) 'Microbial Degradation of Lignocellulosic Biomass.' *Sustainable Degradation of Lignocellulosic Biomass - Techniques, Applications and Commercialization*. *IntechOpen*, May, pp. 207-248.

Sreekrishna, K. (2010) '*Pichia* , Optimization of Protein Expression.' *In Encyclopedia of Industrial Biotechnology*. Hoboken, NJ, USA: John Wiley & Sons, Inc., p. eib480.

Staley, C. A., Huang, A., Nattestad, M., Oshiro, K. T., Ray, L. E., Mulye, T., Li, Z. H., Le, T., Stephens, J. J., Gomez, S. R., Moy, A. D., Nguyen, J. C., Franz, A. H., Lin-Cereghino, J. and Lin-Cereghino, G. P. (2012) 'Analysis of the 5' untranslated region (5'UTR) of the alcohol oxidase 1 (AOX1) gene in recombinant protein expression in *Pichia pastoris*.' *Gene*, 496(2) pp. 118–127.

Steinle, A., Witthoff, S., Krause, J. P. and Steinbüchel, A. (2010) 'Establishment of Cyanophycin Biosynthesis in *Pichia pastoris* and Optimization by Use of Engineered Cyanophycin Synthetases.' *Applied and Environmental Microbiology*, 76(4) pp. 1062–1070.

Stephanopoulos, G. (2012) 'Synthetic Biology and Metabolic Engineering.' *ACS Synthetic Biology*, 1(11) pp. 514–525.

Stephen, J. D., Mabee, W. E. and Saddler, J. N. (2012) 'Will second-generation ethanol be able to compete with first-generation ethanol? Opportunities for cost reduction.' *Biofuels, Bioproducts and Biorefining*, 6(2) pp. 159–176.

Sun, R., Sun, X. F., Wang, S. Q., Zhu, W. and Wang, X. Y. (2002) 'Ester and ether linkages between hydroxycinnamic acids and lignins from wheat, rice, rye, and barley straws, maize stems, and fast-growing poplar wood.' *Industrial Crops and Products*, 15(3) pp. 179–188.

Szostak, E. and Gebauer, F. (2013) 'Translational control by 3'-UTR-binding proteins.' *Briefings in Functional Genomics*, 12(1) pp. 58–65.

Taherzadeh, M. J. and Karimi, K. (2011) 'Fermentation Inhibitors in Ethanol Processes and Different Strategies to Reduce Their Effects.' *In Biofuels*. Elsevier, pp. 287–311.

Teter, S. A., Sutton, K. B. and Emme, B. (2014) 'Enzymatic processes and enzyme development in biorefining.' *In Advances in Biorefineries*. Elsevier, pp. 199–233.

Tramontina, R., Franco Cairo, J. P. L., Liberato, M. V., Mandelli, F., Sousa, A., Santos, S., Rabelo, S. C., Campos, B., Ienczak, J., Ruller, R., Damásio, A. R. L. and Squina, F. M. (2017) 'The Coptotermes gestroi aldo–keto reductase: a multipurpose enzyme for biorefinery applications.' *Biotechnology for Biofuels*, 10(1) p. 4-19.

Tramontina, R., Galman, J. L., Parmeggiani, F., Derrington, S. R., Bugg, T. D. H., Turner, N. J., Squina, F. M. and Dixon, N. (2020) 'Consolidated production of coniferol and other high-value aromatic alcohols directly from lignocellulosic biomass.' *Green Chemistry*, 22(1) pp. 144–152.

Tran, A.-M., Nguyen, T.-T., Nguyen, Cong-Thuan, Huynh-Thi, X.-M., Nguyen, Cao-Tri, Trinh, M.-T., Tran, L.-T., Cartwright, S. P., Bill, R. M. and Tran-Van, H. (2017) '*Pichia pastoris* versus *Saccharomyces cerevisiae*: a case study on the recombinant production of human granulocyte-macrophage colony-stimulating factor.' *BMC Research Notes*, 10(1) p. 148-156.

Türkanoğlu Özçelik, A., Yılmaz, S. and Inan, M. (2019) '*Pichia pastoris* Promoters.' *In* Gasser, B. and Mattanovich, D. (eds) *Recombinant Protein Production in Yeast*. New York, NY: Springer (Methods in Molecular Biology), pp. 97–112.

Urbanowicz, B. R., Catalá, C., Irwin, D., Wilson, D. B., Ripoll, D. R. and Rose, J. K. C. (2007) 'A tomato endo-beta-1,4-glucanase, SlCel9C1, represents a distinct subclass with a new family of carbohydrate binding modules (CBM49).' *The Journal of Biological Chemistry*, 282(16) pp. 12066–12074.

Van Dyk, J. S. and Pletschke, B. I. (2012) 'A review of lignocellulose bioconversion using enzymatic hydrolysis and synergistic cooperation between enzymes—Factors affecting enzymes, conversion and synergy.' *Biotechnology Advances*, 30(6) pp. 1458–1480.

Vanz, A., Lünsdorf, H., Adnan, A., Nimtz, M., Gurramkonda, C., Khanna, N. and Rinas, U. (2012) 'Physiological response of *Pichia pastoris* GS115 to methanol-induced high level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes.' *Microbial Cell Factories*, 11(1) p. 103-114.

Várnai, A., Tang, C., Bengtsson, O., Atterton, A., Mathiesen, G. and Eijsink, V. G. (2014) 'Expression of endoglucanases in *Pichia pastoris* under control of the GAP promoter.' *Microbial Cell Factories*, 13(1) p. 57.

Vassilev, S. V., Baxter, D., Andersen, L. K. and Vassileva, C. G. (2010) 'An overview of the chemical composition of biomass.' *Fuel*, 89(5) pp. 913–933.

Venkitasubramanian, P., Daniels, L. and Rosazza, J. P. N. (2007) 'Reduction of Carboxylic Acids by Nocardia Aldehyde Oxidoreductase Requires a Phosphopantetheinylated Enzyme.' *Journal of Biological Chemistry*, 282(1) pp. 478–485.

Vogl, T., Gebbie, L., Palfreyman, R. W. and Speight, R. (2018) 'Effect of Plasmid Design and Type of Integration Event on Recombinant Protein Expression in *Pichia*

pastoris.' Master, E. R. (ed.) *Applied and Environmental Microbiology*, 84(6) pp. 1-16.

Vogl, T., Kickenweiz, T., Pitzer, J., Sturmberger, L., Weninger, A., Biggs, B. W., Köhler, E.-M., Baumschlager, A., Fischer, J. E., Hyden, P., Wagner, M., Baumann, M., Borth, N., Geier, M., Ajikumar, P. K. and Glieder, A. (2018) 'Engineered bidirectional promoters enable rapid multi-gene co-expression optimization.' *Nature Communications*, 9(1) p. 3589-3602.

Vu, V. H. (2009) 'High-Cell-Density Fed-Batch Culture of *Saccharomyces cerevisiae* KV-25 Using Molasses and Corn Steep Liquor.' *Journal of Microbiology and Biotechnology*, 19(12) pp. 1603–1611.

Wackett, L. P. (2008) 'Biomass to fuels via microbial transformations.' *Current Opinion in Chemical Biology*, 12(2) pp. 187–193.

Walker, J. A., Takasuka, T. E., Deng, K., Bianchetti, C. M., Udell, H. S., Prom, B. M., Kim, H., Adams, P. D., Northen, T. R. and Fox, B. G. (2015) 'Multifunctional cellulase catalysis targeted by fusion to different carbohydrate-binding modules.' *Biotechnology for Biofuels*, 8(1) p. 220-240.

Wang, X., Wang, Q., Wang, J., Bai, P., Shi, L., Shen, W., Zhou, M., Zhou, X., Zhang, Y. and Cai, M. (2016) 'Mit1 Transcription Factor Mediates Methanol Signaling and Regulates the Alcohol Oxidase 1 (AOX1) Promoter in *Pichia pastoris*.' *Journal of Biological Chemistry*, 291(12) pp. 6245–6261.

Wang, F. and Zhang, W. (2019) 'Synthetic biology: Recent progress, biosafety and biosecurity concerns, and possible solutions.' *Journal of Biosafety and Biosecurity*, 1(1) pp. 22–30.

Wang, Y., Chantreau, M., Sibout, R. and Hawkins, S. (2013) 'Plant cell wall lignification and monolignol metabolism.' *Frontiers in Plant Science*, 4, article no: 220.

Watson, J. D., Baker, T. A., Bell, S. P., Gann, A., Levine, M. and Losick, R. (2013) *Molecular Biology of the Gene (7th Edition)*. Boston, Ma, U.s.a.: Benjamin Cummings. 29(11), pp. 4399-4401

Weninger, A., Fischer, J. E., Raschmanová, H., Kniely, C., Vogl, T. and Glieder, A. (2018) 'Expanding the CRISPR/Cas9 toolkit for *Pichia pastoris* with efficient donor integration and alternative resistance markers.' *Journal of Cellular Biochemistry*, 119(4) pp. 3183–3198.

Weninger, A., Hatzl, A.-M., Schmid, C., Vogl, T. and Glieder, A. (2016a) 'Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*.' *Journal of Biotechnology*, 235, October, pp. 139–149.

Weninger, A., Hatzl, A.-M., Schmid, C., Vogl, T. and Glieder, A. (2016b) 'Combinatorial optimization of CRISPR/Cas9 expression enables precision genome

engineering in the methylotrophic yeast *Pichia pastoris*.' *Journal of Biotechnology*, 235, October, pp. 139–149.

Werten, M. W. T., Eggink, G., Cohen Stuart, M. A. and de Wolf, F. A. (2019) 'Production of protein-based polymers in *Pichia pastoris*.' *Biotechnology Advances*, 37(5) pp. 642–666.

Westers, L., Westers, H. and Quax, W. J. (2004) 'Bacillus subtilis as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism.' *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*. (Protein Export/Secretion in Bacteria), 1694(1) pp. 299–310.

Wong, D. W. S. (2006) 'Feruloyl Esterase: A Key Enzyme in Biomass Degradation.' *Applied Biochemistry and Biotechnology*, 133(2) pp. 87–112.

Wong, S. S. C. and Truong, K. (2010) 'Fluorescent Protein-Based Methods for On-Plate Screening of Gene Insertion.' *PLOS ONE*. Public Library of Science, 5(12) p. e14274.

Wu, J. M., Lin, J. C., Chieng, L. L., Lee, C. K. and Hsu, T. A. (2003) 'Combined use of GAP and AOX1 promoter to enhance the expression of human granulocyte-macrophage colony-stimulating factor in *Pichia pastoris*.' *Enzyme and Microbial Technology*. (The 8th Symposium of Young Asian Biochemical Engineers' Community (YABEC 2002)), 33(4) pp. 453–459.

Wu, S. and Letchworth, G. J. (2004) 'High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol.' *BioTechniques*. Future Science, 36(1) pp. 152–154.

Xu, F., Sun, R.-C., Sun, J.-X., Liu, C.-F., He, B.-H. and Fan, J.-S. (2005) 'Determination of cell wall ferulic and p-coumaric acids in sugarcane bagasse.' *Analytica Chimica Acta*, 552(1–2) pp. 207–217.

Yang, J., Cai, H., Liu, J., Zeng, M., Chen, J., Cheng, Q. and Zhang, L. (2018) 'Controlling AOX1 promoter strength in *Pichia pastoris* by manipulating poly (dA:dT) tracts.' *Scientific Reports*, 8(1) p. 1401-1412.

Yang, S.-T., Huang, H., Tay, A., Qin, W., De Guzman, L. and Nicolas, E. C. S. (2007) 'Extractive Fermentation for the Production of Carboxylic Acids.' *In Bioprocessing for Value-Added Products from Renewable Resources*. Elsevier, pp. 421–446.

Yang, Z. and Zhang, Z. (2018) 'Engineering strategies for enhanced production of protein and bio-products in *Pichia pastoris*: A review.' *Biotechnology Advances*, 36(1) pp. 182–195.

YaPing, W., Ben, R., Hong, Y., Rui, H., Li, L., Ping'an, L. and Lixin, M. (2017) 'High-level expression of l-glutamate oxidase in *Pichia pastoris* using multi-copy expression strains and high cell density cultivation.' *Protein Expression and Purification*, 129, January, pp. 108–114.

Yoshida, H. (2007) ‘ER stress and diseases: ER stress and diseases.’ *FEBS Journal*, 274(3) pp. 630–658.

Zahrl, R. J., Peña, D. A., Mattanovich, D. and Gasser, B. (2017) ‘Systems biotechnology for protein production in *Pichia pastoris*.’ *FEMS Yeast Research*, 17(7), pp. 1-31.

Zeng, Y., Yin, X., Wu, M.-C., Yu, T., Feng, F., Zhu, T.-D. and Pang, Q.-F. (2014) ‘Expression of a novel feruloyl esterase from Aspergillus oryzae in *Pichia pastoris* with esterification activity.’ *Journal of Molecular Catalysis B: Enzymatic*, 110, December.

Zha, Y., Hossain, A. H., Tobola, F., Sedee, N., Havekes, M. and Punt, P. J. (2013) ‘*Pichia anomala* 29X: a resistant strain for lignocellulosic biomass hydrolysate fermentation.’ *FEMS Yeast Research*, 13(7) pp. 609–617.

Zhang, Z., Wang, M., Gao, R., Yu, X. and Chen, G. (2017) ‘Synergistic effect of thermostable β-glucosidase TN0602 and cellulase on cellulose hydrolysis.’ *3 Biotech*, 7(1) p. 54-61.

Zhan, C., Yang, Y., Zhang, Z., Li, X., Liu, X. and Bai, Z. (2017) ‘Transcription factor Mxr1 promotes the expression of Aox1 by repressing glycerol transporter 1 in *Pichia pastoris*.’ *FEMS Yeast Research*, 17(4), pp. 1-10.

Zhang, K., Su, L., Duan, X., Liu, L. and Wu, J. (2017) ‘High-level extracellular protein production in Bacillus subtilis using an optimized dual-promoter expression system.’ *Microbial Cell Factories*, 16(1) p. 32-47.