Exploration of Factors Affecting Recombinant Protein Secretion in *Saccharomyces Cerevisiae*

Thesis submitted to the University of Kent for the degree of PhD in microbiology

Samantha Hobbs

Declaration

No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent, or any other University or Institution of learning.

Samantha Hobbs
March 2022

Acknowledgements

I wish to thank everyone around me who has been part of the whole process. From everybody in the lab, KFG and university to friends and family who have all helped in a multitude of ways I couldn't even begin to list. You know who you are, I won't name you all or I'd end up over the word count!

Specific thanks to Campbell, who enabled me to take on this project, and of course to Tobias who accepted me as his student and was a great supervisor all the way through. There is absolutely no way I could have written this thesis without the patient support from Tobias.

There have been many, many times when I began to suspect I would never finish, or that an experiment would never work, and I am so grateful to have had the amazing support systems I have had to get me through each and every pitfall. Thanks to everyone who has been part of the journey.

Table of Contents

Abbreviations

Abbreviation	Full term
mRNA	Messenger RNA
kDa	Kilo Daltons
OD600	Optical density measured at 600 nm
PCR	Polymerase chain reaction
S. cerevisiae	*Saccharomyces cerevisiae*
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
YPD	Yeast extract Dextrose
SC	Synthetic complete
GPx	Glutathione Peroxidase
PDI	Protein Disulfide Isomerase
ER	Endoplasmic Reticulum
rP	Recombinant protein
RNA	Ribonucleic acid
Gluc	Gaussia Luciferase

Abstract

Production of recombinant proteins is a vital industry to overcome limited and costly natural supplies. The huge industry of recombinant protein production and purification relies on multiple cell types and genetic manipulation tools. *Saccharomyces cerevisiae* is a promising organism to produce recombinant protein due to the ability to secrete recombinant protein into the extracellular media. Secretion into the extracellular media allows for simpler purification of proteins, but is hampered by low yields. In this study we approach recombinant protein production in *Saccharomyces cerevisiae* from different angles to study the impact upon protein yields and secretion capability. Initially looking at the effects of genetic deletions, then tuning of the endoplasmic reticulum and finally consideration of different strains and growth conditions. When this data is taken together, it shows the necessity of tuning individual systems to enable the best conditions, as different cell types react differently to conditions which have been optimised for another. The variation in cell types shows that single breakthroughs which can be applied to all systems cannot be expected, and studies must consider effects among a broad selection of strains to enable universal understanding of its affects upon protein production and secretion as a whole.

Chapter 1 Introduction

1.1 Recombinant proteins and their applications

Recombinant proteins (rP) are produced by transcription and translation of recombinant DNA, meaning DNA which has been formed artificially by combining constituents from different organisms and sources to create something which does not occur naturally in the genome. In yeast this includes enzymes that produce or destroy chemicals (washing powders, industrial transformations), structural proteins such as collagen (patent US11028146B2) and spider silk[1] and food (such as milk protein[2]), and medicines. Recombinant proteins have widespread use and often high value. Genetic manipulation of host cells can enable production of virtually any protein desirable. By identifying and restricting the DNA sequences used to create a protein of interest, it can be inserted into a new organism to create fast and reliable protein production. Producing proteins recombinantly is a huge industry, valued at $125billion in 2020[3]. Natural sources of protein can be limited and lead to low yields, recombinant DNA technology allows this to be overcome and reliable protein yields can be produced to match demand.

Recombinant proteins are produced by inserting the DNA encoding a desired recombinant protein into a host organism or system. Typically this is a cell, but the potential for cell free production of recombinant protein also shows promise[4]. The choice of the correct host organism can be vital in producing the correct recombinant protein as each production system has unique properties which may be exploited or may need to be overcome[5]. The system used for recombinant protein production will also dictate how the protein is folded and any post-translational modifications that occur, and so it is important to consider the best system. Each of the common systems used for recombinant protein production has its own set of unique

advantages and challenges. Understanding the challenges presented by each recombinant protein production system enables informed decisions to be made on the best system to use depending on the functionality of the final protein product.

Recombinant proteins have widespread use and often high value. Rp and genetic engineering can provide large amounts of otherwise scarce proteins, as well as the opportunity to mutate and modify proteins and molecules currently available naturally to produce ones that have novel functionality. As an example of modifying proteins to create novel compounds is flavonoid producing pathways have been expressed in yeast artificial chromosomes in *S. cerevisiae,* and different precursors have been used to diversify the available flavonoids through this technology[6].

1.2 Drugs and antibodies

Many of the highest value products created via recombinant protein production are biopharmaceuticals, of which antibody production is one of the highest value and fastest growing biopharmaceutical products. A 2017 review showed that roughly half of FDA approved therapeutic proteins were monoclonal antibody based, roughly a fifth were coagulation factors, around 1 in 10 were replacement enzymes and the remainders were fusion proteins, hormones, growth factors and plasma proteins[7]. Many of these have uses in oncology, haematology and cardiology, but protein pharmaceutical uses are widespread within other diseases also[7.]

As therapeutic antibodies have been suggested in treatments this wide variety of conditions from autoimmune diseases[8], cancer immunotherapy[9], respiratory diseases[10], to infection[11] and multiple sclerosis[12], production has increased massively in recent years both for the potential application and the financial

prospects. Many monoclonal antibodies are available for human use currently. A 2014 study citing FDA approved biological medicines in the USA[13] contained a table citing 34 monoclonal antibody products, but many more are in the trial stage. In 2010 a large review was published in Nature considering data on 147 monoclonal antibodies which had been entered into clinical study[14]. More recent figures of approved antibody products on the drug market indicate further monoclonal antibody products, some of which are whole molecules and others of which are antibody fragments such as Fabs[15] .

The specific binding of antibodies to their targets is what makes them so desirable as medicines. Rp systems allow the creation of these novel antibodies and fragments. Further engineering can then be done to amplify medicinal use, for example conjugation to drugs that will only activate at the specific antibody target. To use antibodies as human drugs they need to be capable of action within humans without causing unwanted immunogenic effects. Originally murine antibodies produced in living mice were proposed as antibody therapeutics however it was quickly found that rejections occurred due to humans producing anti-mouse antibodies causing the degradation of the therapeutic. Techniques improved using grafting techniques to reduce immunogenicity, however the use of bioprocessing reduces the need to work with “imperfect” animal antibodies as human antibodies can be created and purified from immortal cell lines. Since this many other methods of humanising antibodies have been developed, including phage display cloning of human antibody fragments and transgenic mice[16].

As antibodies are so important to modern medicine and industrially, it is important to create recombinant protein systems which allow the creation of high quality

recombinant monoclonal antibodies. In general the simpler the organism, the cheaper protein production can be achieved which could make a massive difference to cost and availability of these cutting edge therapies, but currently is challenging as antibody production relies on the mammalian antibody assembly machinery. There are now limited examples of engineered yeasts producing full length antibodies[17,18], but in general yeast systems remain currently better suited to express simpler antibody fragments[19], and much of the antibody production remains in mammalian cell lines such as the Chinese hamster ovary cell.

1.3 Natural methods

Some proteins are able to be purified from their natural sources such as milk proteins. Although this method may have advantages such as potential reduced costs, increased safety, and integrity of product, supply of some natural plants and proteins are insufficient to meet demand, and increasingly ethical concerns around usage of animals has driven production of engineered proteins over natural sources. Genomic analysis can identify pathways for the synthesis of these compounds and be engineered into host organisms. One example of this is an extract from *Erigeron Breviscapus* which has now been produced in *S. cerevisiae*[20]. This pattern of the identification of useful medicinal products and identifying the gene or pathway to produce the products before implanting it into a host organism has been utilised in multiple examples[21]. However, many proteins remain largely purified from natural sources including Bovine serum albumin[22], polyclonal antibodies and milk proteins[2].

As sourcing proteins from natural sources can be costly, time consuming, environmentally damaging and can lead to products which are of low quality, purity or yield, methods to increase production and quality have been utilised. By producing

proteins in cellular systems, high quality products can be created quickly to high yields while reducing the environmental impact and improving the quality of the final product. This occurs as recombinant proteins can be created using optimised genetic sequences and produced to high yields in optimised systems and therefore is often preferable to the purification of native protein.

1.4 Host systems for recombinant protein production

Cell types to produce recombinant protein both need to be able to express the protein accurately, and provide a good yield. Prokaryotic cells and simple single celled organisms can create proteins with different modifications and folding patterns which can reduce the quality of the final product. To enable different organisms to produce recombinant proteins, they also need to be able to be transfected with the DNA encoding the protein of interest.

The glycosylation profile of an organism needs to be considered as differing glycosylation occurs in different cell types[23] which can affect the structure and function of the final product as well as become immunogenic. Glycosylation in microorganisms forms mannose N-glycans, and many mammalian cell lines produce N-glycolylneuraminic acid, both of these can be immunogenic in humans. Although prokaryotic cells can be used in recombinant protein production, and provide fast and cheap expression, they do not allow for introduction of human post-translational modifications, and in situations where recombinant proteins are intended to be introduced into the human body may therefore not be the most appropriate hosts for the use in human pharmaceutical markets.

Cell type	Bacterial	Fungal	Mammalian
Growth rate	Fast	Fast	Slow
Cost	Low	Low	High
Post translational modifications	Forms mannose N-glycans which can be immunogenic Incapable of complex post translational modification	Forms mannose N-glycans which can be immunogenic but are capable of some more complex post translational modifications	Most similar to human, can produce N--glycolylneuraminic acid which is immunogenic
Safety	Can be pathogenic	Pathogenic fungi may cause risk but main strains very safe	May become virally infected
Ease of genetic modification	Easy	Easy	Harder

Figure 1: Table to summarise the major differences between cell systems

1.5 *E. coli* and other Prokaryotes

Prokaryotic cells are commonly used in recombinant protein production. For the production of simple recombinant proteins, prokaryotic systems are often the quickest and cheapest expression system to begin production. *E. coli* is a very common system used for protein production that is generally considered to be safe[24] (GRAS status). *E. coli* is fast growing to high cell densities in inexpensive media. *E. coli* is easy to manipulate genetically and it is relatively simple to incorporate labelled or non-natural amino acids into proteins produced. It is also useful as a host organism due to the large amount of experimental data and public knowledge available.

Prokaryotic recombinant protein is not solely produced in *E. coli*, another prokaryotic system which can produce proteins with suitable speed and cost is *Bacillus megaterium* which can secrete proteins that would not be secreted in *E. coli*, and also has GRAS status which is useful when producing proteins for human use. *Pseudomonas*, part of the Archaea domain, has also been used for heterologous protein expression. It has been used to produce oxidoreductases, in this case it is

useful as *Pseudomonas* has an aerobic only metabolism enabling proper folding and stabilization of the proteins. To give further examples a bacteria species called *Streptomyces lividans* has been used for production of secondary metabolites and drug modifying enzymes. This system produces many antibiotics. Myobacterial proteins can be produced in *Streptomyces lividans* which may be difficult to express well in other systems.

The biggest disadvantages of all prokaryotic Rp production is that the cells are often unable to produce higher order proteins with the correct folding and post translational modifications for human use. The advantages and challenges of prokaryotic Rp systems are a huge area of study with new advances all the time[25], however with many eukaryotic proteins, particularly those used in human pharmaceutical markets, needing special post translational modifications, chaperones, and folding mechanisms, eukaryotic cells may be required to produce them. The simple expression systems used in prokaryotic cells are, indeed, often the least costly and fastest methods of expressing and collecting a recombinant protein; however the lack of these, often vital, post translational modifications means they can struggle to produce properly folded and modified, high value human pharmaceuticals, or biologically accurate models for the medical research industry.

1.6 Cultured metazoan cells

Mammalian and insect cells are more complex in many ways, and are often more expensive to culture and manipulate. The complexity of mammalian culture leads to far higher costs compared to bacterial cell culture. One of the biggest advantage of using mammalian cells is that they can produce protein with similar post-translational

modifications to humans, however even with mammalian cells, post-translational modifications can still vary between organisms.

Mammalian cells

Mammalian cells still remain relatively easy to grow and genetically manipulate, particularly Chinese Hamster Ovary (CHO) cells. The complex processing of some recombinant proteins may be better suited to mammalian cells as the necessary pathways are not present in natural bacterial or yeast cells. The use of mammalian cells allows for production of properly folded, modified and assembled proteins[26]. The proper production of these proteins relates directly to function, for example, a recent study showed mammalian produced SARS-CoV-2 protein produced in mammalian cells as a vaccine produced a better response than the same protein produced in prokaryotic systems[27] indicating that mammalian cells are able to produce protein that is more functionally useful in human therapeutic usage.

Alongside producing better quality protein, the use of mammalian cells hopes to avoid issues surrounding immunogenicity, either from modifications left on the protein which cause an immune response, or from elements of the cell remaining in the protein collected. Bioprocessing and recombinant protein production can never fully remove all elements of the host cell; it is believed use of mammalian cell may be safer than potentially pathogenic microbial cells. On the other side, viral loads could remain present in mammalian cells and may be more likely to cause harm to a patient upon receipt. Regulations on production will give strict limits in what may remain as a contaminant in a protein purified for human use; these may vary dependant on methods used and any considered danger of the cells. These regulations are suggested by organisations such as the world health organisation,

and reviewed by safety departments in most countries such as the FDA and the European medicine agency. The safer the cell type used is considered to be, the easier the protein can be to purify to safe and satisfactory standards.

1.7 Fungal cells

There is a long history of using yeast as a means for producing recombinant proteins. Yeast contain eukaryotic organelles, which allow compartmentalisation of metabolic pathways which can be further engineered for distinct benefits[28] including the localisation of important proteins and the maintenance of oxidation and reduction states. *Saccharomyces cerevisiae* has high efficiency homologous recombination, genetic tractability and is a robust fermenter[29], and there are now many methods and applications for it[30]. *Saccharomyces cerevisiae* was the first yeast cell factory, however increasing and evermore specialised needs of protein production pathways and bioprocessing requirements has led to a divergence into many other species as expression systems also[31,32]

Although yeast cells are simpler and cheaper than mammalian cells, many core processes are conserved. This allows cheaper and faster ways of studying processes which then apply to higher organisms due to conservation in evolution. In fact, yeast cells have been used in many examples to enable study of disease mechanisms, particularly in the context of neurological diseases such as Alzheimers[33] and other degenerative diseases[34], but yeast use in modelling diseases has wide scope and includes mitochondrial disease[35] and calcium related diseases[36]. Much of this is possible due to the ease of genetically engineering yeast cells and co expression modules[37,38]. The study of these processes in yeast has

allowed high levels of understanding of yeast processing which has made the manipulation and utilisation of yeast as a host organism for Rp easier.

Recent years have brought many developments to the field of yeast hosts for secretory recombinant proteins[39], but there are still ongoing attempts to improve protein quantity and quality. Tools have become available for improved protein expression via engineering of codons, synthetic signal peptides, integration systems with copy numbers controlled, engineering of the ER, protein trafficking pathway, glycosylation pathways, and cell walls[39]. Yeast glycosylation pathways differ to human ones[40–42]. This can lead to glycan structures on recombinant proteins which can be allergenic or alter protein structure and functionality. Many different attempts have been made to modify yeasts to produce humanised glycans[39]. Within *Saccharomyces cerevisiae*, there is huge variance between strains. For example some will produce more protein, grow at different rates, change in enzyme expression patterns, tolerate stress or ferment in very different ways[43,44.] The differences in yeast glycosylation pathways compared to human cells leading to glycoproteins with high-mannose glycan structures may change the behaviour of a human recombinant protein and require further downstream processing compared to the same protein produced in a mammalian cell. Engineering of the glycoproteins to enable better protein functionality and reduced immunogenicity has been carried out in many yeast systems[45]. Identifying and engineering of final proteins with the correct glycosylation is important to create functional proteins which do not cause immunogenic reactions in humans.

Secretion of produced protein is preferable to intercellular expressed protein as this reduces downstream processing steps. As yeasts do not naturally secrete large

amounts of endogenous protein, using yeast as a host to produce secretory recombinant protein can simplify downstream processes[39]. This prevents the need for timely and costly processes such as cell lysis and the resultant protein has high purity to begin.

Although yeasts have many advantages as host organisms, they also have their limitations. The major limitations of yeasts for rP are inefficient secretion, improper folding, species-specific glycosylation that differs from glycosylation in humans, and non-standard proteolytic processing. . Much work has been done on engineering yeasts to overcome these limitations[39,44,46–48]. As synthetic biology has improved secretion and folding in yeast, systems biology has contributed to greater understanding of cellular mechanisms which enable efficient protein production through choosing targets to design better hosts. Computational models have been used accurately to predict mutants which will have increased recombinant protein production[49]. Metabolic models can be created to predict cell behaviours and improve performance of hosts, such as improvement in the ability of P. pastoris to produce secretory human serum albumin[50]. These models give the basis necessary to increase the utility of yeast as an organism to successfully produce rP that can be useful to humans, however there is still work to be done.

An increase in the overall expression and secretion of recombinant protein in cells such as *Saccharomyces cerevisiae* is desirable for industrial processes as well as illuminating to the functions and inner workings of these cells. The secretory pathway in yeast has been modified in many ways for example to produce higher yields of protein[51], to produce proteins with different glycan profiles[42,52] and to secrete numerous non-native proteins.

As yeast can be industrially useful by the production of high yields of protein products which are human pathogen free and carry out many of the necessary post translational modifications, they can be seen as an ideal protein expression tool industrially. Conversely protein production in *S. cerevisiae* leads to the previously discussed hyper mannosylation, which needs to be reduced to avoid immunogenicity and to increase activity in a human recipient. As hypermannosylation is one of the largest drawbacks in *S. cerevisiae* proteins, leading to reduced half-life and efficiency, modification of the glycosylation pathways can be vital for the field. This has been done in a variety of ways including removal of mannosyltransferases[53], enhancing N-glycan homogeneity[52], and CRISPR/Cas9 engineering[54].

Previously, many strains of baker's yeast have been analysed to identify pathways which enable some strains to produce higher levels of recombinant proteins[55]. As many cellular processes are still unknown, studying genome wide changes as a whole has been used to understand the implications of protein secretion upon yeast strains[44]. One way in which genomic differences can be analysed is by RNA profiling. RNA sequencing has been used to study genome wide transcriptional responses to secretion in mutant yeast strains to study which cellular processes are changed in support of protein secretion[44]. This can lead to generalised observations such as altered energy metabolism in which respiration can be reduced and fermentation increased, the balance of increasing amino acid biosynthesis and reducing thiamine biosynthesis has also been flagged as important in optimising producing strains[44].

The attributes of different common yeast species used to produce recombinant proteins were reviewed in 2014[48]. It is important to understand the key attributes of

each species of yeast before selecting one as a host organism for the production of a recombinant protein. One useful attribute in a host species is a fully sequenced genome as this can be important in choosing genetic manipulations used to improve the system. The complete sequencing of the *S. cerevisiae* genome in 1996[56] was the first eukaryotic species to be fully sequenced. The genomes of other yeast species were not sequenced and available in the public domain until after the millennium[48], giving *S. cerevisiae* genetic knowledge almost a decade head start. *S. cerevisiae* is the most well-known yeast, and has GRAS status, meaning the FDA generally recognises it as safe to use. It is a simple eukaryote and it is possible to manipulate the genetics of S. *cerevisiae* by a wide range of techniques, or alternatively many mutant strains are already available[57].

Although many systems are based on *S. cerevisiae*, it is only one of many yeasts used in protein production. Different yeast species have varied glycosylation pathways leading to different structural glycans present on final proteins[58]. Fungal systems also differ in ER and Golgi body structure which can give a secretory pathway more similar to higher eukaryotes[39,59].

Pichia pastoris is another commonly used strain in the bioprocessing and bioengineering industries. *Pichia* has GRAS status and allows high dilution rates and biomass yields. The growth rate in inexpensive media is still high and vectors have been developed which can integrate, enabling high genetic stability of recombinant proteins. *Pichia* leads to less hypermannosylation than *S. cerevisiae* with no 1,3-linked mannose residues, meaning it has the potential to create proteins which are less immunogenic to humans. The genome sequence is also now complete[48].

Some of the other emergent yeast hosts for recombinant protein production are Crabtree negative (i.e. they do not start producing ethanol over biomass) which enables higher biomass than *S. cerevisiae* and can use a wider selection of carbon sources[39]. *For example H. polymorpha* can grow using only methanol as a carbon source, resisting temperatures up to 50^{o}C and stressors such as oxidative stress and heavy metals[60]. *Hansenula polymorpha* has now got a comprehensive toolkit to enable production of secretory proteins[61]. *Hansenula polymorpha* is a further yeast system which can be exploited with current GRAS status, with strong promoters this yeast allows for high yields with stable integration of foreign DNA into its chromosome. *Hansenula* also has a lower level of hypermannosylation compared to *S. cerevisiae* with no terminal 1,3-linked mannose residues present and the genome is fully sequenced.

Generally, the core functions leading to protein secretion are conserved through from yeasts to higher eukaryotes, however there are significant functional differences[62]. Regardless of which species is chosen, it is vital to correctly screen and select the strain to be used[61] for optimum outcome.

1.8 Filamentous fungi

Filamentous fungi are of interest primarily due to their ability to secrete large yields of protein[63]. Filamentous fungi produce some extracellular enzymes naturally. One reason to use filamentous fungi is their ability to produce high amounts of extracellular protein[64] with examples cited such as *Aspergillus niger* which has been noted to produce up to 30g/L or glucoamylase and *Trichoderma reesei* which has been shown to produce 100g/L of extracellular protein. The production of extracellular protein is hugely useful for downstream purification as there will be

fewer secreted proteins than intracellular ones, and so by avoiding lysis of the cells, the secreted protein enters purification pathways at a higher level of purity.

Filamentous fungi, such as *Aspergillus, Trichoderma, Penicillium* and *Rhizopus*, can produce proteases which can then degrade the recombinant protein being produced. Protein glycosylation patterns differ in filamentous yeast[65], compared to other cell types, and so end products are less similar to mammalian cells. There are, however, synthetic pathways which have been established in filamentous fungi which have distinct advantages, and with adequate bioengineering these pathways could either be transferred to new host cells which are more favourable for recombinant protein production, or engineering the filamentous fungi to be more suited to recombinant protein production[64]. Filamentous fungi have been suggested to have certain disadvantages including a low frequency of transformation, risk of morphological defect and modification of proteins either due to native protease activity or local pH.

Utilising the right method of genomic expression can enable production of the non-native proteins of interest, for example whilst *Saccharomyces cerevisiae* has been engineered to produce penicillin using genetic manipulation of genes taken from filamentous fungi, leading to a large increase in yield and therefore enabling bakers yeast to quickly produce a non-ribosomal peptide antibiotic[66], penicillin is classically made in filamentous fungi[67].

1.9 *Saccharomyces Cerevisiae*

The use of yeast to produce high value products has been of great interest for a long time[68,69]. As previously mentioned, sourcing proteins from natural sources can be costly, time consuming, environmentally damaging and can lead to products which

are of low quality, purity or yield. This led to the development of cellular systems creating high quality products with high yields whilst reducing environmental impact and improving final product quality through the use of yeast. Humans have centuries of experience using yeast in different contexts, and so manipulation and handling of yeast is well documented and hence *S. cerevisiae* is generally regarded as safe to use. Now many mutant strains of *Saccharomyces* are available which have been engineered for favourable characteristics. For example, whilst natural strains use hypermannosylation leading to the addition of immunogenic terminal α-1,3-linked mannose residues, engineering has been done to change glycosylation to less immunogenic forms[45]. Synthetic biology is advanced in *S. Cerevisiae*[70] with multiple proven methods of genetic manipulation. Promoters, terminators, transcriptional regulators, post-translational regulations, synthetic genomes and drug production have been well defined in *S. Cerevisiae* [70] through years of research to allow these manipulations leading to a better host organism.

S. cerevisiae is a good host for recombinant protein production as it is the most understood fungal species, there are inherent advantages to its use as a recombinant protein expression system. Many recombinant proteins are produced by *S. cerevisiae* already[71]*. S. cerevisiae* grows well on simple media at large scale, whilst still enabling production of proteins with eukaryotic cell post-translational modifications*.* Growth conditions, genetic manipulation and protein secretion pathways are all well understood. Previous studies can give vital information on how to engineer strains to improve yield. Analysis of mutant strains of *S. cerevisiae*, via RNA sequencing, with higher protein secretion capacity have shown higher levels of fermentation, changes to amino acid biosynthesis and reduced thiamine biosynthesis[44].

Overall, *S. cerevisiae* has many benefits as a producer of secreted recombinant protein. Compared to mammalian cells, growth is fast on inexpensive medium, and compared to bacterial cells yeast has the ability to make some more complex post translational modification. A major advantage of *Saccharomyces cerevisiae* is the understanding within the pathways described above and the toolkits available to genetically manipulate cells so that the shortfalls can be overcome.

1.10 Pathways of recombinant protein production in *S. cerevisiae*

To produce a recombinant protein it is necessary to insert DNA encoding the protein of interest into the host organism. This DNA can then utilise host transcription and translation machinery to be encoded first into mRNA and then into a polypeptide chain. The yeast endoplasmic reticulum is then used to correctly fold the recombinant protein and signal sequences dictate where it is then directed.

The ability to easily genetically manipulate a host organism to express protein has a large impact on productivity. For example, the ability to engineer strains with codon optimised sequences can improve protein production. Codon usage can vary between species, this is known as codon usage bias. Some codons (optimal codons), are decoded by more abundant tRNAs and are therefore preferred for use in highly expressed genes. Editing a sequence to match the host codon bias can therefore achieve higher levels of recombinant protein expression, however this is not the sole determiner of rate as interactions between codons and wobble pairing also impact translation elongation and efficiency[72]. Codon usage can also affect ribosome speed and protein folding, and slower decoding can lead to mRNA instability[73].

Traditionally, codon optimisation was based on individual codon usage bias, which relies on the usage preference for codons in coding, but codon pair context bias has also been considered more recently which looks at organisms preference for specific codon pairs to increase protein expression[74]. With constant advances is computational biology, the ability to model synthetic genes for translation speeds is ever improving. Codon pair context bias has been compared to individual codon usage bias for secretory protein production of C*andida antartica* lipase B where it was found codon pair context bias was more relevant[75]. Many factors have to be considered for successful codon optimisation including GC content of host genome, secondary mRNA structure and the tRNA pool[76].

1.11 Delivering DNA (vectors/ integration)

Not all microbes have the natural ability to take up foreign DNA and so a variety of methods have to be employed to enable different cell types to express a gene of interest[77–79]. Recipient cells can sometimes damage DNA constructs in the attempt to transform cells using endogenous restriction modification methods[80,81], some of these difficulties can be overcome by DNA methylation which has been shown to improve the successful transformation rate[82,83]

There are currently several strategies used to integrate foreign genes into host cells. Firstly DNA can be added with selection markers. These can be used to screen for strains containing the target DNA. It must be noted, however, that selection markers can hamper integration due to limited numbers of dominant markers, so marker-less integration can be preferred, especially when multiple fragments of DNA wish to be inserted. Secondly, optimisation of homology arm lengths can be used to insert DNA into a host. Homology arm (HA) lengths help recombination efficiency. The length of

this depends on the species but *S. cerevisiae* may only need 50BP. DNA can be integrated using sequence specific introduction of double strand breaks. Some programmable nucleases can create a double stranded break at a directed site. These methods include CRISPR/Cas9. Finally introduction of foreign recombination systems can be used. This only needs short homologies and allows DNA integration in double stranded DNA.

Homologous recombination can allow DNA exchange between regions of identical sequence. Homologous recombination is a highly utilised method in inserting DNA into yeast hosts. Homologous recombination methods need regions of homology inserted via PCR to the target DNA and the linearised vector. PCR products can then be transformed into yeast and the vector can be assembled through homologous recombination as by creating DNA fragments with homologous sequences to the host, the host DNA can be exchanged for the recombinant sequence desired**.**

Non homologous end joining can be used to repair double stranded DNA breaks without the need for homologous templates in cells and is part of the natural repair systems. This results in random insertion or deletion of nucleotides and can lead to translocation of the loci. This can be used to insert DNA into a host genome, or to create mutations of specific genes within the host genome. Microhomology mediated end joining is a typically error prone system used to repair double strand breaks, this uses micro homologous sequences and often results in random DNA mutations[84]. This method can be manipulated using short homologous sequences to flanking the desired gene locus to integrate DNA sequences into a host[85]. This method is often used in mammalian cells and less so in industrial microorganisms[86], but could be

used in tandem with CRISPR/Cas9 in the future in organisms which lack homologous recombination.

There are currently a variety of vectors and selection markers. These can integrate at a specific locus by having plasmid with sequence that is slightly mutated in cells, e.g. plasmid with ura gene inside inserted into ura locus of cells with slight ura deletion. When this happens recombination can occur between the tandemly repeated sequences (i.e. *URA3* and *ura3* flanking the plasmid sequences), leading to loss of the integrated DNA fragment such as described here[87]. As recombination can occur in different places, the selection of the strains with marker does not guarantee the presence of DNA. Recombination can lead to flanking repeat genes such as at the ura site which can lead to URA-inserted DNA-ura. The flanking repeats can then recombine and excise the gene inserted. This means the expression is not stable over time. Integration at HO locus has been shown to reduce this risk of tandem repeat[88]. HO codes an endonuclease which is not required for growth and many lab strains have a mutation already at this locus.

It is important in producing a protein in yeast to choose the correct DNA vector. Vectors used must have a target sequence for homologous recombination/integration, a multiple cloning site for foreign gene insertion, a promoter, a selection marker and a secretory signal[89]. Yeasts have a variety of cell vectors such as episomal vectors. These episomal vectors can replicate in cells by autonomous replication. Vectors can also be integrated into the host genome and this gives higher stability, but often a low copy number compared to episomal vectors. Centromeric plasmids can also be used in yeast and they exploit

endogenous replication and chromosomal segregation to exist like a mini-chromosome[90].

Recombinant technology allows genes to be edited to produce the best possible protein and yield. There is a wide variety of genetic manipulation tools to help optimise yeast production of protein, and also optimisations can occur in the environment surrounding the growing cells. In 1991 it was shown that fusing wild type genes from an IgK chain mouse protein to a peptide containing “yeast-preferred” codons improved yield 50 fold[91]. The use of codon optimization has been thoroughly reviewed[92], and there are many reports of extreme translation speed increases due to engineering DNA sequences optimal for the available transfer RNA in the species being used.

Of the 2 major pathways for recombinant protein production in yeast (plasmid- based and chromosomal integration) generally integration of a gene is preferable over plasmid based gene expression[93], particularly for stable industrial protein expression. Plasmid expression can cause genetic instability, segregation instability and structural instability which implicate the final yield. Plasmid based expression is the easiest way to manipulate and regulate compared to integration, however it has disadvantages such as limited cloning sizes and the need for selection pressure. Plasmid based expression therefore gives a fast, easily regulated method of protein expression, compared to integration giving a long term stable method. As plasmids have inherent genetic instability and allele segregation, cultures can end up with variations from cell to cell protein production[94], integration is therefore preferable for the long term, reliable protein production required to pass thorough regulatory checks within industry. Chromosomally integrated constructs tend to give more

consistent gene expression as they do not vary in copy number or segregation[95], however there can be a metabolic burden of maintaining the genetic construct and over production of the final product. There are now a variety of methods of genomic integration in microbial expression models[93].

Integration, although preferred, can lead to altered gene expression if regulatory elements are interfered with, or disrupt certain genes due to integration into protein coding regions. To avoid non-targeted integration having unwanted secondary effects, it is preferable to insert genetic constructs into a gene locus which does not affect normal growth or function of the organism, such as regions on the *S. cerevisiae* chromosome that do not affect growth[96]. It has been shown that the method of protein expression (whether through plasmid or integration) can affect the final protein functions and properties[96.]

Forming double stranded breaks in genome can target integration[87], these same methods can be used to rescue chromosomal alleles on plasmids for subsequent molecular analysis. This is important as the position chosen to integrate a gene into a host can have a large effect on the expression levels[97–99]. The changes in expression dependant on loci are due to DNA compaction, distance to DNA replication initiation site and regulatory factor availability. Positioning the integrated genes in a non-coding genomic position can avoid disrupting regular gene function e.g. in part of the rDNA gene[100], or within the Ty retrotransposon repeats[101]. To ensure there is stable and adequate protein expression, DNA introduction to multiple sites is ideal over single insertions, although the relationship between copy number and yield is not necessarily linear[102], and so levels of integration must be carefully considered.

Synthetic biology and genetic engineering have been advanced by multiple techniques such as Golden Gate assembly and CRISPR/Cas9. These techniques can be used to encode genetic data into yeast systems which can be manipulated to produce recombinant proteins. Synthetic biology is advanced in *S. cerevisiae*[70] with multiple proven methods of genetic manipulation. Promoters, terminators, transcriptional regulators, post-translational regulations, synthetic genomes and drug production have been well defined in *S. cerevisiae*[70].

'Golden Gate' cloning attempts to solve the problem of recombination site sequences elongating the final protein. Type IIs restriction enzymes are used which cut on the outside of the recognition sequence and cleavage sites can be designed to ensure that ligases can repair the DNA without incorporating the original restriction site. 'Golden Gate' cloning uses this to create a simple and efficient genetic engineering/recombinant cloning method without adding extra codons, and therefore amino acids, to the host organism[103].

Yeast artificial chromosomes have also been used to develop novel pathways from drug production using combinatorial genetics[104], allowing *S. cerevisiae* to produce novel compounds which may be used in current drug discovery methods.

As well as genetic manipulation of the DNA or RNA sequence used to create the protein of interest, or methods of improving the host organism to better produce, enzyme behaviour can also be modified to improve recombinant protein production[21]. This can help localise enzymes to key sites, prevent toxic intermediates and the sequestering of product into other cellular pathways. Compartmentalisation of enzymes can lead to faster and more efficient reactions within the cell. As proteins can be compartmentalised into organelles to enhance productivity, but may lead to

unwanted interactions and functional alterations, encapsulins have been used in *S. cerevisiae* to create synthetic non-endogenous organelles proving nanoscale compartments for protection and localisation of proteins[105] .

Expression of recombinant proteins can also be engineered in a manner that allows control of expression. For example riboswitches and ribozyme switches have been engineered to form a complex with a ligand to prevent ribosome binding and moving, this prevents the expression of the gene. This can also be used to prevent pre-mRNA splicing inhibiting gene expression[106]. The riboswitch is an mRNA regulator which adopts a defined structure known as an aptamer. The aptamer is in the 5' untranslated region of the mRNA. When the ligand binds this region it impedes normal ribosome activity. Riboswitches have been engineered in *E. coli* to act as pH sensitive genetic devices to control gene expression dependant on environmental pH[107]. This enabled the engineering of cells to tolerate acids. In yeast a riboswitch has been applied to allow control of pre-mRNA splicing. This can be done with multiple aptamers to give further control of gene expression and splicing[106]. It is possible to engineer a variety of ligands and aptamers resulting in engineered riboswitches which can respond to a ligand of choice[70]. Ribozyme switches, comparatively, contain a aptamer sensor domain and a ribozyme regulatory domain to control genetic expression. These ribozyme switches have been used in a variety of eukaryotic cells[108].

Generally, when cloning genes into yeast, they are first cloned into bacteria for amplification. Not all genes can be cloned into a bacterial host without toxic side effects, and as so many gene cloning methods rely on bacterial cloning to amplify DNA, this can cause problems. Recently, methods have been developed which

enable cloning of integration of expression cassettes within *Saccharomyces cerevisiae* without the need to bacterial cloning[109]. This has enabled integration of genes into yeast without bacteria either with modified Gibson assembly or direct assembly and integration of linear PCR products.

Although integrated gene expression is much more stable compared to extra chromosomally expressed, as there is less variability in copy number, this depends on method used to integrate. Transposition-mediated integration and gene duplication amplification are thought to be less stable versions[93]. Chromosomally integrated gene expression can be weaker than plasmid based expression. To counter this it is possible to put multiple repeats of the gene into a locus, however without continued selective pressure, the repeated genes may be lost over time[110].

If multiple copies of a gene are inserted, it can be important to know how many. There are some methods which can detect how many copies of a gene have been integrated into the host such as qPCR, DNA microarrays and next generation sequencing, giving rise to various methods of analysing gene copy number[111]. Often integration of DNA creates a unique join point which can then be detected.

Shuttle vectors can be used to integrate gene expression. These vectors can be maintained in both E*. coli* and *S. cerevisiae* The makes construction, analysis and amplification of plasmid easy in *E. coli* which can then can be transformed into yeast where genes can be cloned in bacteria. For those that cannot be expressed in bacteria, methods have been developed that skip bacterial steps and gene cassettes are assembled and integrated directly[109]. In one study to achieve this 3 PCR amplified fragments were assembled by homologous recombination. Fragments

were generated with overlapping sequences in the 5’ sequence of the primers to enable assembly of the full DNA fragment within the yeast cells.

1.12 CRISPR-Cas based systems for DNA delivery

In CRISPR-Cas based systems, an RNA nuclease is guided by CRISPR and CRISPR associated (Cas) systems to allow adapted immunity. In heterologous organisms, the CRISPR-Cas9 system from streptococcus pyogenes was made functional by introducing components into cells[112–115], This can introduce double stranded breaks at specific loci. These double strand breaks can then become an insertion site for recombinant DNA. CRISPR is ideal for yeast to create high efficiency gene knock ins and knockouts without a selection marker as the double strand breaks promotes DNA recombination strongly in yeast[116]. Several studies have shown extremely high efficiencies in gene disruptions using CRISPR systems[117].

CRISPR-Cas systems can be class 1 or class 2. This is determined by the configuration of the effector modules. Class 1 systems have multi-subunit CRISPR RNA – effector complexes, whereas class 2 have a single protein carrying out the function of the effector complex. Within these classes there are a further 5 types and 16 subtypes[118], with different tools for different applications[115]. Type 2 systems, as the more compact class, require RNA guided nuclease such as Cas9, alongside crRNA, tracrRNA, and RNaseIII. Cas9 is induced at the target DNA site as 2 RNA molecules hybridise and then can cleave the target DNA with a protospacer adjacent motif (PAM) sequence. To express this in yeast multiple guide RNAs (gRNAs) can be expressed from a single plasmid containing multiple gRNA cassettes or different crRNAs for multiple gene disruptions[117].

Genome editing can lead to off-target effects. This can happen if additional sites are cleaved by the Cas9 and leads to mutations caused by non homologous end joining. The probability of mutation in yeast due to these mechanisms are low as yeast doesn't often use non homologous end joining to repair double strand breaks and so negligible off target effects are usually observed, with one study showing a similar level of genome variability in CRISPR-Cas9 engineered strains compared to wild types[119].

The CRISPR system, now well known for its ability to precisely manipulate DNA, has been applied to yeast species in multiple studies[54]. This has also been developed to use Cpf1 for genome editing of *Saccharomyces cerevisiae*[120]*.* The Cpf1 system described by R. Verwaal et al in 2018 claims to be a quick and reliable method for introducing donor DNA into the *S. cerevisiae* genome[120]. Genomic engineering has also been carried out in more automated environments[121] to enable further application and scale-up as recombinant proteins and products of genetically modified organisms become more popular and necessary to enable production of protein products to markets. Many different CRISPR systems have been used in yeast to modify the host genome.

CRISPR-Cpf1 is a type V CRISPR system which has been considered for use in yeast[120,122] as well as humans[123]. CRISPR-Cpf1 recognises PAM sequences rich in T bases, compared to Cas9 which would struggle to target AT rich sequences[124]. Cpf1 also contains an RNaseIII activity[124] which can be used for further gene engineering, compared to Cas9 which can only be used to induce double strand breaks. CRISPR can increase transcription of target genes and can be used with Cpf1[122].

The crRNA and tracrRNA can be fused via a linker to form a single guide RNA making a simpler process, and there are also examples of gRNA expressed as a transient fragment to deliver the gRNA quickly and simply[125]. Cas9 of the *s. pyogenes* recognises 5'-NGG-3' as the PAM in the target sequence. The gRNA can then bind target DNA using a homologous sequences with the PAM sequence at the 3' end of the binding sequence.

Generation of double stranded breaks using the CRISPR system can increase homologous recombination efficiency, this has been particularly noted in integration of single genes and pathways[116,126], but also been used to allow multiple integrations[127]. Marker less platforms have also been developed using Di-CRISPR (delta integration) for integration of metabolic pathways across multiple loci, using native Ty retrotransposon delta sequences [101]. Ty retrotransposon sites are often unstable and so may be sub-ideal for usage as genomic integration sites for recombinant protein production. The wicket system was developed to be integrated at a designated loci to allow integration of DNA when treated with a nuclease[128]. This should allow highly efficient integration without the need for selective markers as well as the ability for simultaneous integration of different genes, in various copy numbers. Other similar methods have been developed which allow higher control of copy number[129] using DNA landing pads which introduce a CRISPR-Cas9 integration system.

DNA assembler is a method which aimed to allow targeted integration of multigene pathways with a selection marker[130], this uses interlaced overlaps introduced into a gene expression cassette, a helper integration fragment is then used to remove

inserted genes such as ura3, the cassettes are then co-transformed using electroporation allowing assembly of the pathway into the chromosome.

The δ-integration method, which is a multi-copy integration method, can be used to engineer yeast cells to produce recombinant proteins[131]. Conventional δ-integration has been shown to integrate fewer copies of a gene compared to when it is combined with CRISPR[131]. The system works by a pre-breakdown of the δ-sequence on the yeast chromosome by the CRISPR system before integration.

To improve yield of a target protein or substance, the carbon flux to a competitive pathway can be suppressed, however many of these pathways are necessary for the functioning of the yeast and so cannot be completely destroyed. CRISPR interference (CRISPRi) can potentially be used to suppress pathways containing essential genes. CRISPRi interferes with transcriptional elongation, RNA polymerase or transcription factor binding as an inactive Cas9 and gRNA bind to the DNA recognition complex[132–135].

CRISPR-Cas systems combined with the intrinsic double stranded break repair pathways now enable versatile genome editing in *S. cerevisiae*[93,99,101,120,121,136].

1.13 Gene Integration sites

The location of the gene will influence expression, as this can alter compaction of DNA, the distance to the initiation of replication site, and the availability of regulatory factors[137]. Within *Saccharomyces cerevisiae*, reporter gene levels have been shown to vary widely dependant on chromosomal location[97,98]. Differences in the expression of genes at different positions is known for example it has been found that the different loci the gene is integrated into can vary the intensity of red fluorescent

protein 13 fold[98]. It is believed that changing the promoters and carbon sources can even have less effect on gene expression than chromosomal position[98]. Genomic location has even been shown to be conserved and important in certain non coding sequences[138]. We are aware of many ways in which chromosome structure can impact gene promoters[139,140], DNA replication[141] and transformation efficiency[142]. When reporter GFP gene was placed at different locations in yeast, mean protein expression levels changed by up to 15 times and expression noise up to 20 times[137]. This shows the importance of considering gene location when expressing recombinant protein in yeast.

Genes repositioned to be near telomeres have been shown to drastically change in expression[143]. Gene knockout collections have been used to analyse position effects and shows that chromatin differences are associated with gene position and therefore change activity. It is also believed that chromatin regulation is governed by interactions between chromatin and genetic factors[143]. Essential genes generally have low expression noise[137], and so areas close to essential genes are likely to be good target locations if stable expression is needed.

Some integration loci have been analysed individually. Integration at HO locus has been shown to have no effect on growth[88] . This may be important in highly regulated systems where it is important not to interrupt the cell metabolism. Choosing an area of interest in which to integrate genes may be very specific to the individual gene, setting and application.

Methods can also be employed outside the main genome of species. For example mitochondrial DNA can be edited[144]. As many organelles such as mitochondria are impermeable to RNA and DNA, it can be difficult to edit the DNA in methods such as

CRISPR. Although not a common method of expression, the ability to express proteins in the yeast mitochondrial DNA is not new through plasmid based systems[145], however more recently microprojectile transformation has been used to insert plasmids containing Cas9, guide RNAs and donor DNA into organelles[144]. Organelles such as mitochondria produce most of the cellular energy, and play a large roll in cellular metabolism. As organelles already produce some of the cells metabolites such as amino acids, lipids and nucleic acids, they may be used to synthesise important biotechnology products with, for example, chloroplasts in plants already able to synthesise therapeutic proteins[146]. Compartmentalisation of production into organelles may also increase yields by reducing dilution in the cytoplasm[147]. Organelle DNA can allow high expression and high copy number. Homologous recombination activity is still relatively high in organelles but the use of CRISPR allows precise genetic engineering. Long term this could even be used to treat human mitochondrial disease. Double membranes encasing organelles can post an issue due to prevention of nucleic acid import. As CRISPR requires a selection of guide and donor nucleic acids, this can make the application difficult. Using plasmids to replicate in organelles containing the expression of the Cas9, guide RNA, donor DNA and a marker allows CRISPR systems to be assembled in the mitochondria[144]

Improvements in understanding of both host organisms and DNA technology due to further research has enabled purposive changes to original sequences, or addition of extra sequences to further improve proteins produced.

1.14 a Brief Overview of Protein Synthesis

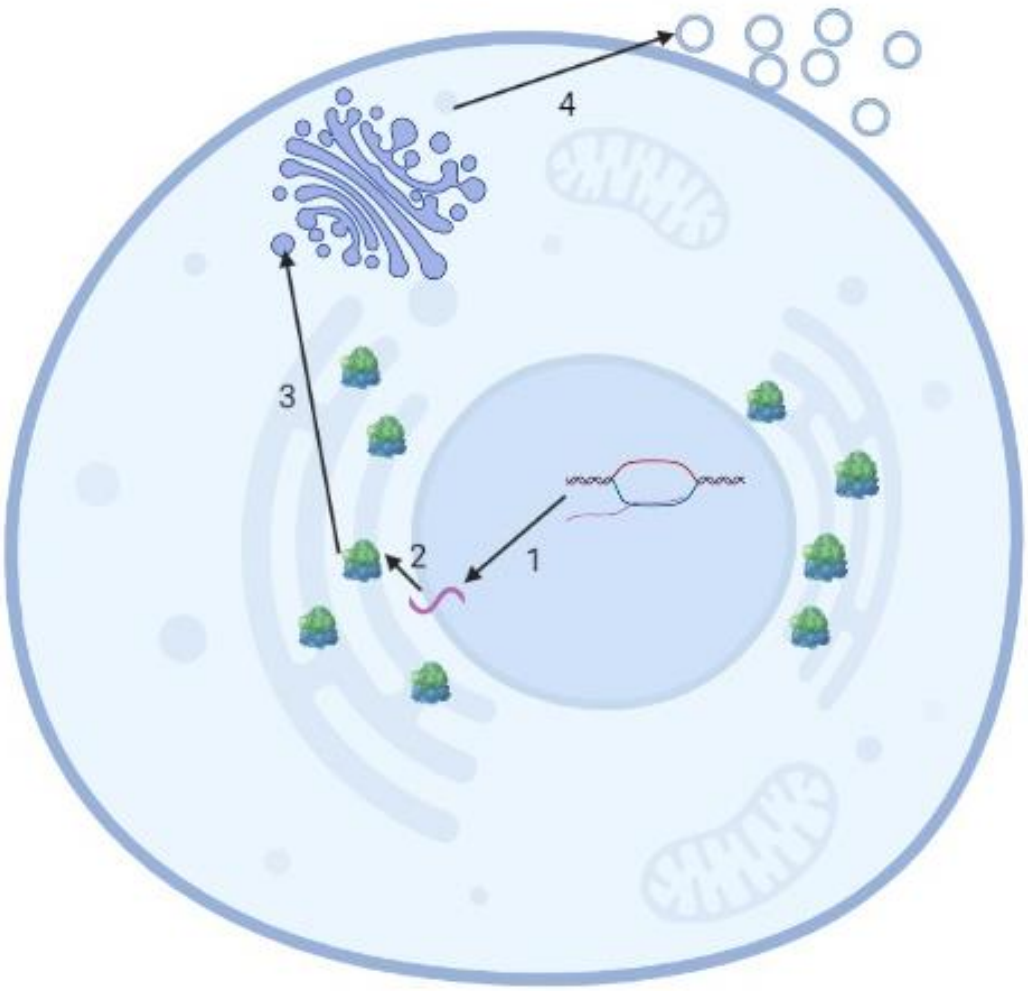

Figure 2: In yeast protein synthesis DNA is first transcribed into mRNA (1) to allow exit through the nuclear pore. This RNA can then be translated at a ribosome on the rough endoplasmic reticulum (2). The protein is then transported to the golgi body (3) which regulated post translational glycosylation before the protein can be secreted from the cells.

In yeast DNA is transcribed into mRNA for translation by the ribosomes into a polypeptide unit. The polypeptide then undergoes binding to BiP to ensure correct folding within the endoplasmic reticulum. PDI is present within the ER to allow

isomerisation and formation of disulphide bonds in the formed protein. At this point if the protein is correctly folded it can be transported to the golgi body for post translational glycosylation and then can be secreted from the external cell membrane.

The first step in production of secretory protein is transfer through ER membrane either ribosome coupled (co-translationary) or uncoupled post translation. This depends on the hydrophobicity of the signal sequence peptide[148]. At the ER membrane Sec61 recognises the signal peptide and forms a channel[148]. BiP binding to the peptide improves efficiency of transport. Inhibiting this translocation pathway has been noted as useful in anticancer and antimicrobial treatment[149]. *S. cerevisiae* uses 2 different translocation pores, Sec61 and Ssh1, The Sec61 pore has 3 units of which the Sec61 subunit is the largest. Ssh1 is a vital and functionally distinct complex[150] of similar function, which also interacts with ribosomes via the 28S ribosomal RNA. Post-translational translocation may be slow and therefore cause a bottle neck in production upon translation, translocating proteins post translation can therefore be favourable in fast growing species[151]. In S. *cerevisiae*, approximately 30% of Sec61 complex is ribosome-associated, however this varies between species[152]. Proteins in this pathway are released from the ribosome and remained unfolded using chaperone proteins, then the heptameric SEC complex is used to translocate across the ER membrane[62].

Sec63 has a few roles in translocation, it can bind Sec62 and stabilise the SEC complex in post-translational transport, it can cooperate with Kar2 in the role of gating the translocational pore, and it can assemble the translocation complexes. The polypeptide is pulled into the ER using a ratcheting mechanism which involves

Kar2 and co-chaperones. The molecular chaperones are believed to bind sequentially to different regions of the same polypeptide and working in a nucleotide binding and exchange manner[153]. Molecular chaperones are present during protein folding, with each organelle utilising different chaperones, although chaperones from multiple compartments may be used to create the final folded protein. Heat shock proteins such as Hsp70s assist in protein folding, degradation, translocation and interactions with functional diversity the driven by J proteins[154]*. S. cerevisiae* contains 2 types of HSP70 chaperones in cytosol and one in the ER (Kar2).

When proteins build up either through misfolding or overload, the ER lumen becomes burdened and so the ER becomes stressed activating the unfolded protein response pathway. The ER is able to degrade misfolded proteins under conditions of the unfolded protein response[155]. Many chaperone proteins and redox enzymes within the ER have been manipulated in attempt to enhance secretion in yeasts[156], but most effective strategies combines multiple engineering strategies to the secretory pathway[51]. As the production of secretory protein by yeast requires trafficking of proteins, both to the ER, from ER to golgi and from golgi to membrane, the modulation of these vesicle trafficking pathways can improve protein secretion in yeast. SNARE proteins are often required in membrane fusion events as they allow fusion between protein transport vesicles, organelles and the plasma membrane. Overexpression of certain SNARE proteins has shown to increase secretion in studies[157]. ER to golgi vesicles are generally COPII coated and bud at the ER exit sites. One of the major proteins of this type of vesicle, Sec16, has been shown to enhance secretion by increasing ER exit site numbers which then leads to a decrease in ER stress[158]. Larger screening studies have identified several proteins which, when mutated, drastically impact secretion capacity[46,159] .

Metabolic stress is a big consideration in using yeasts as host organisms. Bioethanol is a promising industry to reduce reliance on fossil fuels and is one of the promising possible uses for yeast in the modern world. Biomass can be converted into bioethanol using enzymes. Yeast can be used to produce starch hydrolysing enzymes, this requires recombinant strains[160–163]. Expression of these strains can lead to metabolic stress on the cell, and so growth and protein yield can both end up negatively affected. Growth parameters in yeast may be significantly altered to ensure recombinant protein production, as this can shift the energy input of the cell from growth and other metabolic activities to protein production. However some strains do not show metabolic burden, and can show significant alterations in metabolomic profiles without changes to cell viability[164]. Here multiple delta integration of glucoamylase in an industrial yeast was used and the metabolism was examined using Fourier Transform InfraRed Spectroscopy assays (FTIR) with or without ethanol. Metabolites were then examined by LC/MS and indicate that genetic engineering of yeast can have huge impact on the metabolome.

Some vacuole mutants have also been shown to produce recombinant proteins more efficiently – perhaps due to presentation of mis-localisation to the vacuole[165,166]. The vacuole is used within yeast to bulk degrade misfolded proteins. To produce quality recombinant protein in yeast it is still necessary to ensure proteins are properly folded, however it is reasonable to assume a small proportion of correctly functioning protein will be mis-located to the vacuole, particularly under stress conditions from the UPR. With correct engineering, cells should be able to more efficiently and accurately process these proteins to increase yield.

Protein production can also involve synthesis of enzymes within a pathway to produce a product, rather than the protein being the end product itself. Opioids have been successfully produced in yeast[167], expression of 2 enzymes was necessary to enable full biosynthesis of the full opioid within the yeast system. A Cas9-based method was used to improve enzyme expression, building a library of sgRNAs, promoters and tags to improve expression of a problematic enzyme[168]. Polyketide synthase subunits from model systems have been assembled in yeast to produce a diverse library of compounds in *S. cerevisiae*[169]*.* Yeast have been used to produce signalling biosensors and reporter cassettes to transmit signals inside yeast cells upon exposure to create a model system for the angiotensin II type 1 receptor[170], as this signalling does not respond naturally in yeast cells as it does in human cells (*S. cerevisiae*).

In yeast a secretory signal peptide can be used to express proteins which are secreted outside the cell, this signal sequence is frequently MFα1 which has a role in secretion of mating factors[171]. Nearly 10% of proteins in *S. cerevisiae* are predicted to have a secretory signal peptide

Secretory protein production leads to easy purification, and therefore is advantageous for the use of yeast as hosts, however low protein yield and plasmid instability have limited the commercial use of *S. cerevisiae* [32,172]. There are many strains of *S. cerevisiae* used both industrially in research. S288c is the strain in which much of the genetic information has been gained. It has a low sporulation rate and inability to grow on maltose[173] and lack of filamentous growth on nitrogen deficient media[174].

1.15 Protein Secretion

One of the critical factors which has been identified for optimal recombinant protein secretion in yeast is the efficiency of secretory signal peptide sequences to translocate proteins to the Endoplasmic reticulum (ER). α-mating factor is a common signal peptide used in yeast, and can be further modified to increase secretion56. Secretion efficiency with any given secretory signal peptide can vary between proteins, hence screening of these sequences for an individual protein may be necessary to optimise secretion fully[176], however other factors such as protein N-terminus effect, length of signal peptide, secondary structure and interactions must also be taken into account. Proteins which have been translocated to the ER then need to be folded properly. The ER has a limited folding capacity and the yeast ER is particularly ill equipped for efficient folding of proteins with large and complex tertiary structures which includes some of the higher value pharmaceuticals such as antibodies. This is because compared to their natural animal cell hosts, yeasts have lower numbers of chaperones and have limited extension capabilities of the ER membrane.

Yeast cells often lead low abundances of secreted proteins, although around 245 proteins have been identified in the yeast secretome[177]. It has been suggested that secretion is in part regulated by glycosylation[177]. Secretion is important for cell communications, interactions and immune response. Secreted proteins can also give information on the developmental and disease status of cells, and therefore secreted proteins have been suggested in humans to be vital biomarkers for diseases such as cancer[178,179]. Secretion can occur in classical or non classical secretion pathways. Non classical secretion is rare but has been reported for proteins which do not

contain a signal peptide. During classical secretion proteins with a signal sequence on the N terminal get synthesised in the ribosome and then sent to the ER for glycosylation, modification and folding. Modifications continue in the Golgi before being sent to the extracellular space by secretory vesicles[177,180].

N-glycosylation has a huge role in the folding and trafficking of protein, as well as often being a determinant in protein stability and protein interactions. N-glycosylation can be inhibited by tunicamycin enabling studies to assess the direct relationship between glycosylation and secretion[177]. Computational biology can help to identify proteins with signal peptides, however these sequences are not always fully defined and many other pathways exist through non classical secretion which do not involve these peptides.

Assessing protein secretion in mammalian cells is challenging as many of the secreted proteins are indistinguishable from those present in foetal bovine serum, present in mammalian cell media, and high presence of proteins from the media makes it difficult to identify the low abundance of secreted proteins coming from the cell. Identification of secreted proteins is also challenging as death of cells within the media during growth leads to protein release and can therefore mask presence of secreted proteins at low levels.

Analysis of protein glycosylation can be challenging due to the diversity of glycans [181,182]. This creates problems in enriching the proteins specifically and also identification by preferred methods such as mass spectrometry.

1.16 Oxidative folding and PDI related systems

It has been proven that engineering the secretory pathway of yeast such as *Saccharomyces cerevisiae* can improve protein production[183]. When target genes were evaluated for their effect on protein secretion, genes in the secretory and trafficking pathways were sometimes shown to increase amylase secretion compared to wild type strains, and this can also alter intracellular proportions of protein[183]. Particularly, changes to the endosome/golgi trafficking and histone deacetylase complex have been shown to increase protein secretion which can be further improved by multiple gene deletions[183]. Therefore understanding and manipulating the endoplasmic reticulum (ER) is important in engineering strains with high recombinant protein production ability. There are many factors within the ER that impact protein production. Efficiency of translocation to the ER lumen while avoiding saturation of the network is needed. High levels of protein traffic to the ER can lead to accumulation of misfolded proteins leading to cellular stress. Controlling the copy numbers and secretion signal sequences can be used to try and optimise this pathway by reduction of traffic to the ER. Low copy numbers and less efficient secretion sequences can reduce build up in the ER, however can also reduce overall protein production. Balancing these needs is vital in a good protein production system.

Maintenance of the ER to ensure proteins are properly folded is vital for secretion of properly folded proteins. Where the ER becomes stressed or mutated, it can become saturated leading to improperly folded proteins or aggregates which are not secreted. This can occur in *E. coli* as well as yeast cells and those of higher eukaryotes. Overexpression of chaperones or foldases may help to prevent

aggregation and promote secretion[184]. These chaperones and foldases can be used to help encourage prompt and proper protein folding to reduce ER burden.

Engineering of the yeast cells can reduce the effects of bottlenecks in the protein production pathways, for example in tailoring the endoplasmic reticulum[185] where BiP and PDI have been over expressed to reduce the unfolded protein response[186], overexpression of SEC16 to improve translocation from the endoplasmic reticulum[158] but also wider studies such as the use of microfluidics to identify many genes which increase yields of secreted proteins[187].

Disulphide bonds occur within the ER via Protein disulphide isomerases (PDI). . PDI1 is well conserved throughout species, for example in *Aspergillus* species, PDI assists in the folding and maturation of secretory proteins[188] such as by catalysis of the refolding of denatured or reduced RNase A, and in humans an extended PDI system exists to enable protein folding[189]. Protein disulphide isomerases cause formation, rearrangements and reduction of disulfide bonds within proteins. PDI proteins have thiorexin-like domains which can be active or inactive. The number and arrangement of these domains differ between different family members.

The function of PDI in forming disulphide bonds is regulated via feedback, hence when the endoplasmic reticulum has many misfolded proteins PDI levels increase. Pdi1 protein and mRNA have been raised by accumulation of unfolded proteins in the endoplasmic reticulum, but this is not believed to be as part of the primary stress response[188].

Overexpression of PDI and other chaperone proteins may increase activity of some protein production in a variety of species. In *Pichia pastoris* a gene encoding a

peroxidase was synthesized following integration into the genome. The recombinant protein produced was increased in activity with overexpression of PDI and endoplasmic reticulum oxidoreductase 1 (Ero1)[190].

Disulphide bonds stabilize and maintain structures within the protein, which is particularly important in proteins with more than one subunit. PDIs are common to many eukaryotic species. There is evidence that PDI is flexible in solution, which has been simulated and mapped[191]. This flexibility is believed to help PDI interact with the wide range of proteins it works with, in this sense yeast and mammalian PDI are believed to be very similar[192].There are 5 members of the yeast Pdi family, but only Pdi1 is needed for yeasts to survive. There is also a membrane bound PDI member called Eps1 which prevents protein misfolding and unfolded protein response induction[193].

After PDIs transfer disulphide bonds to proteins, they are reoxidised by Ero1, leading to the eventual formation of H_2O_2 and it is currently unknown how the resulting reactive oxygen species are detoxified[194]. H_2O_2 has been shown to be able to reoxidise PDI in vitro and also its presence can help to promote disulphide bond formation in mammalian cells[195], but it is currently unknown whether this remains true in yeasts.

Ero1 is necessary for maintaining a proper redox balance in the ER through reduction and oxidation of regulatory bonds which occurs via PDI[196]. In simple terms reduced PDI activates Ero1 and oxidised PDI inactivates it. PDI responds to the amount of free thiol groups and levels of reduced and oxidised glutathione in the ER to ensure the balance of redox and the generation of sufficient disulphide bonds for

protein folding[196], contrary to mammalian cells where Ero1 responds to the levels of oxidised/reduced glutathione[62].

Many studies indicate increased PDI expression can increase recombinant protein production and secretion[190,197–199]. It has been reported that PDI overexpression lead to a 3 fold increase of a recombinant protein in *S. cerevisiae*[200], even when the protein produced does not contain disulphide bonds[198]. This is likely to be due to the fact PDI can act either to form disulphide bonds, or as an isomerase[191,201,202]. To determine whether the PDI improves secretion due to acting as a chaperone or whether it forms a transient disulphide with a single cysteine present in a chosen protein, the cysteine was mutated to serine. The serine mutant still showed improved secretion when PDI was over expressed, indicating PDI may have a chaperone-like role in protein folding[200]. This further shows the increased protein production and secretion seen in PDI1 overexpression cells could be due to its role as an isomerase either as well as, or instead of its role in de novo disulphide bond formation, however these results could also occur as other proteins produced do require these bonds, and correctly folded proteins enable the endoplasmic reticulum to work at optimal speed. PDI1 and EMC1 have also been shown to be upregulated in some of the increased secretory strains[44]; both genes involved in efficient protein folding within the ER. When Reactive oxidase levels were measured to evaluate oxidative stress, the high production strains showed higher ROS and therefore indicating ER stress was present. Activation of the UPR and enhancing protein folding can reduce this, and strains which were thought to be more efficient in these areas had lower ROS levels[44]. Due to the mutations within the strains creating a more favourable environment for protein production and secretion, the overall ROS/unit of protein was reduced in the optimum mutant strains compared to the wild type[44]. In mutant strains

found to yield higher protein production levels, the chromosome containing PDI was also duplicated[183].

However, PDI levels are not always the magic solution to low protein yields or poorly folded proteins. Yeast is often considered to be inefficient at producing active human virus surface glycoproteins, potentially due to the folding issues in the endoplasmic reticulum. When genes expressing human endoplasmic reticulum chaperone proteins including protein disulphide isomerase were expressed with genes for these proteins, the PDI levels did not appear to improve the quality of the recombinant protein[203].

Overall, high producing strains have been associated with increased specific growth rate, increased glucose uptake rate, with reduced biomass on glucose but increased yield of ethanol on glucose[44]. The mutant strains shown to have a higher reporter protein titre in the medium also show lower intracellularly compared to a reference strain showing the secretory pathways were working at a higher efficiency, and therefore the strains had a higher secretory capacity[44]. The overall increases in secreted protein are thought to be associated with a global change in gene expression, rather than the large adjustment of few genes[44]. One of the strains and its descendants showed trisomy of chromosome three and many of the genes on this chromosome showed a roughly 2 fold increase in transcription, some of the genes present on this chromosome are those identified as being high secretory strains. Some of the genes also upregulated in higher secretory strains such as those required for anaerobic growth, through the Rox1p pathway and the UP1p pathways[44], those for nutrient signalling, nucleotide synthesis and phosphate metabolism were significantly upregulated. Downregulated genes such

as those involved in phosphate responsiveness were linked to phosphate utilisation and regulation, as well as those involved in activation of respiratory genes such as Hap1p36. Gene regulation was therefore leading to a hypoxia like state. Genes also downregulated include Bas1p[44] (which usually causes expression of nucleotide synthesis genes), and Mbp1p and Swi4p 36which regulate the expression of genes during the early cell cycle (specifically G1/s). These genes were analysed for causation in high producing strains by deletion or overexpression to confirm the data[44]. More generally genes concerned with mitochondrial function, energy, respiration, amino acid metabolism were downregulated and genes related to ribosomes, translation within the cytoplasm, organelle organisation, golgi vesicle transport, protein lipidation and lipid metabolic processes were upregulated[44].

The UPR (unfolded protein response) and the ER associated degradation pathway, controlled primarily by Hac1, can affect protein production. Typically a healthy ER under low stress would be expected to produce more protein. Increase in Hac1 leads to increased protein production[204] , due to increasing the ER capacity for protein production. Reporter proteins have been used to assess protein folding[205,206] to indicate how these pathways are performing, but it is important to note that different proteins could favour different conditions and hence reporter proteins may not always give the same results as each other.

As high levels of protein production can lead to oxidative stress in the ER, ways to reduce this stress have been studied, such as the use of antioxidants[207], which have been shown to reduce stress in the ER. It could occur that higher producing strains of cells have a method to reduce cellular stress and increase functionality of the ER, and so the study of ER and surrounding pathways in high producing strains can be

particularly interesting. When high producing strains were analysed it was found that increased expression of HAC1 and ERO1 could be seen[44]. HAC1 is a transcription factor used when the unfolded protein response pathway is activated.

As the UPR is used to reduce cellular stress in situations which cause heavy burden on the ER, it may enhance protein secretion when overexpressed by increase of KAR2 expression[204]. Kar2 works in the protein transport systems and so in theory increased Kar2 would enable more protein to be transported away from a stressed ER.

Increase in NADPH supply can also reduce oxidative stress via the pentose phosphate pathway [208], and although no obvious overproduction of NADPH was seen from a expression point, reduced biomass observed from higher producing cell lines may indicate a relocation of NADPH resources from biosynthesis to maintaining redox balances[44]. Protein disulphide isomerases cause formation, rearrangements and reduction of disulfide bonds within proteins.

The chemical environment within the cells is usually strongly reduced, compared to the atmosphere which is highly oxidised. Biological redox needs to be maintained to prevent oxidation by the environment. Oxygen use is essential for energy production in multicellular organisms as the aerobic respiration releases more energy compared to anaerobic systems. This allows the mitochondria to fuel cells by generation of ATP in form of the Krebs cycle and electron transport chain. The electron In this oxygen acts as an oxidising agent and hydrogen as a reducing agent which are able to generate ATP from ADP In the electron transport chain. Cells have to be protected from oxidation by use of the NADH and NADPH system which act as a storage of reducing molecules. These can then defend multicellular organisms from oxidation.

NADPH is used to reduce oxidised glutathione which acts as the largest redox defence system in humans. Glutathione Peroxidase (GPx) proteins GPx7 and GPX8 reside in the ER of mammalian cells and work with PDI and ERO1 to maintain the redox status of the cell. Redox needs to be maintained for cell function. ROS compounds include hydrogen peroxide which is reduced by GPx proteins to water. Oxidation of proteins in the ER contributes to the generation of ROS in the cells[209]

GPx7 and GPx8 are thought to support ER processes including oxidative protein folding, elimination of H2O2 and preventing lipotoxicity in mammalian cells[210]. GPx proteins deficiencies have also been linked to tumour formations, and so have been studied as an insight into human function, potentially because build-up of reactive oxygen species (ROS) can trigger cell death and affect life span in humans[211,212]. GPx proteins are part of a complex mammalian system which converts and contains ROS within cells, these ROS scavengers include catalase, superoxide dismutase and thioredoxin reductases[210].

The effect of GPx upon hydroperoxides with reduced glutathione was first described in 1957 by Gordon C Mills where it was described has having protective effects upon haemoglobin[213]. GPx7 and GPx8 have since been found to use thiols more efficiently than GSH, for example in the PDI [197,214] .

GPx proteins generally have 4 catalytic amino acids (seleno)/cysteine, glutamine, tryptophan and asparagine[215], however GPx8 has a serine in place of the glutamine. Both GPx7 and GPx8 reside in the ER. Human GPx7 contains 187 amino acid residues with a 19 amino acid long N-terminal signal peptide, whereas GPx8 is an ER anchored type two membrane proteins containing 209 amino acids[210]. In human cells the loss of GPx8 appears to lead to ER stress[216] and is it suggested that H2O2

derived from the redox reactions of Ero1 then leaks into the cytosol and leads to cell death.

The hydrogen peroxide produced via Ero1 has been shown to be used by GPx7 to enable oxidation of PDI in mammalian cells as well as in vitro[217]. Oxidation of GPx7 causes formation of disulphide bond between C57 and C86[210] . Treatment of GPx7 with H2O2 leads to it's oxidation and triggers binding of GPx7 to a heat shock protein via intermolecular disulfide bonding[218]. Loss of GPx7 reduces functionality of the heat shock protein known as GRP78 which in turn leads to accumulation of unfolded proteins driving increased oxidative stress.

1.17 Chaperone Proteins

One of the problems in using yeast or fungal host systems is glycosylation and other post translational modifications. These modifications differ between species and can lead to problems in using proteins between species. Chaperone proteins help to correctly fold proteins and begin the process of post translational modifications. Glycosylation is one of the most common forms of post translational modifications, but there are many different types of modifications that can be integral for protein function. In recombinant proteins, often formed in host cells different to the donor species, glycosylation is complex. In some proteins it is vital to have the correct glycosylation pattern, but systems used in the host species may be quite different to those used in humans.

Deletion of chaperone genes can massively reduce intracellular protein retention such as deletion of VPS5[183], which is involved in trafficking from endosome to late golgi.

The concentrations of calcium ions in the ER lumen is variable dependant on calcium use e.g. in signalling. Calcium ions can alter hydrophobic interactions and so may affect protein folding. Many endoplasmic reticulum molecular foldases and chaperones are low affinity and high capacity calcium ion binders and so changes to calcium present in the ER lumen can inhibit the function of chaperone proteins. Perturbation of ER-luminal Ca^{2+} inhibits chaperone function[219]. PDI has been shown to interact with calreticulin at low calcium ion concentration but dissociate as calcium levels get high[220].

Bioindustry has developed almost countless methods and cell lines to produce and purify recombinant proteins. Choosing the correct method to produce and purify these proteins depends on a multitude of factors including cost, complexity of protein, purity requirements, time scale, safety, genetic expression available, and technique availability. With time the increasing advances in microbes may increase their ability to produce perfect human proteins, however this is still a massive area of research. In an ideal world the creation and secretion of valuable pharmaceutical proteins could be conducted in yeast and massively reduce both upstream and downstream processing costs compared to gold standard current methods. Further research into yeast and human protein modification, yeast secretion and optimised conditions will be necessary to make this a reality.

This project aims to explore some of the factors which affect the production and secretion of recombinant proteins. This will be explored through different methods including the effects of individual genes as well as optimisation of the growth conditions. Any generally applicable strategies found to improve recombinant protein

secretion would then be utilised in the future to further improve yeast as a system for recombinant protein production.

Chapter 2 Materials and Methods

2.1 Materials, Strains and Vectors.

Strain and plasmid tables

Table 1: Yeast strains used throughout project.

Strain	Description	Reference
CGY424 = BY4741	MAT**a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	Brachman et al [221]
CGY839 = BY4741 *ste12::HIS3*	MAT**a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 ste12::HIS3*	Research Genetics Deletion collection wild type
CGY384	Matα *ura3-52 his3Δ200 Leu2-3,112 lys2-801 ade2-101*	Lappalainen *et al* , 1997
CGY384 *cof1::LEU2*	Matα *ura3-52 his3Δ200 Leu2-3,112 lys2-801 ade2-101 cof1::LEU2*	Lappalainen *et al* , 1997
YTH183 = NCYC3590 = SK1	MAT**a** *ura3Δ0*	SGRP
YTH185 = NCYC3591 = YJM978	MAT**a** *ura3Δ0*	SGRP

YTH207 = NCYC3629 = YPS128	MAT**a** *ura3Δ0*	SGRP

Table 2: plasmids used throughout project.

Plasmid name	**Plasmid description**	**Expected protein Size**	**Reference**	**Comment**
pCG495	pBevy-U-*gluc1*	~19kDa	Smales lab[222]	Producing *Gaussia* Luciferase
pTH815	pBevy-U-rluc1	~36kDa	Von Der Haar lab[223]	Producing Renilla Luciferase
pTH644	pBevy-U	N/A – empty vector	N/A	Blank Background
3406	GPx7	~22kDa	Von Der Haar lab[223]	GPx7
3407	GPx8	~24kDa	Von Der Haar lab[223]	GPx8
pCG162	Sec63-GFP	~105kDa	Received from Campbell Gourlay lab[224]	Produces GFP signal

Primers used

Table 3: primers used throughout project.

Description	Sequence
StaGluc f	AGAAACAAGCAAAACAAAAAGCTTTTCTTTTCACTAACGTATATGATGAGATTTCCTTCAATTTTTACTGCAGTTTTATTCGCAGC
Stagluc r	AATGGCTTTTGAAAAAAATAAAASAAGACAATAAGTTTTATAACCTTTAGTCACCACCGGCCCCCTTGATC
Gluc prs306 f gib	gggctgcaggaattcgatatcaagctTCATTATCAATACTGCCATTTCAAAGAA
Gluc prs306 rev gib	tcgaggtcgacggtatcgataagcttGCTTTTCATAGGGTAGGGGAATTTC
SUC2_up_f	GGCTCTATAGTAAACCATTTGGAAGAAAGATTTGACG
SUC2_down_r	GGCAAAAAGGTCCATCCTAGTAGTGTAAGGC
Gateway f	GGGGACAAGTTTGTACAAAAAAGCAGGCTGCATGAGATTTCCTTCAATTTTTACTGCAG

Gateway r	GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAGTGATGGTGATGGTGATGATC

Antibodies and dilutions

Table 4: antibodies used throughout project.

Antibody raised against	Species raised in	Dilution factor for working use in blots	Dilution factor for IF working use	Source
Gaussia Luciferase	Rabbit	1:1000	Na	New England Biolabs
PGK	Rabbit	1:5000	NA	York Biosciences
Rabbit	goat	1:10,000	NA	New England Biolabs
HA	Rabbit	1:10,000	NA	Sigma
Anti Rabbit TRITC	Goat	NA	1:100	Sigma
GPX7	Rabbit	1:5000	1:200	St Johns Labs

Anti-Gaussia luciferase antibody was purchased from New England Biolabs and raised in rabbit, the antibody was used at a 1:1000 concentration for primary detection.

Anti-PGK antibody was purchased from York Biosciences and raised in rabbit, these antibodies were used at 1:5000.

2.2 Medium

Yeast Media

Yeast Extract Peptone Dextrose (YPD)

YPD was prepared with 1% yeast extract (Difco) 2% peptone (Difco) and 2% glucose (added from 40% sterile stock). The resulting media was autoclaved to ensure sterility.

Selective medium (SC)

Selective medium was prepared with 0.67% Yeast Nitrogen Base without Amino Acids (Sigma), 2% glucose (added as 40% sterile liquid after sterilisation) and appropriate volume of amino acids as per selection used at manufacturers recommended concentrations (Sigma). The resulting media was autoclaved at 121°C at 15psi for 15 minutes to ensure sterility.

FOA medium

FOA medium was used to remove CRISPR plasmids by selecting cells which do not contain an active URA3 gene. To do this 5-FOA was added to SC-complete medium (2% Glucose, 0.67% YNB without amino acids, 0.2% SC complete amino acids 2%

agar). 5-FOA was then filter sterilised to a final concentration of 5.74mM and added once medium had cooled to ~60˚C and then poured into plates.

Escherichia Coli (*E. Coli)* Media

Luria broth (LB)

LB was prepared using 1% Tryptone (Difco), 1% yeast extract (Difco), 0.5% NaCl (Fisher), Appropriate antibiotics were added after sterilisation.

Agar

To make any media in agar/plate form, 2% agar was added before sterilisation.

2.3 Microbial Growth

Yeast growth

For general yeast growth an appropriate amount of liquid medium was added to a sterile tube. A small amount of cell culture from a plate was selected and added to the medium with a sterile pipette tip. This was left to grow shaking at 180rpm at 30°C.

E. coli growth

For general *E. Coli* growth an appropriate amount of liquid medium containing 100µg/mL ampicillin was added to a sterile tube. A single colony from a plate was selected and added to the medium with a sterile pipette tip. This was left to grow shaking at 180rpm at 37°C.

Growth analysis

Cells were analysed for growth by monitoring absorbance over 48 hours to indicate whether there were changes in growth profile for the wild type or ste12 knockout cells and whether this varied dependant on which plasmids were transformed into the cells. Growth readings were taken on a BMG LABTECH SPECTROstar nano plate reader with optical density measurements read at 595nm visible light path. Protocols were set for 3 flashes per well with an 1800 second cycle time shaking at 400rpm and additional shaking for 30 seconds before each cycle. Cells were grown with a target temperature of 30°C.

2.4 DNA Procedures

E. *coli* Transformations of plasmids

Plasmids which were not already present in *E. coli* were transformed into *E. coli* using 50µL competent cells on ice mixed with 1µL plasmid for 30 minutes before a 45 second heat shock at 42°C. 1mL of fresh LB was added to the mixture which was then grown for 45 minutes shaking at 37°C. After growth media was removed and cells were selected for on LB-ampicillin plates.

Competent E. *coli* cells

DH5α cell culture was used to inoculate fresh LB and grown shaking at 37°C until OD_{600} 0.5. 3.75mL of sterile glycerol (100%) was added to 28mL of this culture slowly before the cells were chilled on ice for 10 minutes. The cells were then pelleted at 4000rpm, supernatant was discarded and then cells were resuspended in cold 0.1M $MgCl_2$ 15% Glycerol. Cells were then pelleted at 3800rpm and resuspended in 6.25mL of ice cold T-salts (0.075M CaCl2, 0.006M MgCl2 and 15% glycerol) and

placed on ice for 20 minutes, occasionally mixing. The mixture was then spun down at 3600rpm and stored at -80°C until use.

Polymerise Chain Reaction (PCR)

PCR mastermix was made to roughly manufacturers specifications containing on average; 0.2mM of each dNTP 0.1–1.0µM of upstream and downstream primers, buffer containing 1.5mM MgCl2 and DNA polymerase as specified by manufacturer was added to 0.5µg of DNA to be amplified. PCR reaction tubes were then placed in a thermocycler for PCR reactions. Initially denaturation at 94–95°C for 5 minutes with further steps at 15-30 seconds per cycle. Annealing temperatures were optimised based on melting temperature of primers for 15-60 seconds. Extension was performed at optimum temperature for Taq polymerase as per manufacturers instructions (typically 72-74°C). Extension occurred for 1 minute per kilobase (kb) of DNA amplified. Generally 30 cycles of PCR were conducted before a 5-10 minute final extension.

Miniprep to purify plasmids

QIAprep Spin Miniprep Kit was used to prepare and purify plasmid DNA grown in *E. coli.* Plasmids were generated by overnight growth of *E. coli* in selective medium, and 5mL of culture was used to prepare plasmids.

DNA analysis

DNA was analysed via agarose gel electrophoresis. A gel was made by melting 1% agarose in TAE buffer (40 mM Tris (pH 7.6) 20 mM acetic acid 1 mM EDTA) in the microwave. The hot agarose was poured into a mould and 5uL of 10mg/mL ethidium bromide solution was added and mixed in. The gel was left to cool and set before the

comb was removed. The gel was transferred to a tank and the tank filled with TAE buffer. The gel was ran at 90V to separate the DNA by molecular weight.

Gel purification

Where necessary to purify PCR products the PCR products were separated via agarose gel electrophoresis. The Fragments were then viewed via UV light and cut out of the agarose gel. Gel fragments were weighed and purified via GeneJET Gel Extraction Kit as per manufacturers instructions

PCR purification

Where necessary to purify DNA which did not need to be separated as per molecular weight, i.e. there were no expected mixed weight DNA products, PCR purification kits were used. GeneJET PCR Purification Kit was used as per manufacturers instructions.

Transformation of plasmid DNA into yeast cells

Saccharomyces cerevisiae strains were transformed using standard lithium acetate methods[79], briefly 1mL of overnight culture was treated with a transformation mix of 240µL 50% PEG 4000, 36µL 1M lithium acetate, 72µL ddH_2O, 10µL SS DNA (boiled 10 min) and 2µL of plasmid directly from preparation. Cells were all selected through uracil markers and pure colonies were re-streaked on uracil negative plates before use in assays.

Creation of CRISPR plasmid for cutting at *suc2* locus[225]

Small guide RNA (sgRNA) were targeted at SUC2 position 1084 within the open reading frame with the sequence TTAGTAGCAAAACGAGACCAGGG using 2 oligomers SUC2_up_f and

SUC2_down_r. A plasmid casette[225] was digested with Swa1 restriction enzyme overnight at 25°C before heat inactivation at 65°C for 20 minutes. A second digest with BCL1 was then incubated for two hours at 50°C. The digested plasmid was purified using a commercial PCR purification kit.

Oligomers were hybridised by adding neat 10mM oligomers to ligation buffer, heating to 95° C and being left to cool to room temperature overnight. The hybridised oligomers were then ligated to the purified digested plasmid using T4 ligase incubated at room temperature overnight. Resultant plasmids were transformed into *E. coli*.

PCR products were created with 45 base pair overhangs to the Suc2 gene and used to amplify the Gaussia luciferase gene. PCR products were purified via a PCR purification kit. These PCR products were co-transformed with the CRISPR plasmid in the Lithium acetate method used previously but replacing any water with the purified PCR product.

Gateway cloning

PCR was used to amplify the Gaussia Luciferase gene with attB flanking regions using oligomers Gatewayf and Gatewayr. This was confirmed to be the correct size via agarose gel. PCR was purified by adding 150uL of TE buffer (10mM Tris, 1mM edta pH8) to 50uL of PCR. 100uL of 30% PEG8000/30mM MgCl2 was then added and mixed thoroughly before centrifugation at 10,000g for 15 minutes. The supernatant was removed and pellet dissolved in 50uL of TE buffer and checked on agarose gel for presence of attB dimers.

Entry clones were then developed using a BP reaction. Purified PCR was checked for concentration via nanodrop spectroscopy, and roughly 100 femtomoles off the PCR product and donor vector was added to the BP reaction. 4uL of 5X BP clonase reaction buffer was added to the mixture and the final volume made to 16uL with water. Reactions were incubated at 25°C for one hour. 2uL of proteinase K solution was then added to each reaction and incubated for 10 minutes at 37°C. the BP reaction was then frozen to be used in *E. coli* transformation

After the *E. coli* had grown on plate an overnight culture was grown in selective media and then the plasmid was purified via mini prep. The resulting plasmid was then subject to the LR reaction. 100ng of this plasmid was added to 2uL of the LR plasmid with 4uL of 5X LR clonase reaction buffer and Te buffer added to a final volume of 16uL. Reactions were incubated at 25°C for one hour and then 2uL of proteinase K solution added before incubation at 37°C for 10 minutes. This reaction was then transformed into *E. coli* and the remaining reaction stored in the freezer. The resulting *E. coli* was grown overnight and plasmid was purified via mini prep. This plasmid was digested using Not1 enzyme as per manufacturer's instructions and then transformed into yeast using the standard yeast transformation protocol.

Gibson Assembly

PCR was conducted to give PCR products with overlaps to allow the Gibson assembly to proceed using oligomers Gluc prs306 f gib and Gluc prs306 r gib. PCR products were purified via agarose gel and then gel purification kit. The Gibson assembly plasmid was linearised by digest with HINDIII as per manufacturer's instructions. This was checked on an agarose gel and purified as per manufacturer's instructions via PCR purification kit.

The Gibson assembly reaction involved adding 2µL of each PCR product and linearised destination plasmid to 10µL of Gibson assembly master mix made to a final volume of 20 µL in water. Samples were incubated at 50°C for one hour and then stored at -20°C. The reaction was then transformed into E.coli and miniprepped before usage.

2.5 Protein preparation and analysis

Renilla Luciferase assays

Cells were analysed for luciferase production, cells containing the pTH815 and 645 plasmids should produce luciferase, the pTH815 contains a secretory signal sequence and therefore luciferase should be excreted. The luciferase production was analysed in overnight culture, supernatant and concentrated resuspended cells. This was normalised for absorbance at the time of luciferase measurement where assays were carried out during early growth. Cells containing the PET plasmid were used as a control to indicate readings due to other factors and the luciferase production was measured in both the CGY424 WT and the CGY839 ΔSte12 cell lines.

Gaussia Luciferase assays

Cells were analysed for luciferase production, cells containing a Gaussia plasmid should produce Gaussia luciferase (Gluc) measurable via commercial assay (Thermofisher cat 16160). The luciferase production was analysed in supernatant and resuspended cells. Liquid cultures of each cell type were prepared. Some samples were matched based on optical density of 10, other cells were matched based on time.

Samples were prepared by spinning 1mL sample of liquid culture at 8000rpm for 1 minute and collecting the supernatant. Cell samples were resuspended in 1mL of water. 30µL of the cell or supernatant sample was plated in a black welled 96 well plate. 10µL of 1M Tris (pH8) was added to each well before the addition of 30µL of Gaussia luciferase assay reagent (prepared as per manufacturer instructions). The plate was incubated for 10 minutes benchtop before being read on a BGM Optma Fluostar multiplate reader at 475nm. Settings used a lens emission filter, 3600 gain, top optic reading, a 0.2 positioning delay, 1 cycle in 1 kinetic window with a 2 second shaking cycle at 500rpm before measurement.

Magnetic-bead Immunoprecipitation

Cells were grown overnight and diluted the following morning 1 in 10 and grown for a further 4h to reach log phase. 1mL of cells was taken and centrifuged at 8000rpm for 3 minutes. The resultant pellet was resuspended 125µL yeast protein extraction reagent (Y-PER) as per manufacturer's instructions. Reaction was incubated at room temperature for 20 minutes, before centrifugation for 10 minutes at 13,000 RPM. 100µL of the resulting lysate was added to 25µL anti-HA magnetic beads. Incubate for 30 minutes at room temperature shaking. Separation of the beads from the surrounding liquid was done magnetically and washing occurred by addition of 300µL TBS/T, 1 minute shaking was conducted before magnetically separating and discarding wash solution. Washes were repeated twice. Elution of the bound protein was achieved in 100µL of 50mM NaOH shaking at room temperature. The eluate was added to 50µL of 1M tris to neutralise it before adding SDS loading buffer and analysing 25µL of the resulting sample via western blot.

Cell fixing

A 2% paraformaldehyde solution was made by dissolving paraformaldehyde in PBS for 2 hours at 80°C. The formaldehyde solution was filtered through a 0.22 micron filter. Cells were resuspended in the paraformaldehyde solution at 1OD unit per mL of paraformaldehyde solution. 10mL of this culture was incubated shaking for 7 minutes with the paraformaldehyde solution before being pelleted at 2000rpm for 3 minutes. Paraformaldehyde was discarded and cells resuspended in 0.5mL of ice cold 1.25M glycine in PBS. Cells were span down as before and washed in 1,25M glycine/PBS a further 2 times. Cells were then resuspended in 125µL of Y-PER reagent and incubated for 30 minutes. The magnetic bead protocol was then followed as before.

SDS analysis of proteins.

TCA precipitation of supernatants

SDS page gels were run with both pure supernatant and supernatant concentrated by precipitation using TCA[226].

Cell lysate preparation

1mL samples of overnight cultures of cells were used to make a lysate. Cultures were OD matched to OD5 and centrifuged at 8000rpm for 1 minute. Supernatant was removed and cells were resuspended in 200µL of lysis buffer (0.1 M NaOH, 0.05 M EDTA, 2% SDS, 2% β-mercaptoethanol). The resulting mixture was heated for 10 minutes at 95°C before being removed from the heat. 5µL of 4M acetic acid was added to neutralise. Following this 50µL of loading buffer was added (0.25 M Tris-HCl pH 6.8, 50% Glycerol, 0.05% Bromophenolblue).

Preparation of polyacrylamide gels.

Gels contained a 5% acrylamide stacking gel followed by a resolving gel with varied acrylamide concentration depending on the experiment. Pre-cast gels of 4-12% were also used. The resolving gel was poured first into a BioRad cast system and covered with isopropanol whilst the polymerisation reaction occurred. The isopropanol was removed after polymerisation was complete and then the stacking gel was poured on top and a comb placed to form the sample wells. A typical gel contained 6.5mL of resolving gel and 1.5mL of stacking gel. Gels were prepared as per the table:

Table 5: composition of SDS gels made.

Gel Type	Volume 1.5M Tris/HCL pH 8.8 0.4% SDS	Volume 1M Tris/HCL pH 6.8 0.4% SDS	Volume 40% Acrylamide	Volume TEMED	Volume APS	Volume Water
12.5% resolving	2mL	0mL	2.5mL	10µL	35µL	3.5mL
12.5% stacking	0mL	0.75mL	0.5mL	7µL	35µL	1.75mL
10% resolving	2mL	0mL	2mL	10µL	35µL	4mL
10% stacking	0mL	0.75mL	0.375mL	7µL	35µL	1.875mL

SDS gel electrophoresis protocol

Following gel polymerisation, gels were placed in an electrophoresis tank and covered in running buffer (Tris/Glycine/SDS). 5µL of protein ladder was added and sample volumes between 10µL and 30µL added to individual wells. Gels were initially ran at 90-100V and then voltage was increased to up to 180V until the dye front had reached near the end of the gel. Gels were then either stained using a water based coomassie staining method[227], or transferred onto nitrocellulose membrane for western blotting.

Western Blot

Transfer

Polyacrylamide gels, once ran to completion, were transferred onto nitrocellulose membranes via a Bio-Rad semi-dry protein transfer machine. The membrane and blotting paper was soaked in transfer buffer (0.0029% Glycine, 0.00004% SDS, 0.0058% Tris base, 20% methanol) for at least 10 minutes before transfer. The gel was stacked using 2 pieces of soaked blotting paper, the membrane, the gel followed by 2 further pieces of blotting paper. The stack was carefully pressed between layers to remove bubbles which may be detrimental to the transfer. The top electrode was placed on top of the layers and run using 25V for 30 minutes.

Immunoblotting

After transfer the membrane was blocked in 5% skimmed milk powder made in PBS containing 0.1% Tween. The membrane was left shaking at room temperature for 1 hour in blocking solution (5% skimmed milk dissolved in PBS/Tween) . After blocking the membrane was probed using an appropriate dilution of antibody made up in the same blocking solution at 4°C overnight. The following day the membrane was washed at least 3 times in PBS-0.2% tween. The membrane was then further probed

with a secondary antibody which had an HRP conjugation. This was made up at an appropriate dilution in blocking solution and left turning for 1 hour. The membrane was washed again in PBS tween, changing the wash buffer a minimum of 3 times before it was ready for developing.

Western blot development

Western blots were developed using an ECL detection method. The membrane was left in a 1:1 Ratio of 2 developing solutions. Solution 1 (1mL luminol 250mM (3-aminophthalhydrazide for FLUKA No.09253), 0.44mL *p*-coumaric acid 90mM (Sigma), 10mL 1M Tris-HCL (pH8.5) made up to 100mL with H_2O) and solution 2 (64µL 30% hydrogen peroxide, 10mL 1M Tris-HCL (pH8.5) made up to 100mL with H2O.) were stored in the fridge in the dark. The mixed solutions were left on the membrane for 30 seconds before the membrane was removed and placed in the gel dock and detected automatically using the parameters decided by the software.

Blot stripping

Blots were incubated overnight in PBS. The following day they were rinsed twice with fresh PBS. Blots were then washed in PBS for 15 minutes once and then twice for 5 minutes each. PBS was drained and 10mL neat stripping buffer was added and incubated for a minimum of 30 minutes at room temperature. After draining stripping buffer blots were rinsed twice in PBS before the 3 PBS washes were repeated. The blots were then blocked for 1 hour in 5% milk before proceeding to immunoblotting.

2.6 Cell analysis

Flow cytometry

Cell cultures were diluted to OD 0.1 and grown overnight with or without the presence of 1.5mM H_2O_2. Cells for analysis were grown with 1µM H_2DCFDA which can be used to detect ROS and oxidation of proteins[228], controls were grown without additional staining present. Cultures were grown in plates for 24 hours shaking at 30°C. After growth 25 µL of culture was added to 500 µL of fresh sterile PBS. 2 µL of PI stain was added to each sample before analysis, again excepting some controls. Negative and positive controls were used to effectively gate the flow cytometer readings and then samples were ran and analysed according to these gates. 1000 events were collected per sample according to the live cell gating.

Immunofluorescence

Each overnight culture was diluted 1 in 10 and left to grow for 4 hours to reach log phase. A 2% formaldehyde solution was then made with 5mL of log phase cells and left to stand at room temperature for one hour. Cells were centrifuged for 4 minutes at 4000RPM and resuspended in 2mL of Sorbitol buffer (1.2M sorbitol, 0.1M potassium phosphate, pH7.5). This step was repeated twice to wash the remaining formaldehyde from the culture. Washed cells were resuspended in 0.5mL sorbitol buffer. 1 µl of beta-mercaptoethanol and 20 µl of 1mg/mL zymolase were added before cells were incubated for 30 minutes at 37°C.

Slides prepared in advance by incubation of a drop of poly-L-lysine on each well of a multi well slide for 1 hour, rinsed with water and left to completely dry. 15 µl of cell suspension was then added to each prepared well. Cells were incubated on the slide for 10 minutes at room temperature before being aspirated off.

Each well had 10 µl of 0.1% SDS placed on it and removed after 30 seconds. Wells were then washed with PBS/BSA by adding a drop to each well and then aspirating off with a pipette tip. These washes were repeated 10 times. Slides were kept in a humid environment to prevent them drying out completely.

Primary antibody diluted in PBS/BSA was added at 15 µl per well and incubated overnight in the fridge. The following day 10 further washes in PBS/BSA were conducted. Secondary antibody also diluted in PBS/BSA was then added at 15 µl per well and incubated for one hour at room temperature in the dark. 10 further washes were conducted before a drop of phenylenediamine was added to each well and slides covered with a cover slip and sealed. Slides were then analysed by microscopy or stored at -20 until ready to be viewed.

2.7 Computational

RNA extraction

RNA extraction was performed using E.Z.N.A Yeast RNA Kit (R6870-01) from OMEGA as per manufacturer's instructions. This work was completed prior to the beginning of this project.

RNA analysis computational

RNA analysis was conducted using galaxy software[229], and tools within the galaxy software[230] quality checks were used to analyse the raw RNA reads. Quality checks were conducted using FastQc. FastQC can analyse how the data quality changes, reads can drop in quality over time due to the reaction kinetics. Use of the quality control allows this to be visualised and reads can be trimmed to reduce the presence of poor reads. TrimGalore! Was used to trim the data to keep only high quality reads

based on the FastQC readout. RNA reads were then aligned to the genome. The Galaxy software contains a model *saccharomyces* genome which was used for the work. HISAT2[231] was used to begin the alignment process of matching RNA reads to the genome. The reads were then counted via FeatureCounts which analyses how many of the RNA reads were present in each part of the genome. DEseq2 was used to compare reads between WT and *ste12* mutant cells. After raw data reads were concatenated, quality checked, trimmed and aligned to a reference genome, significant differences between the wild type and ste12 mutant cells were viewed. Gene ontology mapping through the SGD was used to show pathways in which genes were up or down regulated[232]. Gene set enrichment analysis[233,234] was used to indicate which gene pathways had been most affected by the mutation and cytoscape[235] was used to visualise this.

Design of experiments

Design of experiments was done using minitab software choosing a full factorial design and entering the experimental limits into the software enabled generation of lists of experiments to conduct. Data gathered experimentally was then input back into the software to enable the generation of graphs and analysis of factor importance as well as factor interaction.

Statistical methods

Statistical analysis was conducted using Graphpad Prism 9 software. Results were analysed manually and then entered into software and relevant information was entered to enable statistical tests with default parameters. As multiple samples were compared in each experiment, ANOVA tests were used to compare samples, with

one way ANOVA used to compare between samples where only one independent variable is present, and two way ANOVA used with two independent variables, for example when comparing cell and secreted protein levels between cell lines in one experiment. A Tukeys post-hoc test was used to compare all samples to eachother for statistical significance, and then relevant differences were indicated. P values were used primarily to indicate and denote statistical significance. Statistical information was checked manually before inclusion into figures and data presented in this thesis.

Chapter 3 : Investigating the effect of a *Ste12* mutation on recombinant protein production

3.1 **Introduction** to *Ste12*

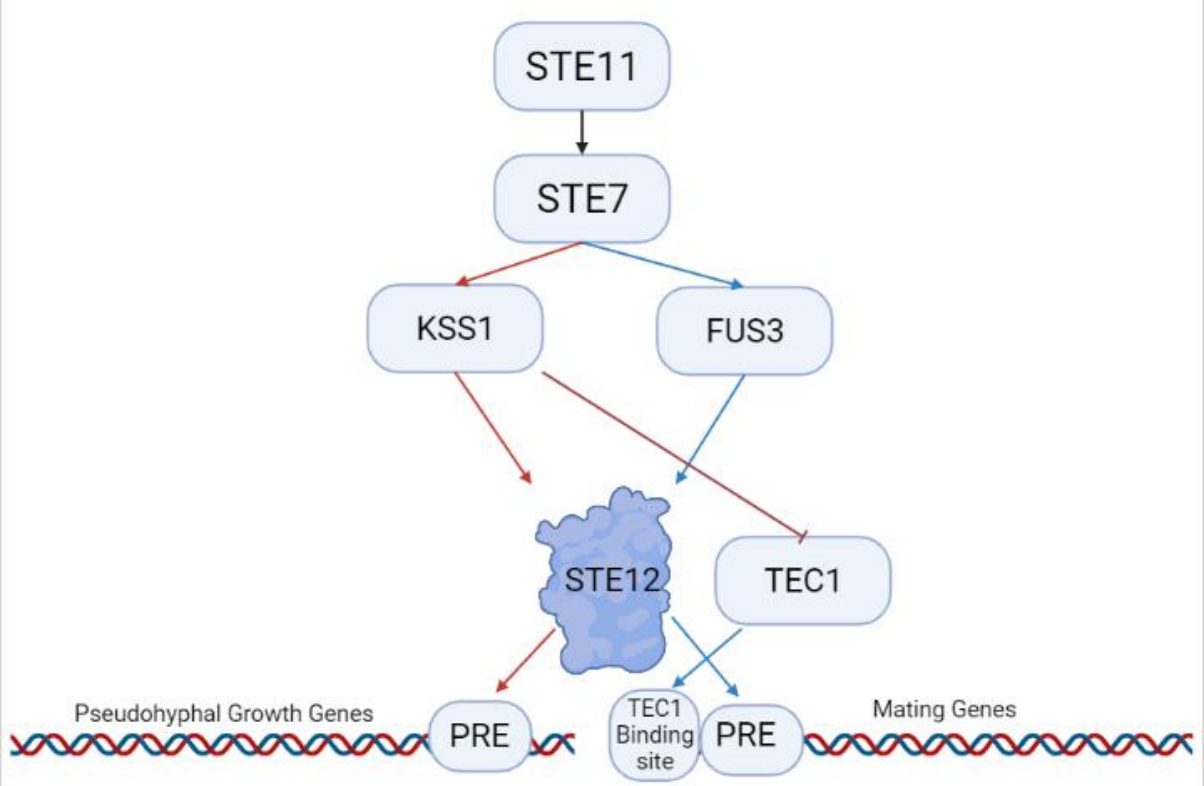

Figure 3: Major pathways in which STE12 is regulated through phosphorylation pathways leading to hyphal growth (red) or mating (blue) gene activation.

The gene which encodes the Ste12 protein (Ste12) is so named due to the origins of identification from a yeast sterile mutant, and was originally identified as a target of the Fus3 Mitogen Activate Protein Kinase (MAPK) pathway to regulate mating in 1989[236], but has since been implicated in pseudohyphal development acting alongside Tec1p[237] (see figure 3). Ste12 has been shown to act as a downstream transcription factor for the execution of mating, haploid division and pseudohyphal delevopment[238].

Ste12 acts through the mating or pheromone response pathway which is activated by the Ste2 pheromone receptor, a trimeric G protein which activates the MAPK

cascade in response to binding to yeast mating pheromones. In *S. cerevisiae* the MAPK kinase pathway is used to enact an effect through Ste12 in the pheromone response pathway. Ste12 enacts its effects via DNA binding for transcriptional regulation. When Ste12 is phosphorylated by FUS3p this regulated pseudohyphal growth however when it is activated by KSS1p this regulates mating, this phosphorylation depends on whether it is acting due to a pheromone response[236] or pseudohyphal growth[239]. Yeast Ste12 proteins contain an N terminal Ste motif used in DNA binding to the pheromone response element, it has been shown that mutation of this site leads to a great reduction in binding affinity[240]. Residues 301 to 335 of Ste12, defined as the minimal pheromone induction domain[238], which depends on the MAPK pathway for induction activity. The serine and threonine residues contained in this domain can be mutated without affecting induction of transcription[238].

Regulation of STE12 in yeast.

Ste12 is inhibited by Dig1 and Dig2 which dissociate upon Ste12 activation[241]. Ste12 can be inhibited by Dig1/Rst1 and Dig2/Rst2 which act individually on separate regions on STE12[242]. Mutation of tyrosine's in the 301-335 region was shown to prevent interaction of Dig1 and Dig2 inhibitors[238]. Kss1 binds to Ste12 when unphosphorylated which leads to Kss1-mediated repression of Ste12. When Kss1 is phosphorylated by Ste7, the interaction between Kss1 and Ste12 is weakened, stopping the oppression. Fus3 binds Ste12 less strongly and therefore is a weaker inhibitor of growth[243]

For mating, Ste12 works as a homodimer, whereas in the invasive pathway leading to pseudohyphal development, Ste12 acts together with Tec1p[244] to activate genes.

Signal specificity is gained by Fus3 phosphorylation leading to degradation of Tec1 which is used in filamentation response genes (figure 3). Tec1 is phosphorylated at threonine 273[245] which leads to destruction by Gia2 and Cdc53 ubiquitin ligases[245] causing mating responses over pathogenesis and invasion responses.

When activated by pheromones, phosphorylated Ste12 homodimers bind a pair of pheromone response elements. The pheromone response element has 4 copies upstream of the *FUS1* gene and leads to regulation of Fus1 via Ste proteins. Deletion of the pheromone response regions prevents all transcription of FUS1.Residues 216 to 688 of Ste12 have been expressed to cause FUS1 induction, however it is believed this occurs due to the residues causing activation of endogenous Ste12[242]. Overexpression of residues 262 to 594 has been shown to cause FUS1 induction and can bind Dig1 inhibitor but not Dig2[242]. When hybrids of Ste12 were generated with DNA binding domains and activation domains of Gal4p were created it became possible to define a pheromone induction domain of Ste12 to enable gene transcription[238]. The pheromone induction domain of Ste12 has been suggested to both relieve repression and activate transcription together with another activation domain[238].

STE12-like proteins

In filamentous fungi, Ste12-like proteins have been implicated in sexual development and pathogenicity, these differ from the Ste12 protein found in yeasts as they contain 2 zinc fingers at the C terminal which are absent in the yeast protein. It is suggested that these zinc fingers have a role in virulence but are not necessary for DNA binding[246,247], however this role is unclear. The deletion of the C2H2 domain at the C-terminal does not seem necessary for DNA binding in vitro studies[246], but is

needed for in vivo protein function, as seen by mutagenesis studies showing that the zinc fingers are needed for plant infection[247]. Fungal transcription factors “Ste12-like factors” have been implicated in processes leading to morphogenetic programs[244].It is currently not known whether Ste12 like proteins bind their cis acting regions singly or as dimers[244]. In *Candida albicans*, it is necessary to switch between yeast and filamentous forms to enable pathogenicity in humans which is controlled, in part, by the *Candida albicans* Ste12 homolog. Mutations in proteins surrounding these systems can prevent filamentation and lead to avirulent fungi, and deletion of *STE12* has been shown to prevent the formation of pseudohyphae[239]. In other species of fungi, particularly with Ste12-like proteins, these effects differ[244].

3.2 Results

The main aim of work in this chapter was to follow up on earlier observations in the lab the deletion of the *STE12* gene in *Saccharomyces cerevisiae* may improve yields of recombinant proteins.

3.2.1 Transformation of plasmid DNA into yeast cells

To analyse the effects of *ste12* deletion in *S. cerevisiae*, it was decided to transform cells with a series of plasmids expressing recombinant proteins. These included *Gaussia* and *Renilla* luciferases, as well as PETase[248] (a PET degrading enzyme of interest in other projects in the lab). *Renilla* luciferase and PETase were codon optimised[249] for expression in *S. cerevisiae*, whereas *Gaussia* luciferase was not optimised. These genes provided a variety of substrates for studying the levels of recombinant protein expression by either western blot to detect the protein levels by immunoblotting, or a commercial luciferase assays to show levels of functional

protein produced. These assays will allow for the efficient detection of effects of *STE12* gene deletion on the cells ability to produce recombinant luciferase protein.

Transformation of plasmid DNA into yeast cells was deemed successful where colonies were present on –Uracil plates and not on control plates (that is cells which followed an identical procedure with water added in place of plasmid).

BY4741 Background cells

Growth analysis

3.2.2 Growth of BY4741 cells in YPD

It was believed, from previous laboratory data, that the *ste12* deletion mutant may have an impact on growth rate comparative to *STE12* wild type strains due to the impact it has on mating and pseudohyphal growth. Previous data generated within the lab (not shown) had indicated that *ste12* deletion would lead to faster growth compared to their wild type counterpart in rich medium (YPD). Originally, we wished to reproduce this data by comparing BY4741 cells with and without a *ste12* deletion. Effects of a *ste12* deletion on yeast growth were initially tested by measuring growth rates of wild type and knockout cells, grown in rich medium (figure 4). Rich medium was chosen as the original observations had shown the strongest effects. Cell growth was analysed by optical density at 600nm wavelength over time to enable calculation of the growth rates. These data were collected for 45 hours to ensure we recorded all of the growth phases to ensure we would be able to see any differences in growth rate between the wild type and *ste12* deletion strains.

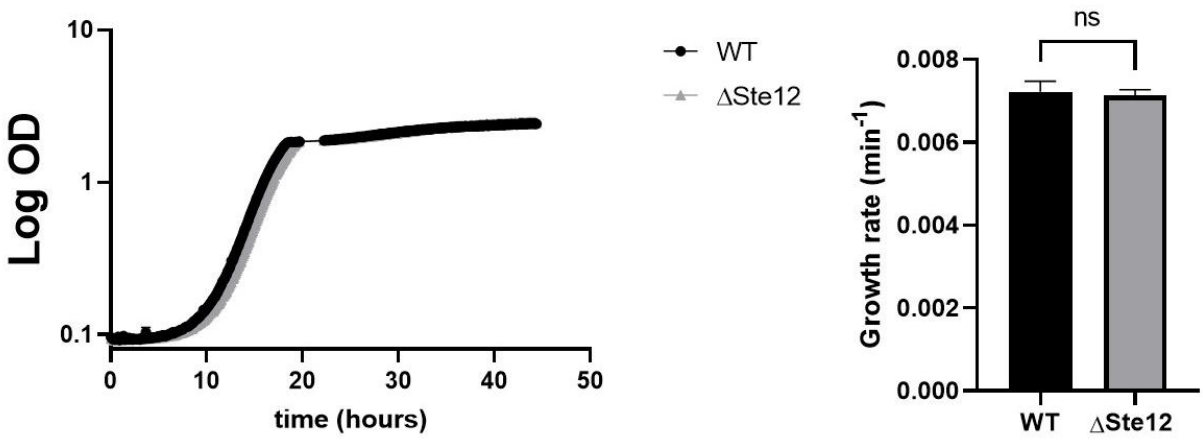

Figure 4: Growth rate of BY4741 WT and ∆ste12 was analysed by optical density readings (600 nm) over 45 hours and growth rates calculated over consecutive seven timepoint windows with the largest value observed reported as growth rate for the culture. Paired T-test showed no significant difference between growth rates.

Growth curves and rates derived from the logarithmic part of the growth curves showed no significant differences (p= 0.3130) in growth rate in the *ste12* mutant compared to the wild type in the same strain. This was unexpected as previous data had suggested that there were growth differences between the strains, however we were unable to confirm this in our study.

3.2.3 Growth analysis of transformed cells in minimal media.

As no difference was noted with the *STE12* wild type cells compared to the *ste12* mutant, it was thought that these effects could be more apparent either in minimal media or upon expression of recombinant proteins. For this reason, growth rate assays were conducted again with both wild type and *STE12* deletion strains after transformation with a set of plasmids, either empty or containing representative recombinant proteins such as *Renilla* luciferase or PETase. Strains were then grown

in SC-ura medium and analysed for growth rates as before. It was thought that the increased demand on the cells due to the production of the recombinant protein might exacerbate differences in growth as extra cell resources were used to produce proteins. As the original experiments conducted before this study had indicated that *ste12* deletion would increase recombinant protein production, it was thought the effects of protein production would have a stronger effect within the *ste12* deletion strains.

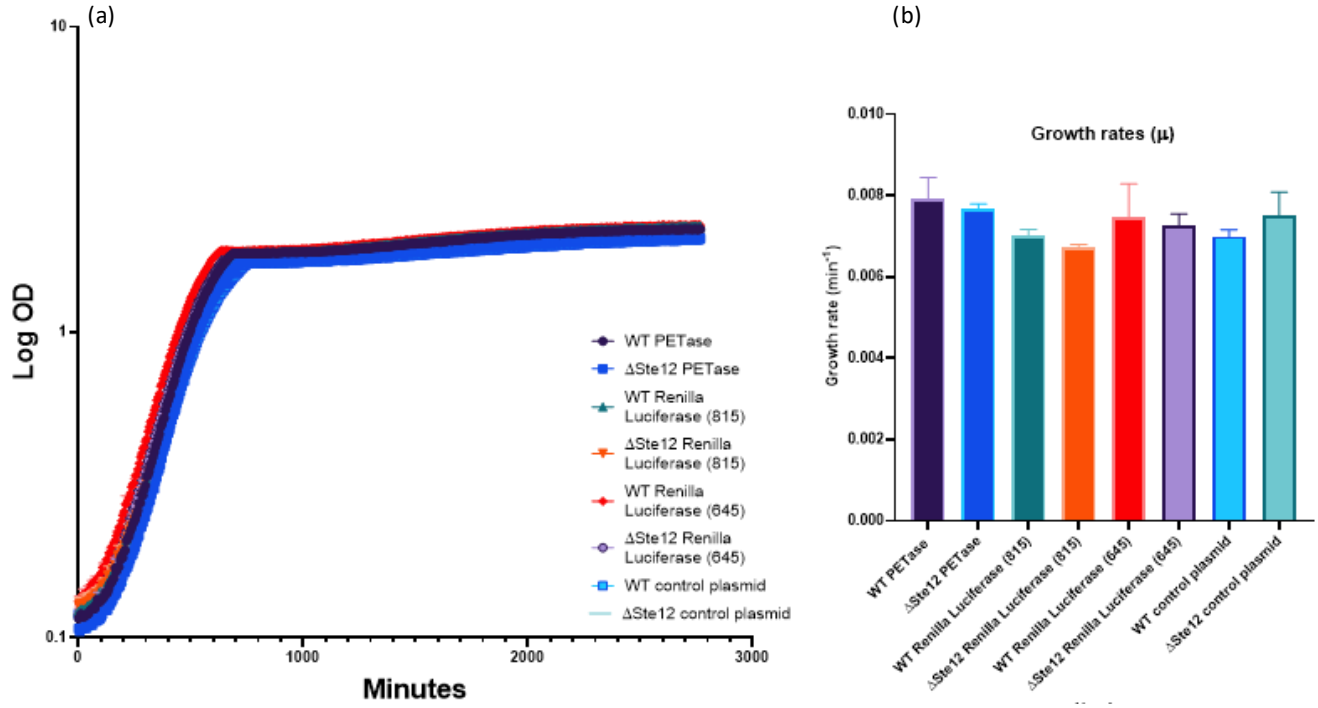

Figure 5: (a) Growth curves for BY4741 WT and Δste12 cells transformed with control plasmids or various recombinant protein expressing plasmids showing error bars for SEM (N=3). Growth rate analysis for each cell type and plasmid were compared. (b) Growth rates were determined as the highest growth rate observed over a seven timepoint series. The time in which the cells ended logarithmic growth was also noted. One-way ANOVA with Tukey's post-hoc test showed no significant differences in growth between samples.

Analysis of growth rates showed that there were no major differences in growth of either the wild type or *Ste12* mutant when expressing recombinant protein compared to a control plasmid which only conferred ability to produce uracil (P= 0.9868).

This was again unexpected given the original observations, but confirms the findings made in YPD media that deletion of *ste12* does not significantly affect growth rates in BY4741.

3.2.4 Renilla Luciferase assays

To analyse whether the *Ste12* mutant showed differing levels of protein production, a luciferase reporter gene had been transformed as an example recombinant protein. Renilla luciferase is a single domain protein containing 311 amino acids to form a 36kDa protein[250]. Following transformation, cells were grown in SC-ura medium to select for the marker, otherwise the cells risk growing without copying the plasmid. The transformed plasmid contained the uracil selection marker and the luciferase gene preceded by a secretory signal sequence and hence should be secreted into the extracellular medium. Samples of cellular culture were taken at different time points to indicate when luciferase production was occurring. Samples were separated by centrifugation to analyse whether protein remained inside the cells or whether it was being successfully secreted. Wild type and *Ste12* deletion cells were compared.

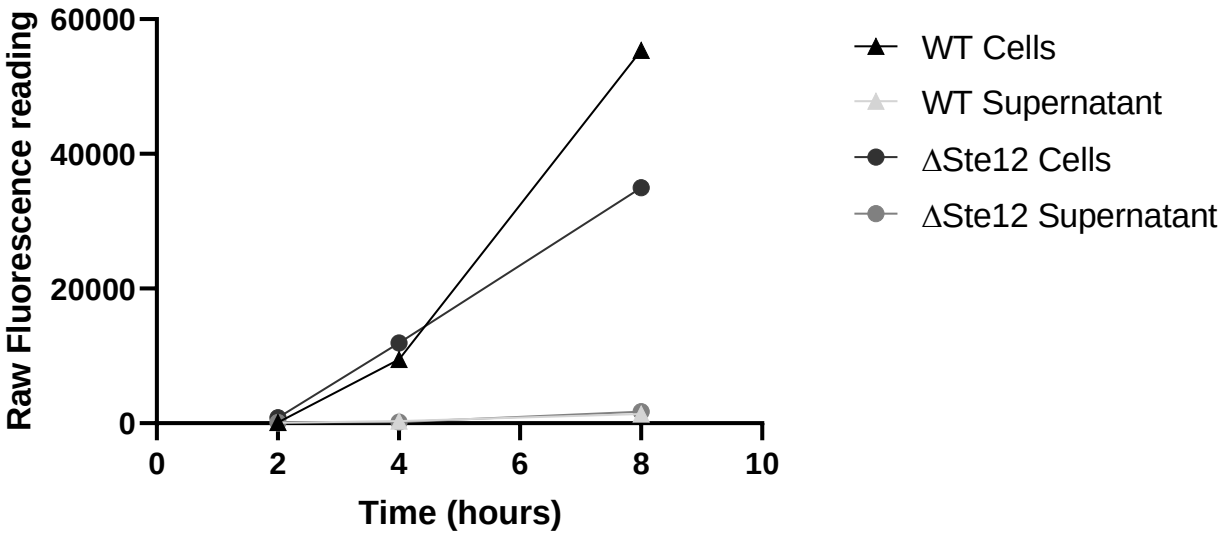

Figure 6:Renilla Luciferase assay showing production of luciferase over time. Assays conducted at 3 time points measured from initial dilution of overnight cultures. Points shown as averages of 3 independent measurements. 2-way ANOVA with Tukey's post-hoc test showed no significance.

Luciferase production was analysed in growing cells and in culture supernatant, following inoculation to a low starting OD (0.1). Results were corrected for changes in cell number during growth by taking an equal number of cells according to optical density at each time point. Over time, as cells began to grow, cellular luciferase increased, however secreted luciferase remained low. The WT cells showed slightly elevated levels of luciferase than the ste2 deletion strain at the assay end point, although this was not significant. These results indicate that the deletion of ste12 is not increasing the ability of the cells to make recombinant Renilla luciferase.

3.2.5 Gaussia Luciferase assays

Gaussia luciferase, compared to the more commonly used *Renilla* Luciferase, is a naturally secreted protein by its host (the copepod *Gaussia princeps*[251]) that appears more stable due to high levels of disulfide bonds. It is also shown that *Gaussia*

luciferase can show higher signal and be secreted more effectively compared to *Renilla* or Firefly varieties[252]. As industrial applications favour secretion to avoid having to employ cell lysis in downstream processing, we decided to focus on secreted products also in this project.

Following the initial observation of low *Renilla* luciferase secretion described in the previous section, we decided to follow secretion of recombinant proteins using *Gaussia* luciferase instead of *Renilla* luciferase. As before we compared WT and *ste12* deletion strains for expression of *Gaussia* luciferase after an overnight growth period and compared cellular levels to secreted levels to test where protein was found. As with the *Renilla* luciferase assay, low signal for *Gaussia* luciferase was seen for cells grown overnight. Moreover, repeat measurements showed that the signal decreased following storage. Following these initial observations we therefore investigated whether activity could be stabilised in the media supernatant. In particular, we investigated whether stabilising the pH could also stabilise luciferase activity, as this protein has a pH optimum around 7.7[253], whereas unbuffered yeast medium becomes acidic with culture growth. Buffering of the culture supernatant with 30% Tris at pH 8 improved the luciferase signal in cells transformed with Gaussia expression plasmids but was not shown to give any increase in readings when added to cells which do not contain a luciferase sequence. Results with the added tris appeared less variable between assay repeats and so future assays all included addition of Tris pH 8 to 100 mM final concentration.

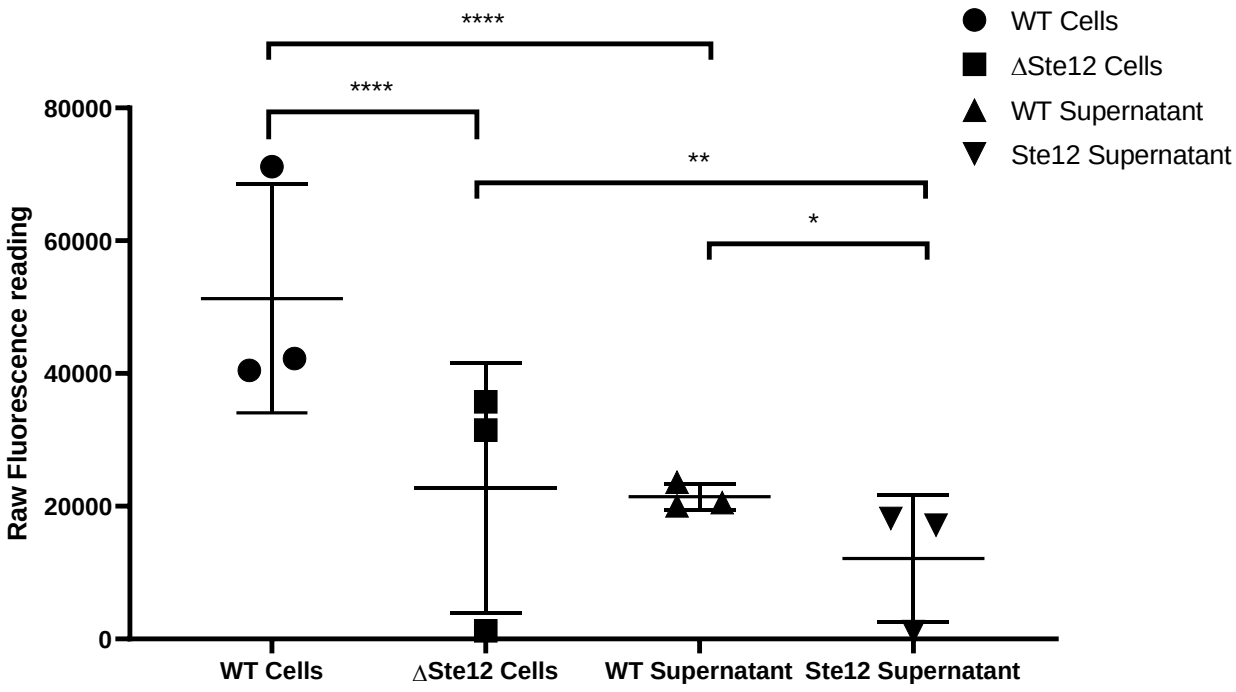

Figure 7: Raw data for Gaussia luciferase assay comparing WT cells (BY4741 background transformed with a secretory Gaussian luciferase plasmid pCG495) and ΔSTE12 (BY4741 cells with a STE12 gene knockout also transformed with the pcg495 secretory Gaussian luciferase plasmid). Samples had 1M Tris pH8 added to correct pH. Data shows mean for each of 3 biological triplicates (n=3). 2-way ANOVA and Tukey's post-hoc test were used to determine significance of differences.

The *Gaussia* luciferase assay in figure 7 is a representative example of *Gaussia* luciferase data gained from the yeast strains and shows how wild type cells are producing and secreting significantly more *Gaussia* luciferase than the *ste12* mutant cells, indicating that if Ste12 has any significant effect on recombinant protein production it supports this process, contrary to initial hypotheses.

3.2.6 Gaussia Luciferase Western blot

Because the *Gaussia* luciferase assay was noted to be affected by conditions such as the pH of the medium, western blots were used to back up and confirm the

findings and ensure the amounts of expressed protein had been assessed accurately.

With a western blot as protein will be denatured and inactive any affects of the culture upon the protein, for example degradation, should be minimised and therefore the western blot may give a more accurate depiction on protein production. For western blots the cell culture was lysed, and the supernatant was added neat to the SDS page gel and separated by electrophoresis. In western blots samples were not corrected for pH before being added into the experiment as the western blot does not need the protein to be active to give a signal. First a control was studied by using 2 different plasmids for *Gaussia* luciferase with different signal sequences. Control cells were also used with no plasmid present to check for any cross reactions of antibodies used.

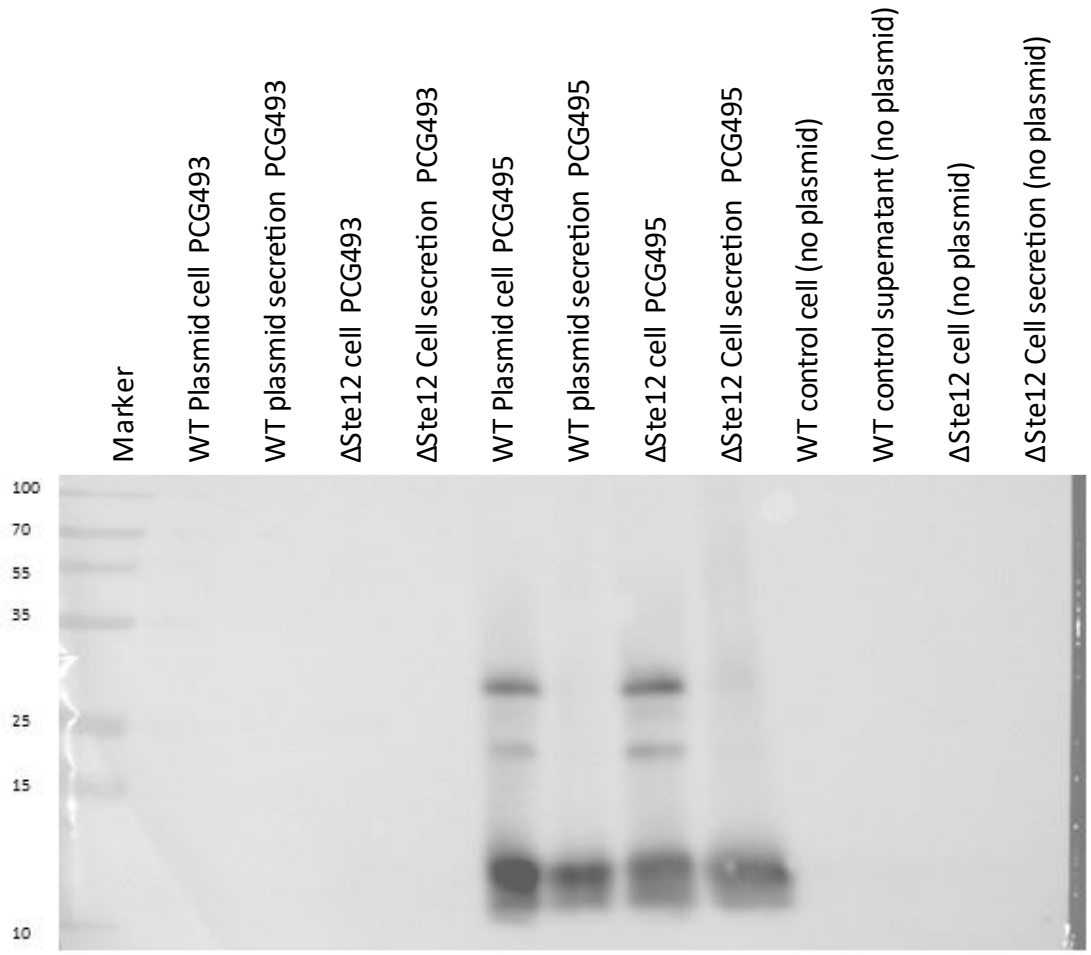

Figure 8: Western blot probing for Gaussia luciferase to check for cross reactivity. Samples 10, 11, 12 and 13 all show no signal indicating the antibodies used have no cross reactivity with the samples. Loading control is absent. Plasmid 1 is PCG493 and plasmid 2 is PCG495.

Only cells transformed with PCG495 gave notable signal, there appears to be a signal in lane 10 for the WT control cells however it is likely this is due to overspill of protein in loading of the original SDS page gel, as this was an anomaly and never seen again. Further testing of Gaussia luciferase either in western blots or commercial assays were therefore conducted using cells transformed with PCG495 plasmid as this gave the most reliable Gaussia luciferase signal.

It was then decided to test multiple cultures of cells for Gaussia luciferase via western blot to try and compare differences between the strains further and gather an accurate understanding of which cell was producing the most Gaussia luciferase.

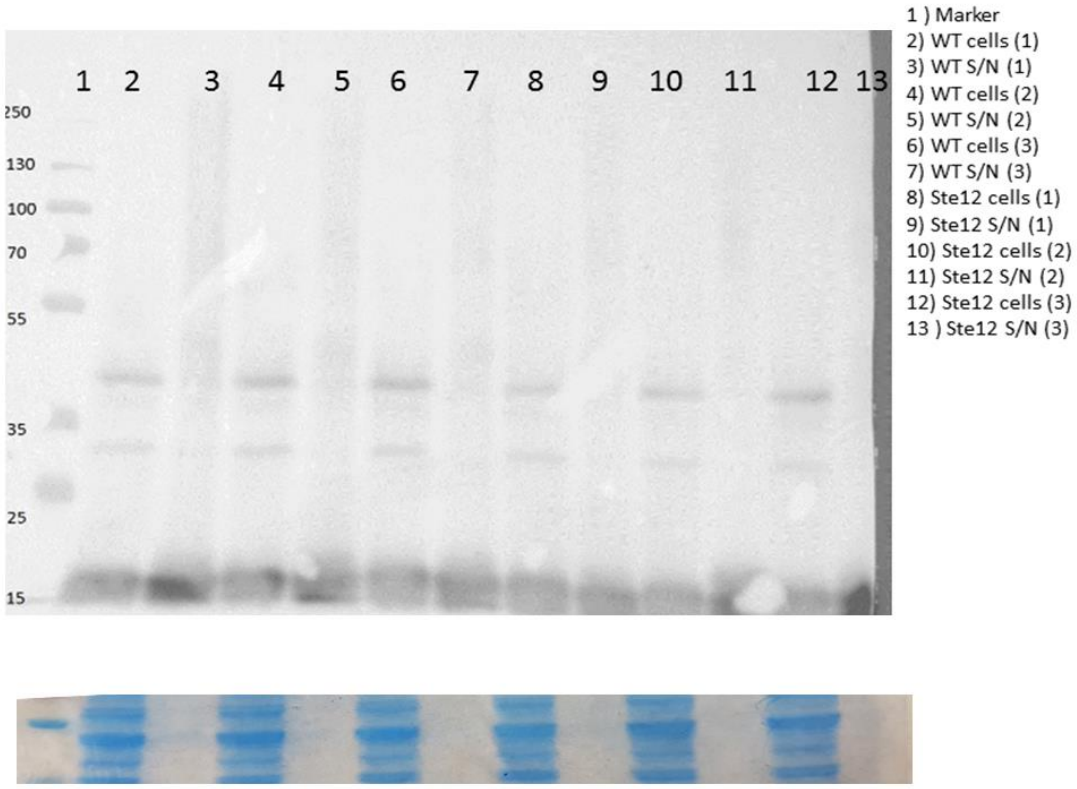

Figure 9 Western blot showing Gaussia luciferase present in either the cell preparation or the supernatant in either WT (BY4741) or ste12 mutant cells with a Gaussia luciferase plasmid transformed. Samples were separated on 10% SDS gel. Cell extracts were prepared by boiling in lysis buffer and supernatants prepared by TCA precipitation. Samples were prepared in triplicate, loading control is absent.

The western blot seen in figure 9 does not appear to show large differences in the amount of luciferase being produced between the WT and knockout cells, however as the Gaussia luciferase protein is quite small at ~20kDa, during gel separation it will travel further towards the bottom of a 10% SDS gel. To enable to see the Gaussia luciferase better, the experiment was repeated with a 12.5% gel which enables smaller proteins greater separation.

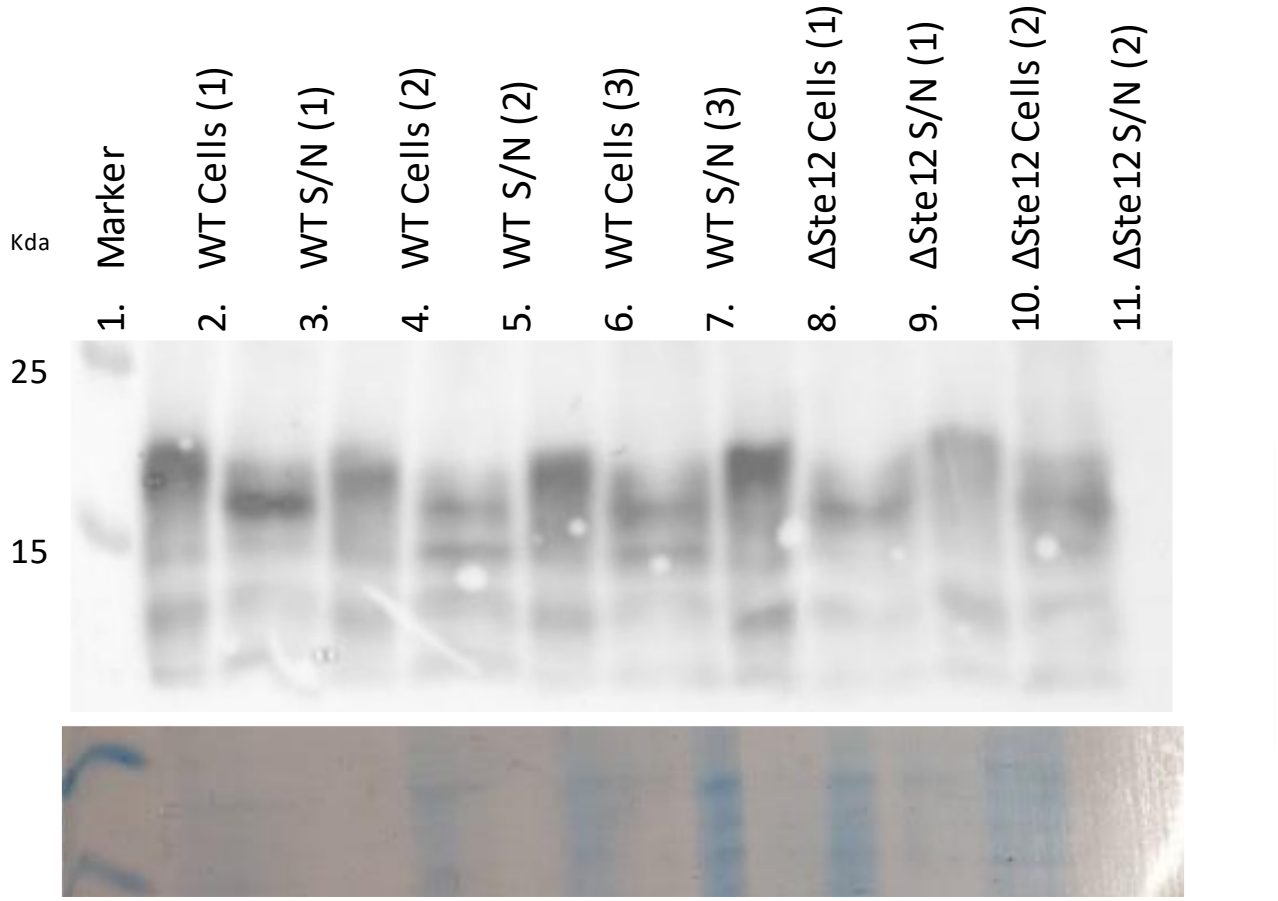

Figure 10: Western blot showing Gaussia luciferase present in either the cell preparation or the supernatant in either WT (BY4741) or ste12 mutant cells with a Gaussia luciferase plasmid transformed. Samples ran on 12.5% SDS gel. Cells extracts were prepared by boiling in lysis buffer and supernatants prepared by TCA precipitation. Samples were prepared in biological triplicate, loading control is absent.

Repeats of the *Gaussia* luciferase western blot on a higher percentage SDS page gel showed that the *ste12* deletion strain appears to give weaker bands at the G*aussia* luciferase size. This matches the findings of the commercial assay in figure 7 and implies that removal of *STE12* impedes protein production.

3.2.7 CGY384 background

A different cell background (Matα *ura3-52 his3∆200 Leu2-3,112 lys2-801 ade2-101 cof1::LEU2*) was chosen as the original BY4741 cell line used appeared to contradict other results and hypotheses . Previous work within the group had indicated this strain may improve secretion and as cof1 acts upon sorting and export of cargo from the golgi body[254],this could allow for higher secretion levels. As initial findings had shown *STE12* deletion to have a negative impact upon Gaussia luciferase production it was hypothesised that these differences may be dependent on cell strain. To analyse this a further Gaussia luciferase assay, western blot and growth analysis was carried out comparing the original by4741 cell line to a different CGY384 cell line. Initially a western blot was conducted to check for differences in Gaussia luciferase production in the wild type cells of the new strain (CGY384) compared to the original strain (CGY424). The original BY4741 background which appears to be a generally poor secretor of proteins and it was thought different cell lines may be better at secretion to further show differences.

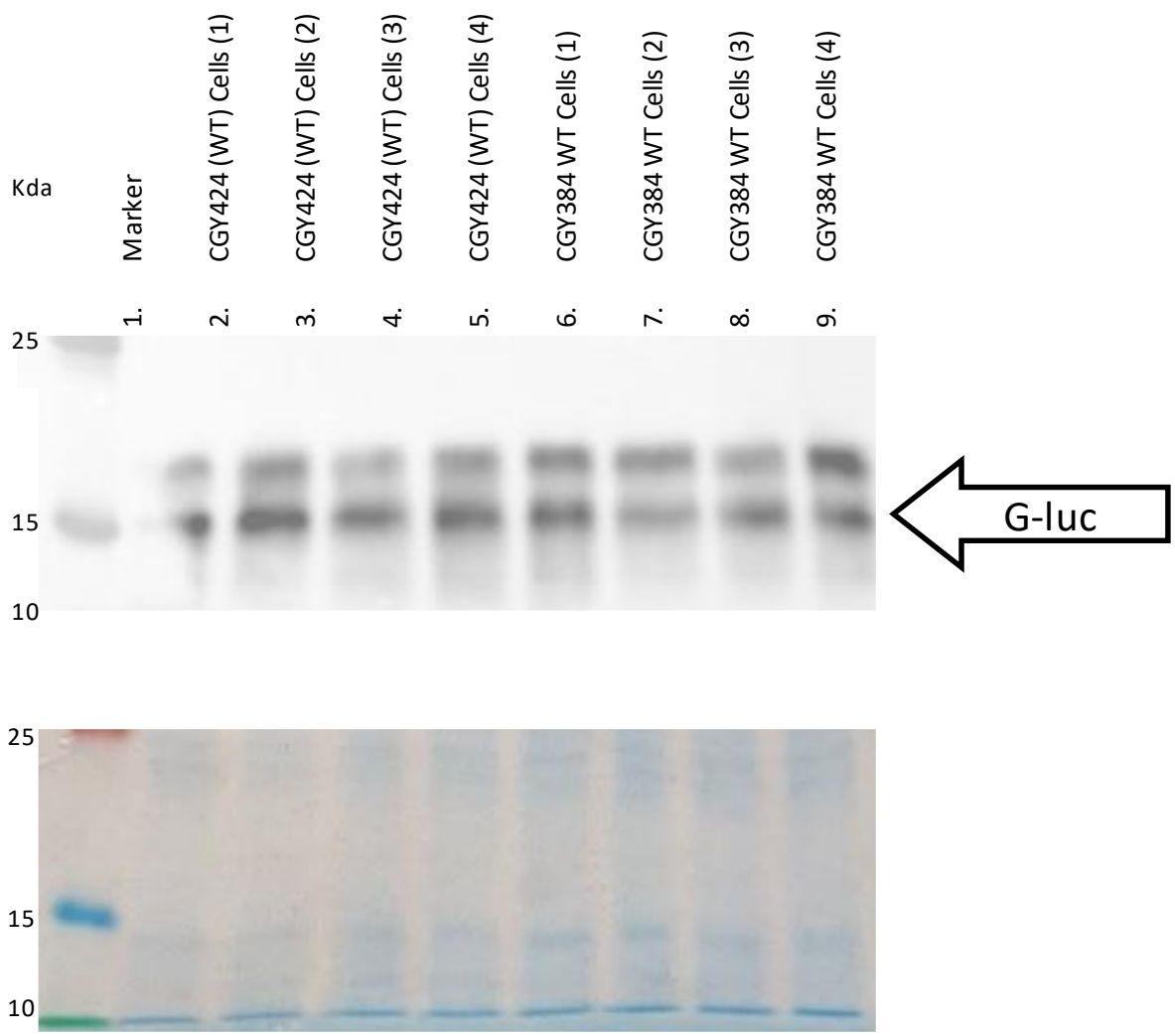

Figure 11: comparison of 4 unique transformed colonies for Gaussia luciferase production via western blot. Each cell samples were boiled in lysis buffer before loading onto a 12.5% SDS Gel. Loading control is absent.

The western blot shows similar levels of *Gaussia* luciferase production in the original background compared to the new background with all cells giving a strong band for *Gaussia* luciferase in the expected molecular weight region.

Previous results had shown that the *ste12* deletion strain grows at much faster rates than the wild type cells, although this was not seen in the BY4741 background used

previously (figure 4 and 5). We therefore repeated the growth rate analyses with the BY4741 background the new CGY384 background in parallel.

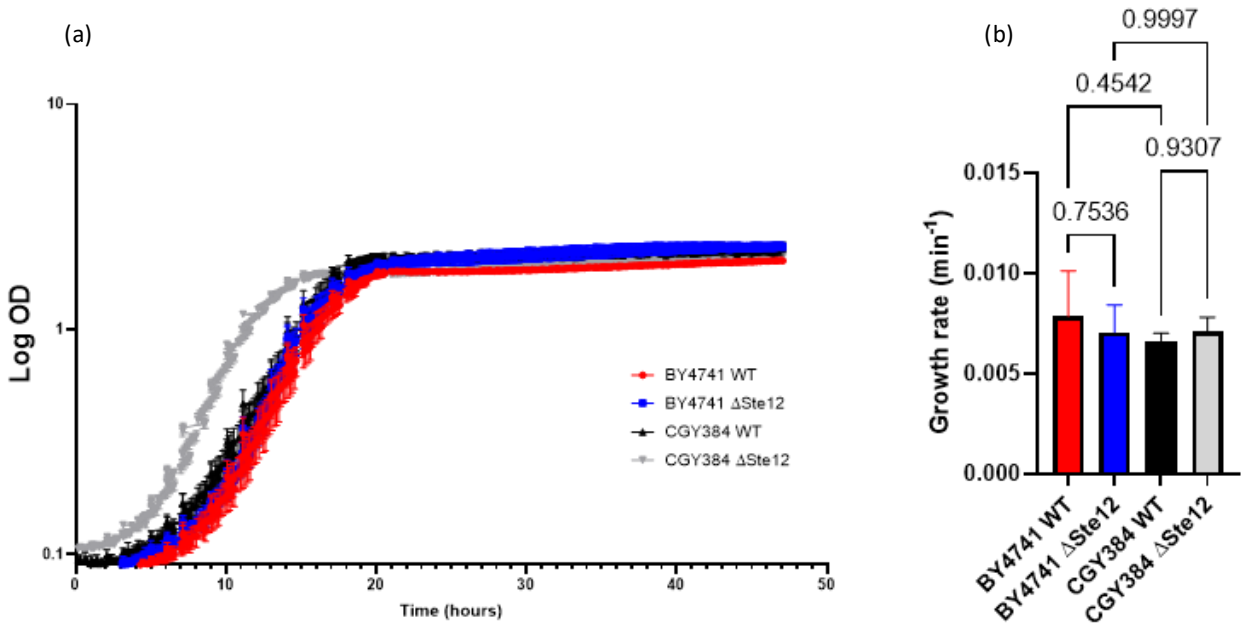

Figure 12: Growth rate comparison of BY4741 cells with CGY384 cells with and without the Ste12 mutation. (a) Raw OD values average of n=6) with error bars for SEM. (b) Growth rates were calculated by selecting the highest growth rate observed over a seven timepoint series. 2-way ANOVA with Tukey's post-hoc test showed no significant significance between the calculated growth rates; P values indicated on bar graph. Error bars show the standard deviation.

Visual inspection of the growth curves showed that the optical density of the CGY384 strain containing the *ste12* deletion increased more rapidly than for the other strains tested, although when growth rates were calculated there were no statistically significant differences. Thus, the *ste12* deletion appeared to shorten the lag period before full growth rates were reached, but did not significantly affect the growth rates themselves. It is possible that the decreased lag time was interpreted

as increased growth in the original experiments, hence leading to the apparent difference in results with our work.

3.2.8 Comparing cell backgrounds for *Gaussia* luciferase production

It was then decided to see whether the production of *Gaussia* luciferase was the same for CGY384 with and without the *ste12* deletion compared to the original BY4741 based strain. The original observation of increased *Gaussia* luciferase production in the *ste12* deletion background had been made in the CGY384 background, and we therefore repeated this experiment I the same background. As before results were used to compare intracellular luciferase to secreted luciferase.

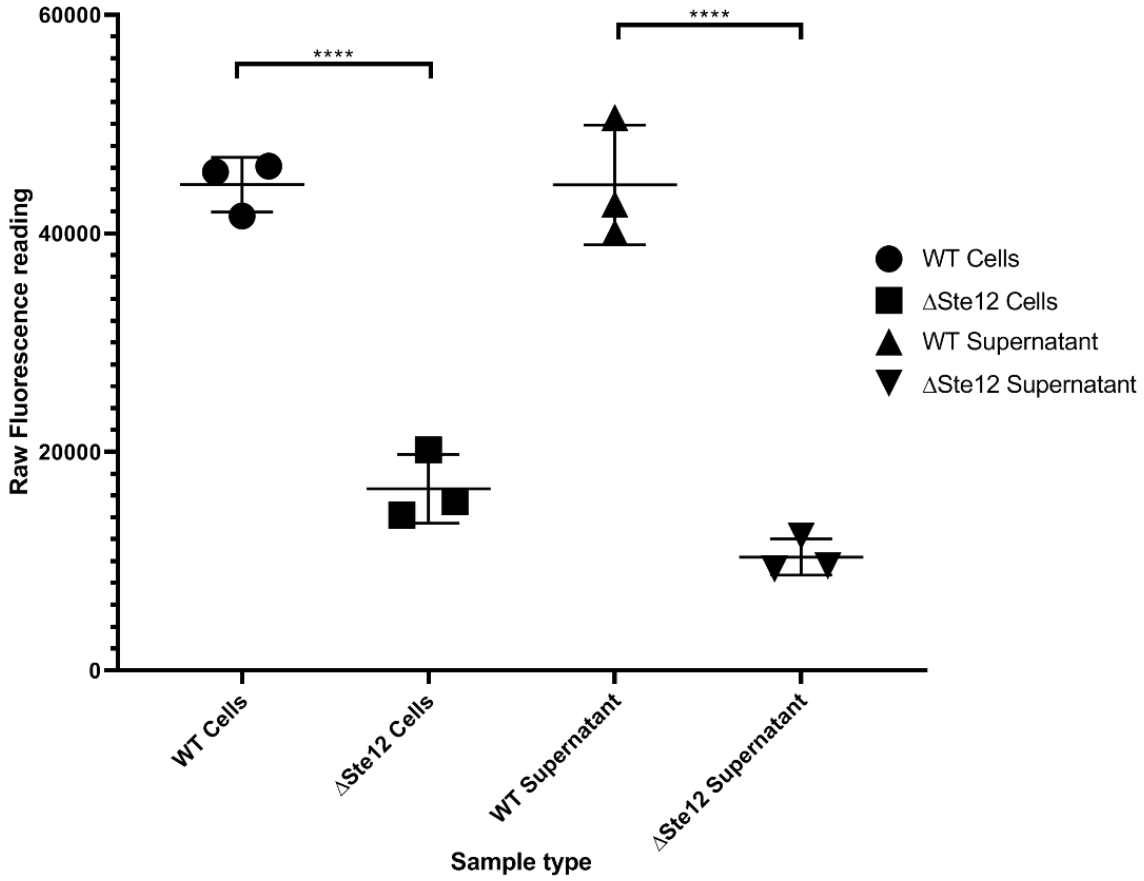

Figure 13: Gaussia luciferase data of 3 biological repeats (n=3) of CGY384 cell line either wild type or ΔSTE12. The figure shows individual colony data for 3 different

transformed colonies of each cell type performed in triplicate (dots), as well as averages and standard deviation. 2-way ANOVA with Tukey's post-hoc test were used to analyse significance.

The new strain still showed significantly reduced levels of *Gaussia* luciferase production upon deletion of *ste12*, which was also confirmed by western blot (figure 14).

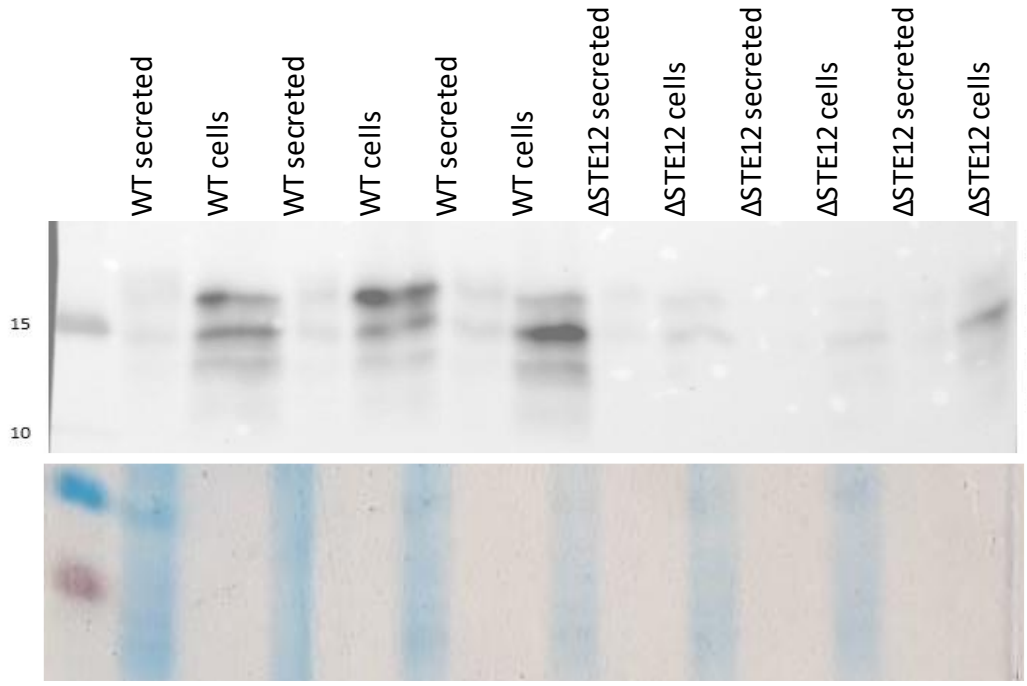

Figure 14: Western blot of proteins ran on a 12.5% SDS PAGE gel transferred onto nitrocellulose membrane and probed with 1:1000 Anti-Gluc followed by 1:5000 Anti-Rabbit-HRP. Key shows cell type, sample type followed by the number of the colony as per the restreak plate. Cells samples prepared by boiling cell extracts in lysis buffer, supernatant samples prepared by 1mL TCA precipitation of proteins. Loading control is absent.

Western blot showed reduced production of *Gaussia* luciferase in the *STE12* mutant. This confirmed data seen in figure 13. The *STE12* mutation seemed to have an overall negative effect on protein production.

Comparing both cell lines, Ste12 appears to reduce production of recombinant luciferase proteins as a rule. This appears in both the active protein detected via the assay and the overall protein levels revealed in western blots. Interestingly with all strains regardless of whether *STE12* was present or absent, the majority of luciferase appears to remain within the cellular mass and so neither strain appears to be secreting this protein particularly actively.

3.2.9 RNA analysis

As follow-up work to the original experiments which had suggested positive effects of *ste12* deletion on growth rates and recombinant protein production, RNA samples of STE12 wild type and *ste12* wild type and *ste12* deletion strains had been collected and subjected to next generations sequencing analysis by Illumina sequencing (RNA-Seq). Despite the fact that the current study was not able to confirm the positive effects on recombinant protein production, we used the availability of these not yet analysed data to further investigate what effects the deletion of *ste12* had on the biology of yeast cells and how this might relate to effects on recombinant protein production.

The RNA-Seq data were analysed using a standard Bioinformatics pipeline provided by the *Galaxy* facility[255]. Reads for *ste12* deletion and comparable wild type cells were initially concatenated into one single dataset per condition.

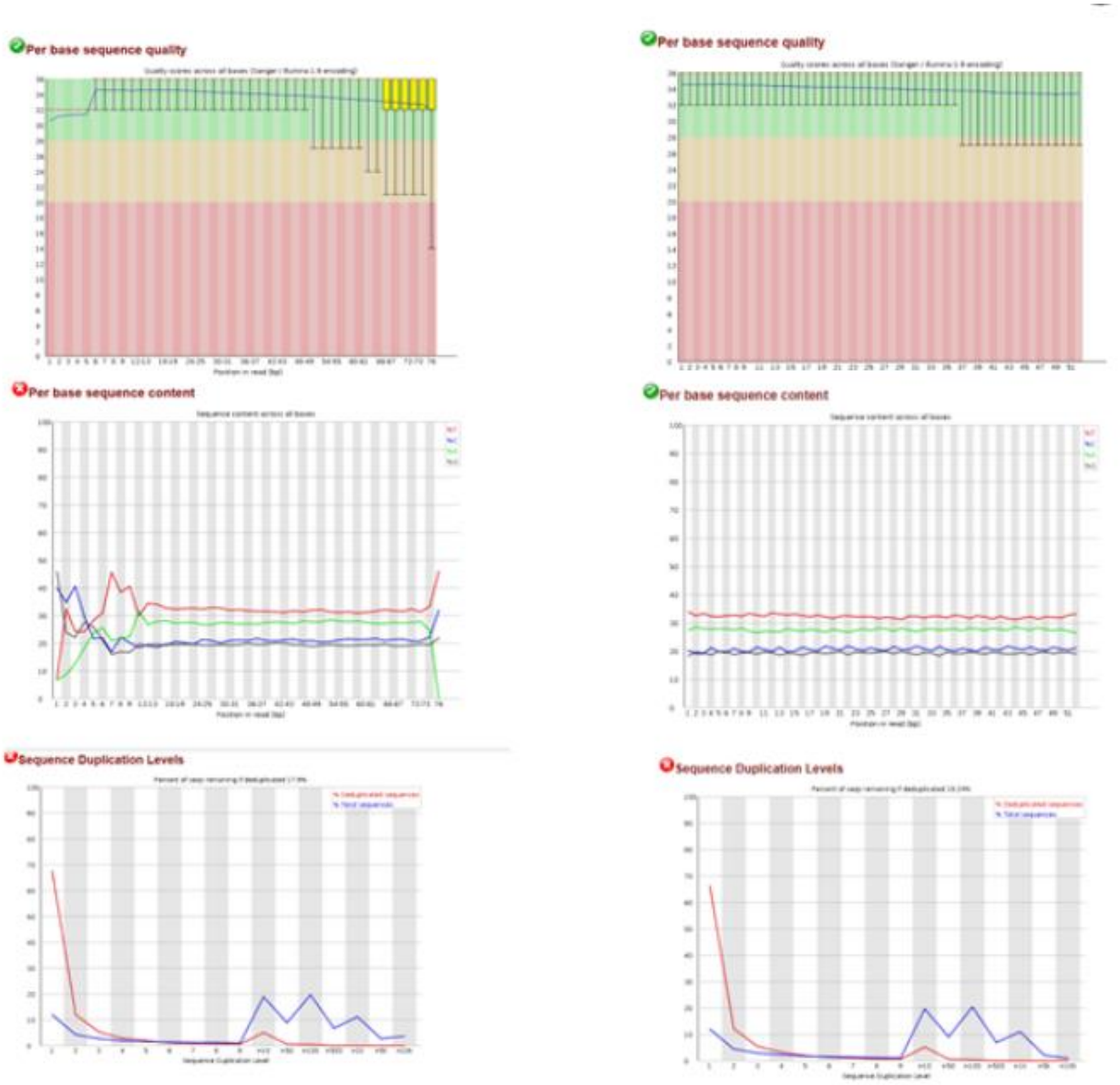

Figure 15: FastQC data showing per base sequence quality (top) per base sequence content (middle) and sequence duplication levels before (left) and after (right) trimming of data.

FastQC was used to determine the overall sequence read quality (figure 15) and showed high read quality along the full read length (~55 nucleotides) following removal of adapter sequences using TrimGalore!. Sequence duplication levels remained high because read start sites are non-randomly distributed due to selection of primers, preparation protocols and high read coverage of highly expressed genes.

Reads were then aligned to the *S. cerevisiae* genome (version sacCer3) using HISAT2[231] and expression levels of individual genes determined using FeatureCounts. Finally, Gene Set Enrichment analyses[256] were conducted to identify cellular pathways that were affected by deletion of the *Ste12* gene.

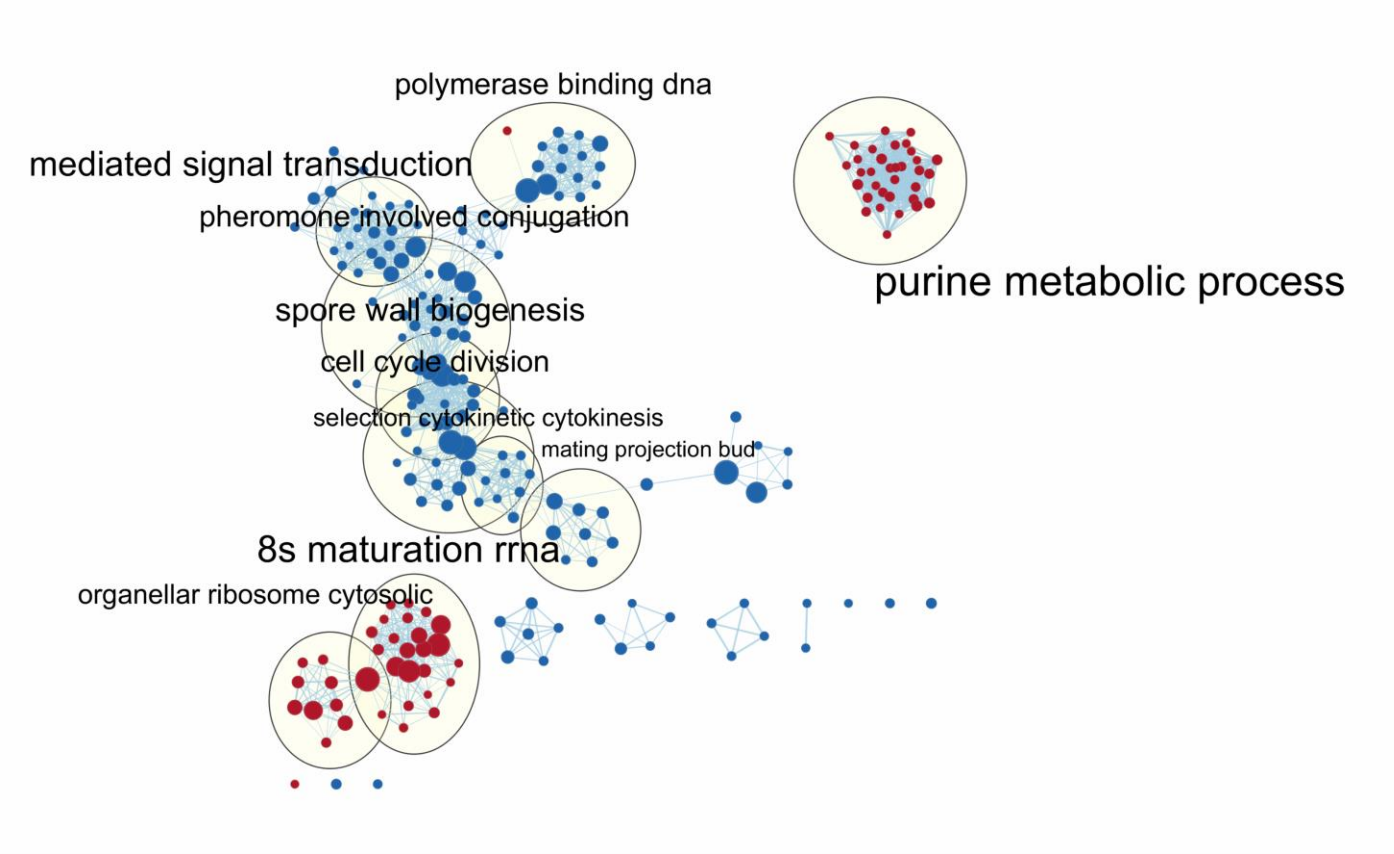

Figure 16: Gene set enrichment map of genes changed in expression upon deletion of STE12 with upregulated genes shown in red and downregulated genes shown in blue.

RNA analysis suggests that the STE12 mutation alters cellular processes involving ribosomes (figure 16), leading to upregulation of ribosomal protein encoding RNAs (cluster “organellar ribosome cytosolic in figure 16) as well as RNAs encoding proteins involved in the maturation of ribosomal RNA (cluster “8s maturation rna” in figure 16). There have also been increases in catalytic activity particularly involving ATPase activity indicating the cell is using more energy. The increase in ATP energy

and ribosomal proteins may increase protein production seen by upregulation of purine metabolic processes. Gene set enrichment shows that mating, cell cycle division and pheromone activity has been downregulated, which is expected as these are functions known to be stimulated by Ste12 activity[257].

3.3 Discussion

3.3.1 The effect of *STE12* mutant on Growth

For the BY4741 strain the *STE12* mutation has no significant effects on growth (figures 4 and 5). For the CGY384 strain the growth of the *STE12* mutant cell line does change (figure 12), although the strongest effects appears to be a shortening of the lag time before logarithmic growth is reached whereas there is no significant effect on the calculated growth rates. Transcriptome analyses showed that cells with a *ste12* deletion show downregulated cell cycle division and mating, but increased purine metabolic processes (figure 16). The downregulation of multiple genes involved in cell cycle would generally imply reduced cell division rates, and therefore it is interesting that cell growth rates do not seem to be significantly affected by the *ste12* deletion. This could be due to higher levels of ribosomes and purine metabolic processes, meaning that even with a downregulated cell cycle, cells are more effective at producing necessary proteins and energy for growth and leading to an overall enhancement.

Purines are metabolic substrates which provide necessary components in DNA and RNA building. High concentration of purine metabolites can be found in tumour cells and purine antimetabolites have been used to block DNA synthesis and halt growth. Purinosome (The multi enzyme complexes thought to carry out de novo purine synthesis within a cell) within purine metabolism is closely related to growth [258],

when purine is depleted in the cell then the biosynthetic purine pathway forms a complex called the purinosome which shows that purine depleted cells have larger purinosome formation in the G1 phase of the cell cycle[258]. This indicated that the purine metabolism has impacts upon the cell cycle. Purine and pyrimidine are heavily involved in energy and synthesis of nucleotide cofactors. Nucleotide synthesis pathways are conserved between animals and microorganisms. Purine metabolism can be impaired in cancer, indicating that impaired purine metabolism can lead to an increase in proliferating cells[259]. As the data in figure 16 shows an increase in purine metabolism but downregulation of cell growth, this indicates that the increase in purine metabolism is either overall more effective at increasing growth than mating and cell cycle genes, or that the increases seen in these genes are overall much larger than those affected by mating and cell cycle genes. This is surprising as Ste12 is known to be implicated in mating and pseudohyphal growth, but not in purine metabolism. The slight increase in growth seen in figure 12 is therefore unexpected and most likely due to the increased RNA profile of genes involved in ribosomes and purine metabolism.

Nucleotides can be synthesised by either denovo or salvage pathways, denovo pathways build nucleotides from simple molecules and has high energy requirements, with 5 ATP requiring steps compared to only one in the salvage cycle[260]. As pathogens within the blood grow, they must adjust the metabolism according to the availability of nutrients. Some of these nutrients have low availability and therefore must be synthesised within blood bore pathogens. *E. coli* species found to have a deletion of a gene within the purine or pyrimidine nucleotide biosynthetic pathways showed a huge drop in viable cell counts[261] after 24 hours growth in human serum, some species of gram-negative pathogens are unable to

grow in human serum when biosynthetic pathways are blocked[261]. As the *ste12* deletion reduces the ability of cells to be virulent due to inhibiting pseudohyphal growth, it may be that these cells acquire more ribosomes and purine metabolism to counteract what they may normally seek in the external environment.

Cell growth and ribosome production processes can also be tied together. As there is some evidence to show the *ste12* deletion strain, at least In the CGY384 background (figure 12) can reach maximum growth rates faster than the WT cells, and the RNA data show that cytosolic ribosomes are upregulated, it could be this upregulation in ribosomal proteins is responsible for the changes in culture growth behaviour.

3.3.2 the effects of Ste12 on protein levels

Ribosomes are responsible for producing cellular proteins and it is believed that transcriptional stress can impact both production of ribosomes and cellular growth. In *E. coli* it has long been known that ribosome content is proportional to growth rate[262], with studies in the 1960's comparing rates of RNA and DNA production. The data shows ribosomal proteins are upregulated, but does not necessarily mean higher levels of functional ribosomes are present[263]. We do not see higher levels of rRNA in the microarray, and with no upregulation of RNA polymerase 1 seen, functional ribosomes may not be upregulated which could explain why increases in recombinant protein levels were not seen. In mammals, impaired ribosomal synthesis has been associated with cancer, this is hardly surprising as many protein synthesis pathways are linked with cellular signalling pathways. Often these pathways rely on c-Myc in mammalian cells with both promotes growth and stimulates Pol for ribosome biogenesis[264]. The connection between increased

ribosomes and increased growth would potentially explain the perceived growth increase in the *STE12* knockout.

Figure 16 Shows a downregulation of polymerase binding DNA, and therefore it could be that this is reducing protein production even with upregulation of ribosomes and purine metabolic processes. As the binding of RNA polymerase to DNA is one of the first steps in protein production, it may be that the reduction in this binding is causing the reduction in the luciferase reported noted in the *STE12* knockout cells (figures 7, 8, 9, 11 and 13). Computational models suggest that within healthy yeast cells availability of ribosomes is the rate limiting step for protein translation[265], and therefore it is unexpected that the *Ste12* mutant cells are producing less luciferase than their wild type models.

The origin of this part of our investigation had been reports that deletion of *ste12* could enhance recombinant protein expression in yeast. Ultimately this could not be confirmed in our data as the findings all indicate that *ste12* deletion reduces recombinant protein expression and hence *ste12* deletion is not a viable option for engineering of higher protein production and secretion in *Saccharomyces cerevisiae* based on this data.

Chapter 4 : The effects of GPx proteins

4.1 Introduction

Production of recombinant proteins within yeast cells is reliant on multiple biological pathways. Following the central dogma through from initial transcription of DNA, through translation into peptide and through folding into the correct 3D structure hundreds of cellular processes can either directly or indirectly affect the final quality and quantity of produced protein. Protein disulphide isomerase has an important role in helping proteins fold correctly and adopt their native tertiary structure. PDI enables formation of disulfide bonds in the ER due to oxidation of thiols[266].

PDI activity is provided by Pdi1 in yeast. Pdi1 becomes reduced in its function and requires reoxidation by a single enzyme: Ero1. PDI1 in yeast has 4 genes which show homology, MDP1, MDP2, EUG1 and ESP1. Generally these do not enable restoration of viability upon deletion of PDI1, excepting MDP1[267], but appear to carry out some of the functions of Pdi1. Exploration of mutants in PDI1 homologues have shown that they can lead to defects in protein folding and glycan modification[267], and show the importance of Pdi1 functions surrounding protein folding within *saccharomyces cerevisiae*.

Pdi1 can have 2 major functions, de novo formation of bonds and isomerisation. These 2 functions allow translated peptides which have been sequestered into the ER to be correctly folded. The active site of PDI transfers disulfide bonds onto proteins and becomes reduced itself. After PDI is reduced it can be re-oxidised by Ero1. This process forms a molecule of hydrogen peroxide. PDI contains 2 domains which are able to be oxidised by Ero1, the a' domain which is believed to catalyse substrate oxidation and the a domain which allows isomerisation of disulphide bonds[201].

The cysteine containing active domains of PDI can be present in a semi oxidised form where the active site in either the a or a' domain is oxidised[202]. The redox state of the thiol groups in PDI depends on which actions it is able to perform. The action of PDI is essential for all disulfide bond creating proteins. This makes it of particular interest when considering targets to improve protein production. Increasing the number of disulfide bonds which need forming, can increase the production of hydrogen peroxide and cause oxidative stress[268]. Manipulation of the oxidative folding pathway could be a method to control this.

In mammalian cells the PDI system is more complicated. There are more PDI related activities in mammals. In yeast, deletion of PDI1 in yeast is considered fatal for cells3, although viability can be restored by expression of the EUG1 gene which has homology with both yeast and human PDI, although is non-essential in itself6.

The more complex human oxidative protein folding pathway has a wider variety of ER resident proteins including peroxiredoxin IV which can act in the oxidative-reductive processes with PDI enabling human cells to remain viable even without a functional ERO gene7, only showing a slight delay in disulfide bond formation. In mammalian cells the oxidative protein folding system with PDI is also benefitted by the presence of Glutathione Peroxidase (GPx) proteins. These proteins allow another method of oxidation for PDI. Increasing the oxidation levels of PDI should enable more PDI functionality and could potentially reduce oxidative stress. The higher functionality of PDI enables more proteins to be correctly folded and removed from the ER.

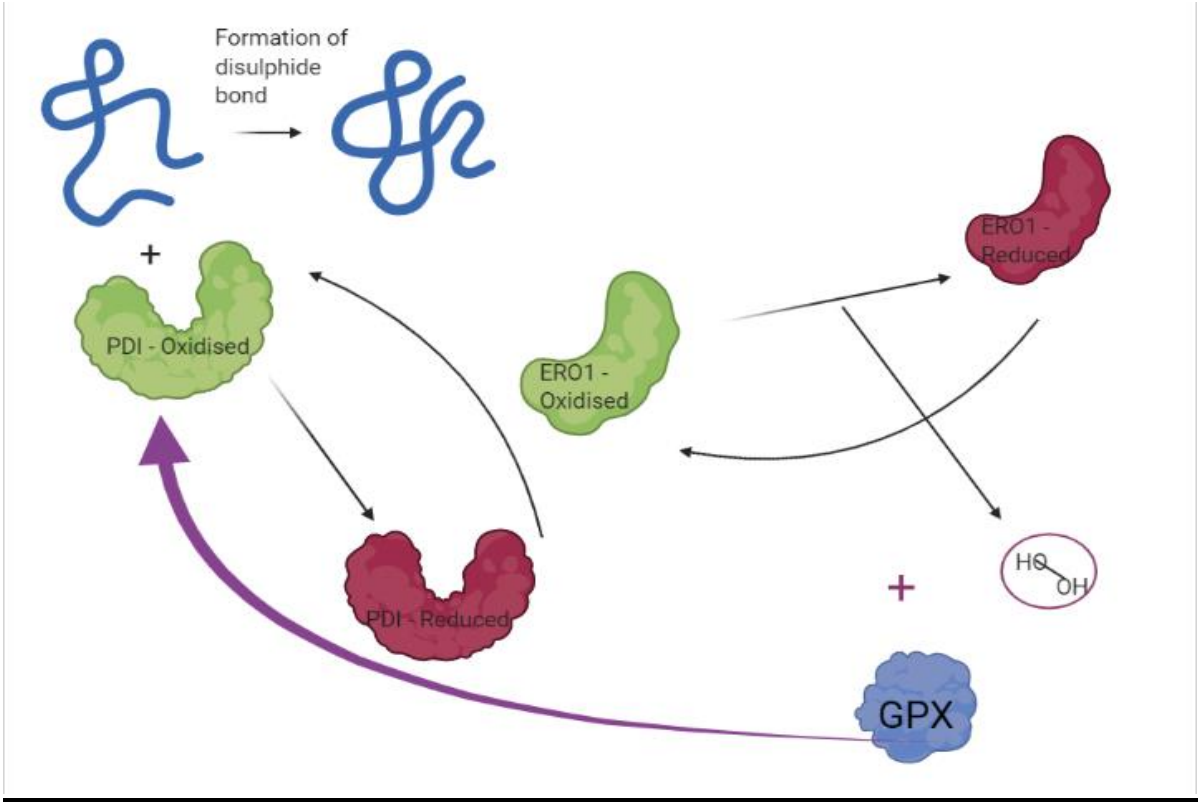

Figure 17: GPX proteins form part of a REDOX system with PDI and ERO1 in mammalian cells.

Glutathione Peroxidase (GPx) proteins GPx7 and GPx8 reside in the ER of mammalian cells. As shown in figure 17 they interact with the hydrogen peroxide produced in the oxidation of Ero1 to directly oxidise PDI. This has the dual benefit of more available, oxidised PDI and the removal of hydrogen peroxide, which can cause oxidative stress to the cell. GPx proteins have been shown to directly use the hydrogen peroxide generated by Ero1 to reoxidise PDI[217], and appears to prevent the diffusion of H_2O_2 created by Ero1 from diffusing around and outside the rough ER[216].

GPX proteins have been known for a long time, being originally described to react with reduced glutathione in 1957[213]. Further research since the initial discovery,

however, has shown that GPx7 and GPx8 actually use thiols more efficiently than glutathione as is seen when they interact with cysteine groups in PDI[197,214].

GPx proteins are vital in many mammalian cells to maintain a redox balance, however not all eukaryotic cells contain GPx proteins. Firstly some mammalian cells themselves do not contain GPx7 and GPx8. Expression of GPx7 or GPx8 in mammalian cells which do not normally contain these proteins have been indicated to reduce H_2O_2 generation, ER stress, and apoptosis induction[269]. With the implications on the protein folding pathway, and the positive benefits seen in mammalian cells upon addition of GPx, the effects of recombinant GPx 7 and 8 upon yeast cells remained unknown.

Increased levels of protein-disulfide isomerase (PDI) have been shown to improve recombinant protein production in yeast[190,200] . As GPx proteins interact with human PDI, and biochemical work outside of the lab (unpublished) has shown that GPx can interact with yeast PDI, we wished to consider how GPx proteins interacted in a yeast system. As increasing Pdi1 can help recombinant protein production we wondered whether introducing the GPx proteins into yeast could have similarly beneficial effects. Biochemical work in the lab has shown that human GPx7 can increase the reoxidation of yeast Pdi1 in vitro (Dave Beal, unpublished), and so it was hypothesised that expression of recombinant GPx in yeast could increase recombinant protein production, and reduce oxidative stress.

4.2 Results

4.2.1 expression of GPx

GPx proteins have previously been shown to improve redox balances in mammalian cells which produce recombinant protein[269] . To produce the GPx proteins successfully in yeast, recombinant sequences had been created which encode GPx7 or GPx8 but have been codon optimised for yeast expression[223]. The sequences encoding GPx proteins were cloned into a centromeric yeast expression vector using a constitutive promoter. The code contained a yeast signal sequence comprised of a MATalpha pro-sequence and Ost1 presequence[270]. The plasmid DNA was also designed to contain a C-terminal HA tag and yeast specific ER retention sequence (HDEL). The plasmid DNA was introduced into yeast and retained using *LEU2* as selectable markers.

GPx proteins were hypothesised to increase yields of recombinant protein in *Saccharomyces cerevisiae* by improving availability of oxidised PDI. To test the impact of GPx proteins upon recombinant protein production in yeast after successful DNA transformation it was necessary to detect whether they were being produced. Initially human GPx proteins were co-transformed into BY4741 yeast cells using plasmid DNA alongside the *Gaussia* luciferase plasmid used throughout this project. To test whether the GPx proteins were being produced by the cells, a western blot was used to detect the anti-HA tag upon them. Samples for GPx7 and GPx8 were compared to controls and size markers were used to indicate whether

the cells were successfully producing the correct GPX protein.

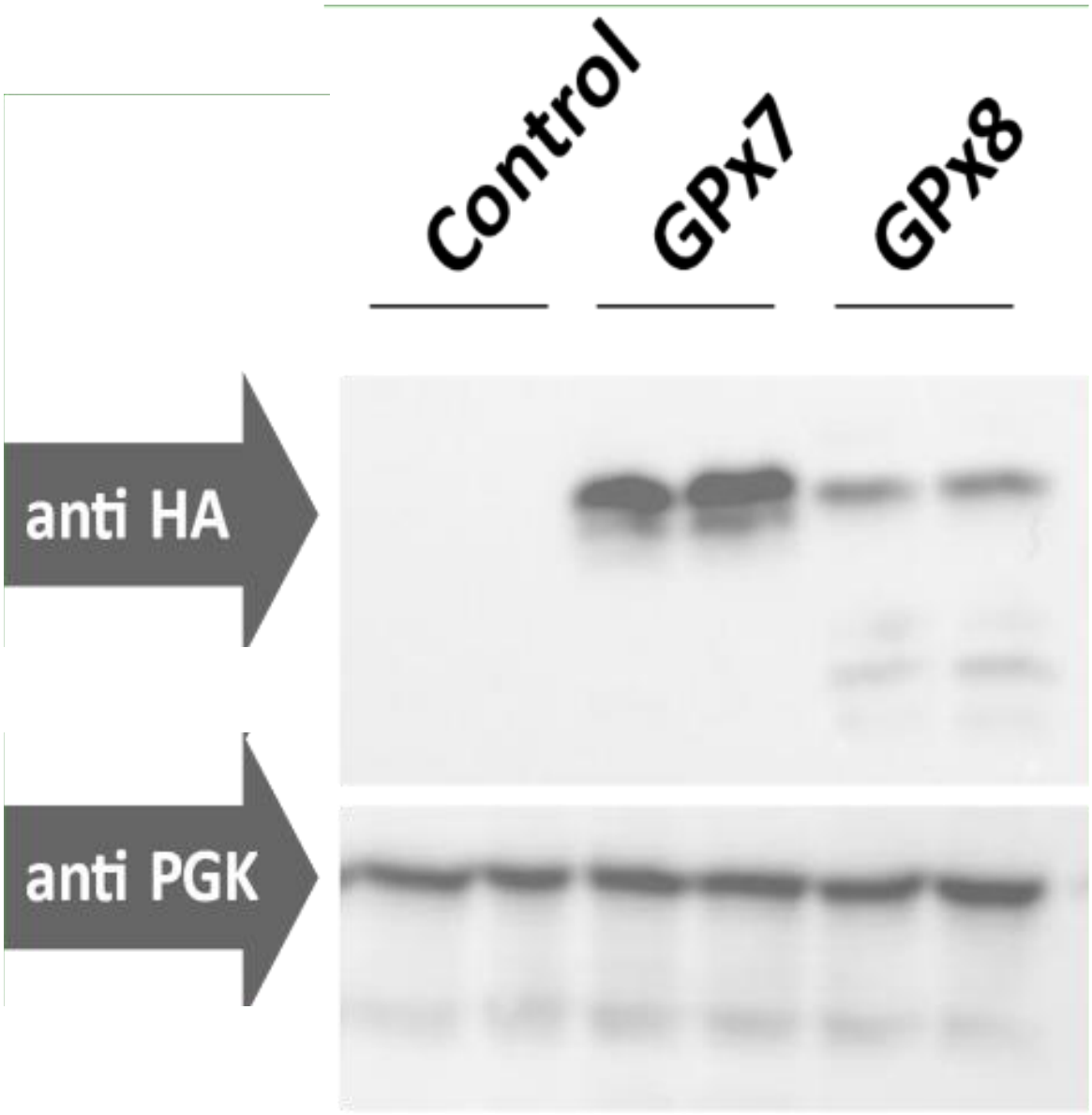

Figure 18: Cell lysates were probed via immunoblotting for presence of GPx proteins. 2 cell lysates from individual transformants shown. Samples were also probed using antibodies raised against PGK to show equal loading.

4.2.2 Expression of luciferase in GPx cells

Western blotting showed that cells appear to be producing GPX proteins successfully and that these are not present in control cells. Next, it was decided to test how the GPx proteins impacted recombinant protein production. As the GPx

proteins interact with the PDI pathway it was thought that higher levels of oxidised PDI may be present and that this could lead to higher levels of protein production. The production of intracellular and secreted *Gaussia* luciferase was measured by commercial *Gaussia* luciferase activity assay to discover what effect GPx had on the cells ability to produce and secrete luciferase proteins. *Gaussia* luciferase is known to contain five disulfide bonds[271], meaning that it is likely to require both the *de novo* folding and isomerase activities of PDI. The cells with transformed GPx proteins were compared to control cells, containing a plasmid that allows them to grow in the same selective media as the producing cells. Samples were tested to give a raw reading of luciferase content and then normalised. To normalise the data the luciferase reading was considered to be 1 for the control cells and supernatant individually for simple comparison upon addition of GPx.

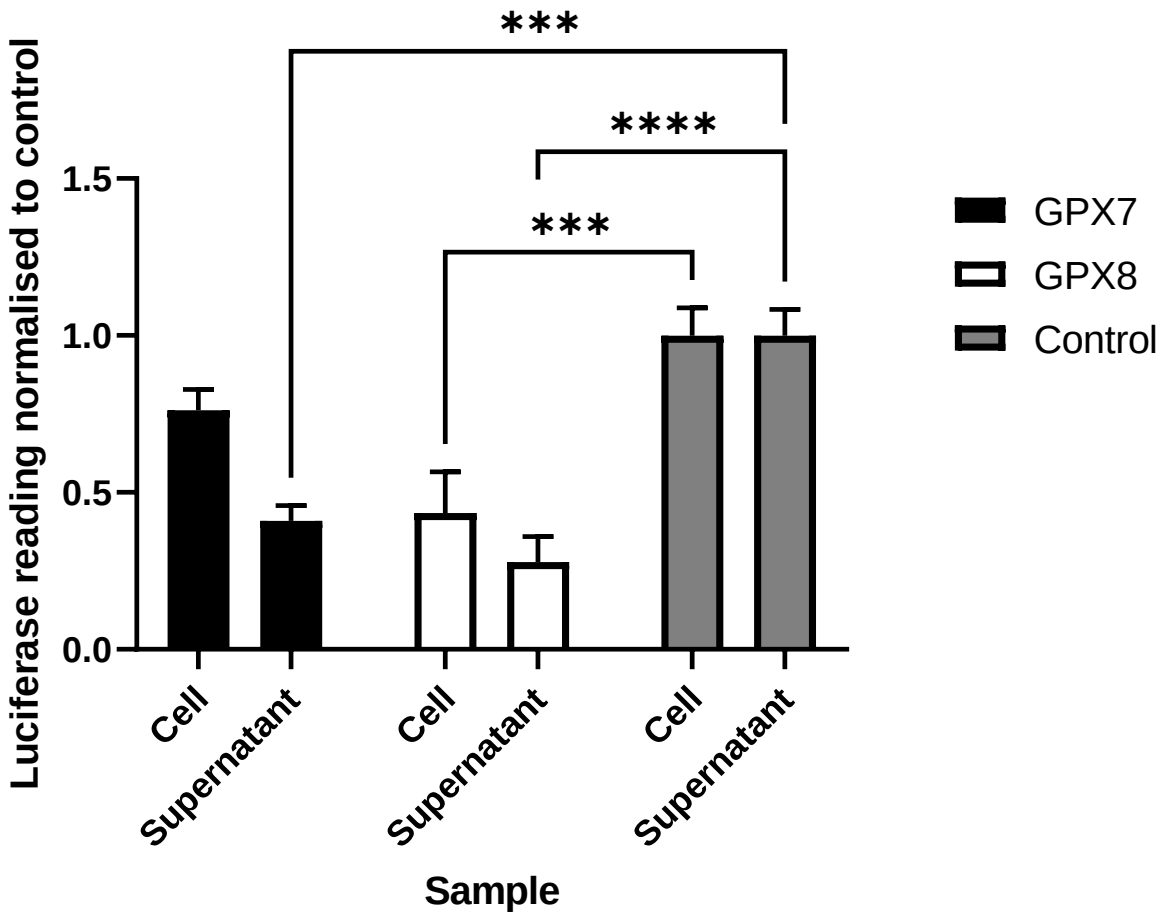

Figure 19: GPx7 and 8 appear to reduce Gaussia *Luciferase production both intracellularly and secreted. Error bars for mean +SEM. N=12. Control samples normalised to 1 for both cellular and secreted signal individually. 2 way ANOVA plus Tukeys multiple comparisons test was used to test for significance of differences*

It was unexpected that GPX proteins appeared to reduce levels of produced luciferase in both the cells and within the supernatant. Previous literature indicated that GPx proteins may support recombinant protein production in human cells, it was though that GPx proteins may allow higher levels of oxidised PDI to assist protein folding in the ER, it was expected this would lead to higher protein yields. The significantly lower levels of cellular and secreted luciferase for cells transformed with GPx8 (P= 0.0003 and P= <0.0001 respectively) and secreted levels of luciferase seen in GPx7 (P= 0.0001) show that GPx do not increase recombinant protein

production. GPx7 appeared to lead to a small reduction in intracellular luciferase, but this was deemed insignificant (P= 0.3909).

4.2.3 Flow cytometry on GPx strains

Presence of GPx proteins was expected to increase PDI functionality leading to increased recombinant protein production, as this did not occur we wished to investigate the effects of GPx further. To investigate these effects and in an attempt to explain the unexpected results, we wished to investigate how the intracellular pathways involved within recombinant protein production may be affected by the GPx proteins. It was thought that the effects of GPx on the Pdi1 and Ero1 redox systems could lead to a lower level of oxidative stress within the cell. To test this we used flow cytometry staining with both propidium iodide (PI) to test for viability, and H_2DCFDA to test for oxidative stress (ROS). Cells containing GPx, control cells containing just a selection plasmid, and positive control cells known to be at high levels of oxidative stress were compared for ROS. Due to low levels of stress in the system 1.5mM hydrogen peroxide was added to all cell lines to increase oxidative stress levels with the aim of identifying the differences in how the strains react to high levels of oxidative stress.

Cells were gated for PI to show viability, dead yeast cells can show high fluorescence and so gating based on PI staining allowed the removal of this signal. The percentage of live cells that were gated for H2DCFDA staining were then counted to give a percentage of live cells which were considered high ROS. This allowed us to compare oxidative stress between the different samples.

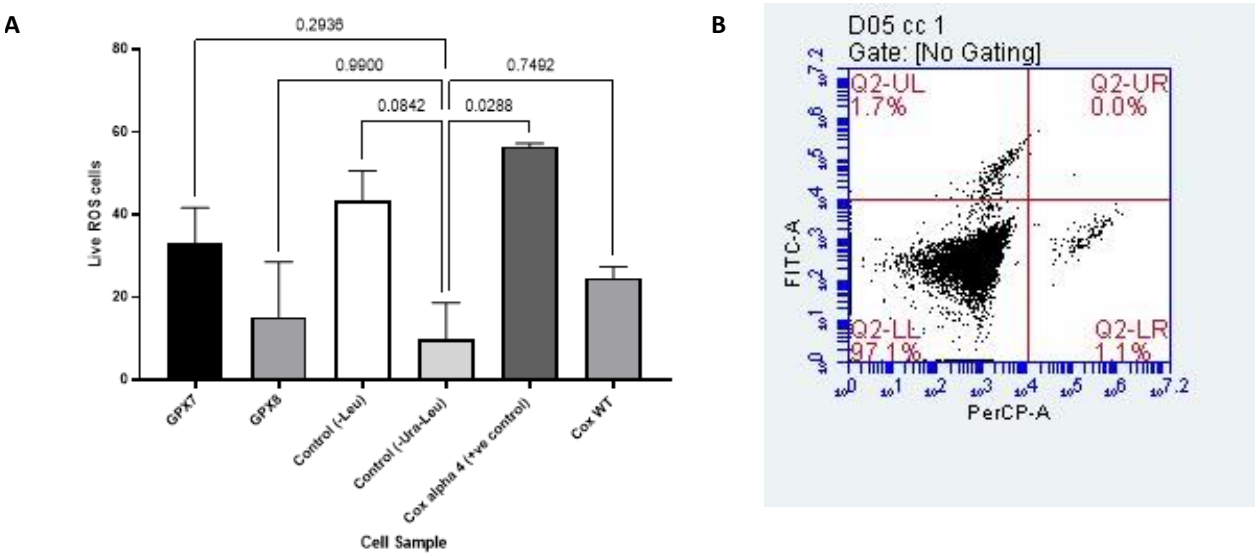

Figure 20: A:Flow cytometry data showing a potentially reduced ROS response in GPx cells. Cells were treated with PI and H2DCFDA and data shown is from cells showing high levels of H2DCFDA staining whilst still alive based on PI staining. One way anova and Tukeys test showed significant differences between control and cox alpha 4 but no significant differences elsewhere. P values indicated on graph. B: Raw data for control cells showing percentage of events for PI (X axis) and H2DCFDA (Y axis).

Flow cytometry appeared to show some reduction in levels of live cells under high reactive oxygen stress, although further experimentation would be needed as differences here were not statistically significant (N=3). Certainly the GPx samples are not under high reactive oxidative stress, however the wild type cells were also showing very low levels of oxidative stress and hence differences may be minimised.

The GPx produced should interact with Pdi1 causing a change in oxidative state. As the GPx was not increasing recombinant protein production, and as the reduction in oxidative stress was not as significant as hoped, it was decided to investigate

whether the GPx and Pdi1 were interacting as expected. To determine whether the GPx was interacting with the Pdi1 it was decided to firstly detect the location of the GPx proteins via immunofluorescence.

4.2.4 Immunofluorescence

To try to locate the presence of the GPx in the cell using immunofluorescence lyophilized cells were fixed onto a microscope slide and incubated with firstly anti HA antibodies and secondary antibodies conjugated to a TRITC molecule. Cells were expressing GFP coupled to a native Sec63 yeast protein resident in the ER membrane as part of the signal recognition particle pathway[272]. We hypothesized that the GPx signals should occur in the same place as the GFP to indicate localisation of the GPx.

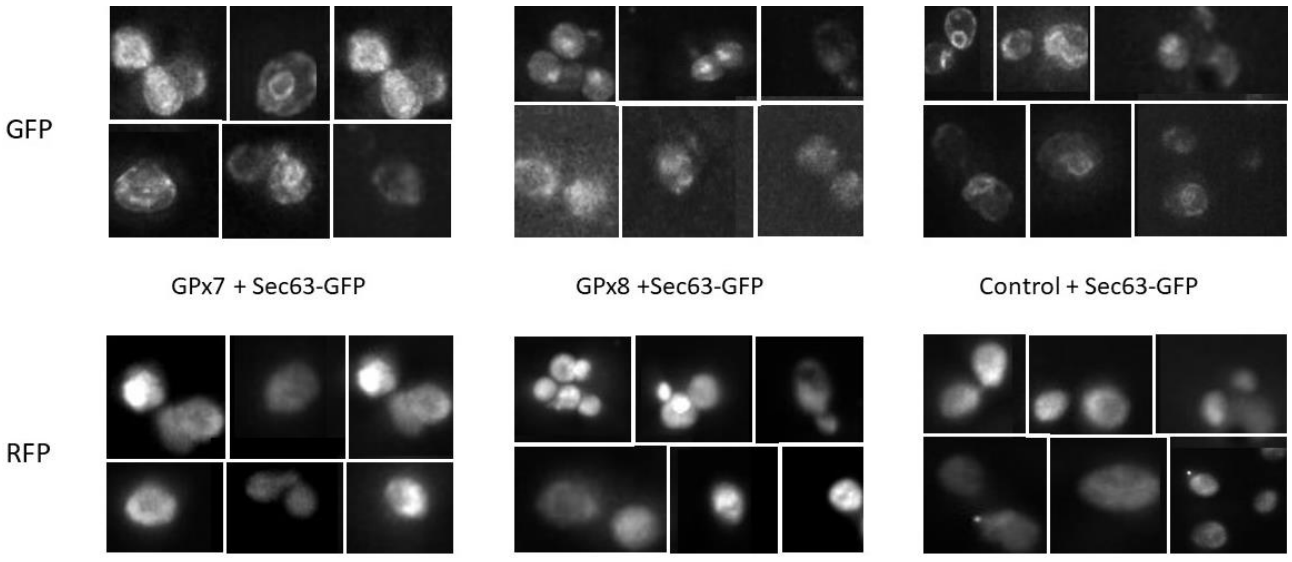

Figure 21: Example of cells showing GFP (top) and RFP (bottom) signals in GPX and control cells. Composite images.

The Sec63-GFP fusion protein showed clear ER-like distribution, visible as a ring-like shape around the cells' nucleus. In the presence of HA-tagged GPx8, visualisation of

TRITC-labelled anti-HA antibody also frequently revealed a concentration of signal around the cells' nuclei, indicating that the signal sequence and ER retention signal fused to GPx8 likely lead to ER-localisation as expected. The detected signal as well as the localisation was less clear for GPx7. We observed that GPx7 did not interact strongly with the anti-HA antibody when the protein was extracted under mild conditions (figure 22), although it was clearly detectable when extracted under denaturing conditions (figure 18). This may indicate that the HA-tag is obscured in the context of folded GPx7, which would explain the lower immunofluorescence signal observed for this protein. However, in GPx7-expressing cells displaying the strongest immunofluorescence signal, this signal tended to show an uneven cellular distribution which could be consistent with ER-localisation also for this protein. In summary, immunofluorescence analyses are consistent with ER localisation of the recombinantly expressed GPx proteins although they do not provide unambiguous support.

4.2.5 immunoprecipitation of GPx

To further analyse the interaction of PDI with the recombinant GPx proteins, it was decided to try and co-immunoprecipate them. To do this magnetic beads conjugated to an anti-HA antibody were used to remove GPx proteins from a cell lysate, and then the precipitate could be probed for presence of PDI and GPx. To enable this analysis, firstly the cells were lysed in Thermo Scientific Yeast Protein Extraction Reagent (Y-PER). After lysis of cells, the resulting protein samples could be separated via gel electrophoresis, and then detected via western blot. A western blot was conducted to show PDI and GPX proteins in the resulting cell lysis. This blot should show that both the GPX proteins and PDI are being removed from the cells in

lysis, and that the control cells do not contain anything which causes a cross reaction.

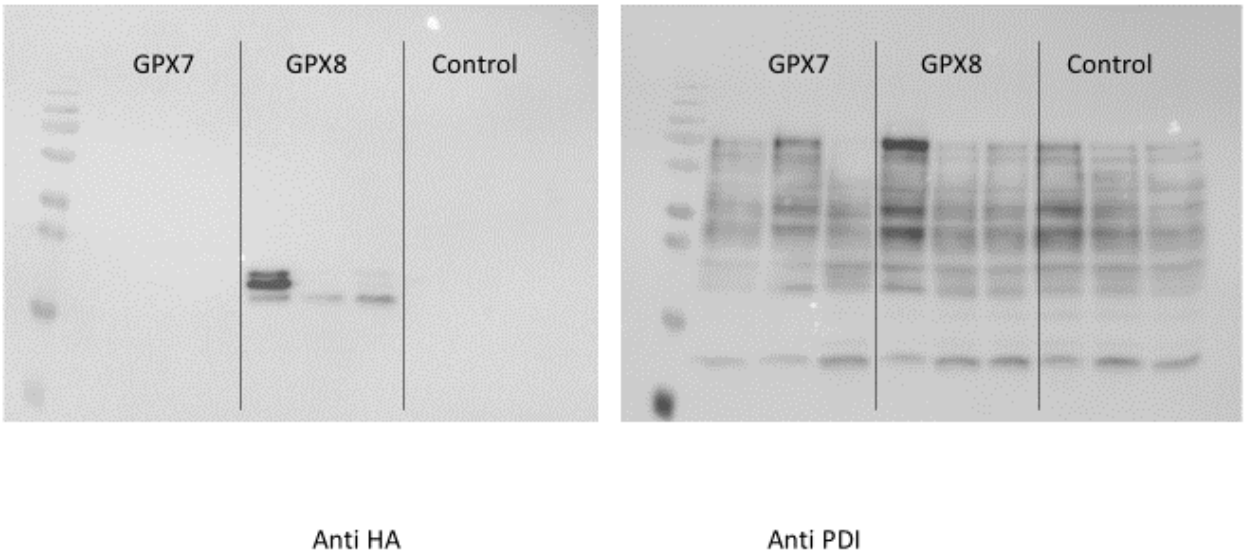

Figure 22: Cells lysed with Y-PER showed only GPx8 gave banding for AntiHA, indicating GPx7 was not removed from cells using this method. Loading control is not shown.

When cells were lysed using YPER reagent, only GPx8 could be detected using immunodetection on a western blot. All cells, including control, showed all cell lysis containing PDI, as would be expected, but the anti HA results indicate that GPx7 is not being removed from the cells in this method. GPx7 is being produced by the cells and can be detected by western blotting as seen previously in figure 18. but the change to a milder detergent lysis reagent appears to prevent either extraction or detection.

Samples could then be analysed after incubation with anti-HA bound magnetic beads. If the GPx and PDI are in complex it would be expected that both PDI and GPx would be present in the eluant from the beads. As the interaction of GPx and PDI is not constant, cells were crosslinked using paraformaldehyde before lysis in an

attempt to maximise PDI in complex with GPx at the time of immunoprecipitation. If the GPx and PDI can both be seen after purification it shows that the proteins are interacting in vivo.

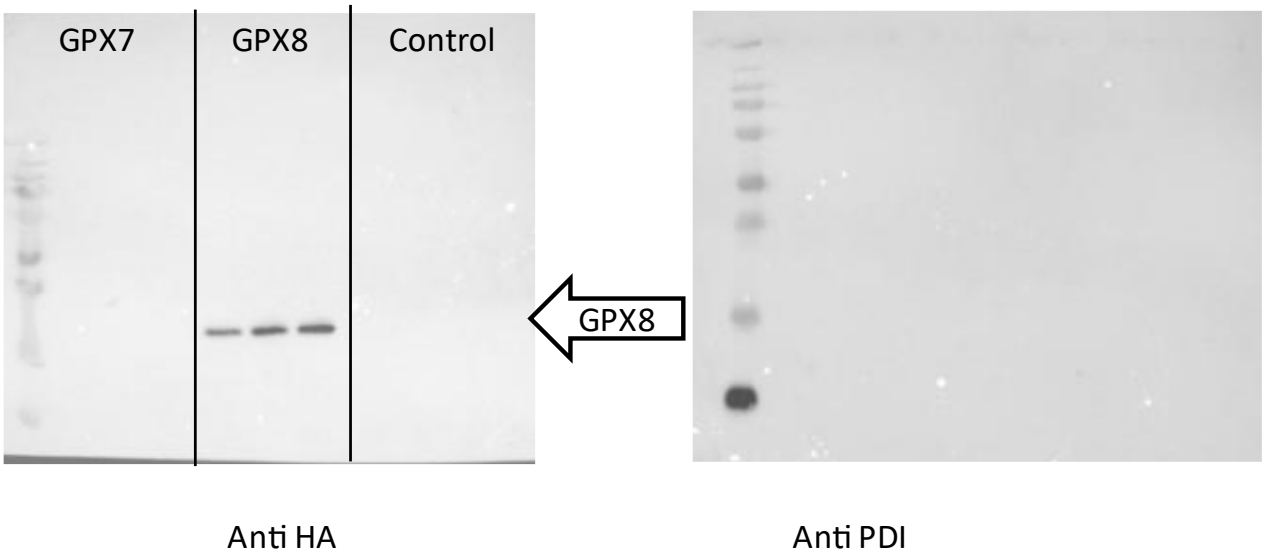

Figure 23: Cells were crosslinked and then lysed with YPER reagent. The magnetic beads were used to purify the anti-HA GPX proteins and then samples were tested for HA and PDI presence. Although GPX8 cells show some purification of GPX8, no presence of PDI could be detected. Loading control is not shown.

After cell lysates were incubated with the anti-HA magnetic beads, GPx8 could be seen however PDI could not be seen. This was unexpected as previous experiments show that the GPX proteins are reducing recombinant protein production and hence it is expected that they are functional in some sense within the cell. For the proteins to be functional within the cell it would be expected that they would interact with PDI as we know this is their method of action[210], however the expected interaction between GPX and PDI could not be seen via this method. This may be because the interactions were either not strong or long enough to remain present after washing the magnetic bead bound GPX, or could be that interactions are too fleeting to be

picked up via the assay. Although crosslinking should help to increase visibility of interaction, if PDI is already oxidised, the GPX would not bind.

4.2.6 oxidative state of PDI

To identify whether GPX proteins were having the expected effects upon the oxidation state of PDI it was decided to look at the oxidation state of PDI. To do this we incubated cell lysate with PEG5000 attached to maleimide. This can bind to free thiol groups in oxidised PDI and increases the molecular weight by ~5000 per molecule bound[202]. When these samples are separated via SDS-page they therefore appear at different positions on the resulting gel. When samples are transferred onto a western blot and detected with anti-PDI antibodies the band shifting can be used to identify levels of oxidised PDI in the samples.

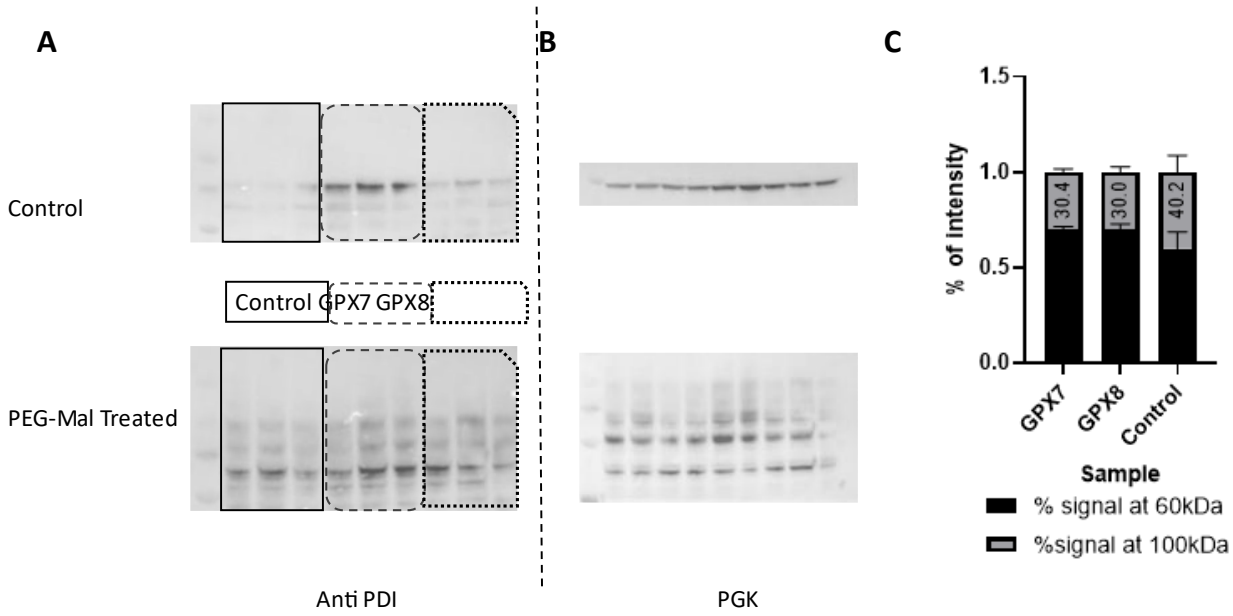

Figure 24: Oxidative state of PDI was analysed by treatment with PEG-mal and detection via western blot. A) Higher banding can be seen for samples which have bound 5K PEG molecules, which is not present in the control untreated cells. B)

showing PGK probing of the same samples to show equal loading between samples. C) Band intensity was analysed in ImageJ and band intensity between the 60kDa and 100kDa bands was compared. Percentage of signal observed at the 100kDa band is noted on graph. Error bars for plotted for SEM. N=3.

The PDI bound to the PEG-Mal can be seen as higher bands shown on the treated blot for anti-HA but are not present in the control blot, showing that they come from the binding of the PEG-Mal. These bands are present for all 3 cell types tested. Band analysis was conducted to identify at what level these changes occurred at using ImageJ to quantify relative band intensity. After analysis it appears a larger percentage of control cell sample is at the 100kDa band. This could indicate more free thiol groups within PDI in the control samples compared to those containing the recombinant GPX.

A further assay was attempted to identify interaction between Pdi1 and GPx using DHFR. DHFR confers a large resistance to methotrexate compared to wild type yeast and can be translated in individual subunits, which only interact when physically close[273]. This can be used by fusing one half of the DHFR complex to each protein of interest, and screening for methotrexate resistance. The assay was designed using a plasmid created to contain both GPx and Pdi1 with complementary halves of the DHFR compound. In theory as the GPx and Pdi1 interact, the DHFR complexes come together and confer the resistance to be measured via growth assay, however the assay was unsuccessful as growth was unpredictable. Fusing DHFR to native Pdi1 would improve this, however this is difficult as Pdi1 is also essential for cell function.

4.3 Discussion

Evidence shows that GPx7 and GPx8 seem to be produced by the yeast, this can be seen as the western blot identifies a HA tagged protein in the transformed cells which is absent in the controls (figure 18). The GPx proteins produced appear to be in the correct size range of roughly 187 amino acids and 22kDa[274] for GPx7 and 24kDa for GPx8[216]. It is unknown whether the produced GPx proteins are identical to those in humans as the yeast system will have different post translational modifications[275]. Post translational modifications could cause structural changes to the final protein produced. Any changes in the active site due to modifications, whether they be structural or changes in charges, could affect the function of the final protein. The effect of any post translational modifications on the structure of GPx was not studied in this project and therefore cannot be known. Analysis of final structure and active site using techniques which could include Xray crystallography, NMR and circular dichroism would potentially give further insight into the quality of the produced protein.

The effects of the GPx proteins upon *Gaussia* luciferase production (figure 4) were unexpected due to the hypothesis that they would improve PDI function. The expectation that more PDI would become oxidised and available for disulphide bond formation should have lead to an increase in recombinant protein expression and secretion in a similar way that over expression of PDI does17. As this pattern was not observed it led to one of the following possible conclusions: The GPx was being produced incorrectly, the PDI was not being oxidised by the GPx, or that the GPx was unable to interact with PDI in vitro.

To consider the effects of GPx upon the cell further, we delved further into the cell biology of these proteins. Knowing that GPx proteins have a huge role in regulating the redox state of mammalian cells[216,269,274,276], it was thought that we may be able to view the differences in oxidative stress using flow cytometry (figure 20). Originally the experiment was designed and conducted with cells grown in their regular media, however in these conditions none of the samples showed significant staining by H2DCFDA. To counter this, and knowing how GPX works with hydrogen peroxide, hydrogen peroxide was added into the media in the hope of observing the effects of the GPX. Signal variance of these samples was high, this made is difficult to clearly and accurately interpret results. This meant that although it appeared at first glance that GPX proteins were reducing the effects of oxidative stress upon the cell, it could not be shown definitively without further experimentation. The changes may have also been minimal or variable due to the lack of natural oxidative stress in the system. Considering cells which had a higher background level of oxidative stress may lead to better results. Although not significant, results are consistent with previous research on the unction and interactions of GPx proteins and therefore suggest that there may be some promise in using these proteins to enhance recombinant protein production

With the evidence that there could be a reduction of reactive oxygen stress upon the cells with the GPx, it was decided to look at how they were interacting with PDI. As the GPx was transformed with a well-documented secretion signal[270], we believed quite confidently that they would be sequestered within the ER, enabling interaction with PDI. Microscopy data (figure 21) compared the ER location in the cells using a Sec63-GFP fusion protein, to the location of the recombinant GPX detected via a TRITC conjugated anti-HA antibody.

Detection of protein within cells using fluorescent antibodies such as this is a well proven method[277] however does contain its limitations. Firstly in our example, digestion of the yeast cell wall and formation of spheroplasts does affect the visibility and morphology of the cell. This means that identifying the exact location of the ER is more difficult compared to using intact cells. This could be seen when looking at the control samples which had not been treated with the anti-HA antibody, as these samples were still difficult to identify strong ER structures in. This made it difficult to accurately identify the location of GFP and TRITC signals within the cell, and made it impossible to determine whether the GPx were ER resident. However, one useful piece of information from this was that it appeared cell ER morphology was not compromised upon addition of GPx. The second issue with immunofluorescence is that the use of large antibodies to detect small proteins can lead to steric hindrance[278] which can cause a masking effect on the signal. We did attempt to use an anti-GPx antibody instead of an anti-HA antibody (data not shown) but when this was tested via western blot, it appeared to have too much cross reactivity with other proteins in the cell lysate to be a viable option.

To improve knowledge of the GPx location, a fluorescent GPx protein could be made by fusing a fluorophore onto the GPx. This would allow much better visualisation of the GPx by microscopy and would hopefully allow definitive proof of location. The evidence we do have means GPx appears to be located in the ER, particularly evident in GPX8, however the control cells have similar signal in some place and so it is impossible to prove this definitively.

The location of the GPx proteins was further considered when we tried to look at co immunoprecipitation. The lysis of cells using the Y-PER reagent showed that the

GPx7 was not removed into the lysate. Y-PER is meant to be a method of yeast lysis that is less harsh, less reducing and simpler than standard lab methods, which is why we chose it originally for the magnetic bead work. We hoped the milder lysis would allow the GPx to be lysed from the cell, without being denatured as we needed it to retain the ability to bind to the magnetic beads and PDI. What became interesting is that GPx7 could not be seen in the lysate after Y-PER lysis. This indicates that it is not free in the cytosol and could add some credit to the view that the GPx proteins are being expressed and localised to the ER as intended.

The Co-immunoprecipitation failed to show evidence of Pdi1 binding to the GPx8 (figure 23). Early exploratory experiments showed some hints of Pdi1 on blots, however it was unclear whether this Pdi1 was purified by co-immunoprecipitation, or just failed to be washed off. There are a few reasons why the Pdi1 may not co-precipitate with GPx. Firstly the interactions of Pdi1 and GPx could be too quick and transient to enable to Pdi1 and GPx to be pulled down in complex. Secondly the interaction of the Pdi1 and GPx proteins could be too weak to withstand binding and washing of the magnetic beads. GPx7 especially has been shown to interact with Pdi1 quickly and, itself, can be oxidised within 15 seconds in hydrogen peroxide[279].

Finally to look more directly at the effects of the GPx proteins studied, we considered their effects on Pdi1 within the cell. The oxidative state of Pdi1 was viewed by binding to 5K PEG-mal and viewing size changes via western blot (Figure 24). When considering the percentage of the band signal at a higher molecular weight, control cells showed a larger percentage of signal, indicating more binding of maleimide compared to the GPx samples. This is really interesting as we would expect that the GPx would oxidise Pdi1 more and therefore see both GPx protein samples binding

more maleimide. Pdi1 can be present in semi oxidised forms with either of the sites being fully oxidised[202]. The limitation of this experiment was that it is unable to identify the exact location of the Pdi1 oxidation. Another limitation of this experiment is that thiol-disulfide exchange could occur during cell lysis or incubation with the PEG-maleimide. Experimental conditions were designed to attempt to reduce the likelihood of this, such as by lysing cells with glass beads in buffered conditions, however it is still possible that some state changes occurred.

If the Pdi1 is less oxidised in the presence of GPX, as suggested by the higher binding of PEG maleimide in the control sample, this could potentially go to explain why lower levels of protein production were seen in the GPX samples. Although we would expect the GPX to oxidise Pdi1 based on the mammalian system, we have no solid proof that they are capable of interacting with yeast Pdi1 in vitro. Further experimentation into how the GPx proteins interact with the yeast proteome would be needed to draw robust conclusions on their actions.

Chapter 5 comparison of strains and medias

5.1 Introduction of SGRP strains

In 2009 over 36 strains of bakers yeast were genomically sequenced with the hope of analysing their evolutionary relationships[55]. The overall variation seen in *S. cerevisiae* was relatively low compared to that in other populations of yeasts, such as *S. paradoxus* which is the closest relative of baker's yeast to live without human domestication. *S. paraxous* showed nearly 3X the number of SNPs (623287) and nearly 2X the number of Insertions and deletions called indels (25267) than *S. cerevisiae*. There are many reasons why large genome databases of strains can become useful. Analysis of differences between cells can help to indicate which differences are relevant, such as those that may lead to higher yields of recombinant proteins. Large scale studies genomically identifying yeast strains can indicate novel open reading frames and hence still could reveal new targets to improve recombinant protein production. Understanding the differences in these cell types, and previously unknown genes within them can, therefore, help to optimise protein production.

Improving the efficiency of cells for protein production is of great interest to the biotechnology industry. With the biology of many cellular processes remaining unclear, studying genetics and genome wide changes can help to understand the implications of protein secretion upon yeast strains[44]. One way in which genomic differences can be analysed is by RNA profiling. RNA sequencing has been used to study genome wide transcriptional responses to secretion in mutant yeast strains to study which cellular processes are changed in support of protein secretion[44]. This can lead to generalised observations, for example that altered energy metabolism can reduce respiration and increase fermentation [44]. The

balance of increasing amino acid biosynthesis and reducing thiamine biosynthesis has also been flagged as important in optimising recombinant protein producing strains[44].

As the use of engineered yeast cells to produce proteins can lead to the secretion of high quality protein-based pharmaceuticals at a fraction of the cost to mammalian cells both in production and downstream purification, yeasts have become common cell factories. One method to increase protein production and secretion is to engineer the yeast cells and reduce the effect of bottlenecks such as tailoring the endoplasmic reticulum[185] where BiP and PDI have been over expressed to reduce the unfolded protein response[186], overexpression of SEC16 to improve translocation from the endoplasmic reticulum[158] but also wider studies such as the use of microfluidics to identify multigenic pathways which increase yields of secreted proteins[187].

When high protein secreting strains of *S. cerevisiae* were examined by high throughput RNA sequencing evidence showed that conserved patterns such as energy metabolism and amino acid biosynthesis are present in high protein producing strains[44], and can therefore be used to create a list of requirements to gain yeast strains with efficient protein secretion. If this was fully optimised it may be that any strain could be modified with genetic deletions or upregulations to enable it to better produce and secrete recombinant protein.

When mutant strains were identified to be high producing using an alpha amylase reported genes, higher alpha amylase production was also associated with increased specific growth rate, increased glucose uptake rate, with reduced biomass on glucose but increased yield of ethanol on glucose[44]. The final biomass

of mutants was generally lower than the reference strain. The mutant strains were shown to have a higher amylase titre in the medium but lower intracellular expression levels compared to a reference strain showing the secretory pathways were working at a higher efficiency, and therefore the strains had a higher secretory capacity[44].

To enable accurate transcriptional profiling, cells were sampled in early exponential phase to reveal factors which may influence secretion[44]. The use of different evolutionary groups by principal component analysis indicated that the evolutionary paths leading to increased protein secretion could differ. The overall increases in secreted protein were thought to be associated with a global change in gene expression, rather than the large adjustment of few genes[44]. One of the strains selected for higher secretion and its descendants showed trisomy of chromosome three and many of the genes on this chromosome showed a roughly 2 times increase in transcription. Genes that were commonly expressed between high producing strains were identified and compared. Some of these genes such as those required for anaerobic growth were significantly upregulated. Downregulated genes such as those involved in phosphate responsiveness were linked to phosphate utilisation and regulation.

When transcription factors in cells producing higher levels of secretory protein were analysed, they showed upregulation of genes involved in hypoxia through the Rox1p pathway and the UP1p pathways[44]. Genes normally regulated by Hap1p which is involved in transcriptional activation and regulation of respiratory gene expression were downregulated in high producing mutant strains[44]. Gene regulation was therefore leading to a hypoxia like state. Genes involved in nutrient signalling,

nucleotide synthesis and phosphate metabolism were also upregulated[44]. TEC1 and MSS11 involved in the starvation response pathway were found to be upregulated in high producing cells[44], indicating that high expression of protein may lead to nutritional strain on the cells. Genes also downregulated include Bas1p[44] (which usually causes expression of nucleotide synthesis genes), and Mbp1p and Swi4p [44]which regulate the expression of genes during the early cell cycle (specifically G1/s). These genes were analysed for causation in high producing strains by deletion or overexpression to confirm the data[44]. More generally genes concerned with mitochondrial function, energy, respiration, amino acid metabolism were downregulated and genes related to ribosomes, translation within the cytoplasm, organelle organisation, golgi vesicle transport, protein lipidation and lipid metabolic processes were upregulated[44].

Carbohydrate metabolism was investigated in higher producing mutant strains, genes for high affinity hexose transport were generally found to be upregulated which may indicate abnormal glucose sensing within the strains[44]. This may occur via SNF3 which is normally involved in the regulation of such receptors and was found to be mutated in the high producing strains[44]. Many genes involved in the central carbon metabolism, including the TCA cycle, were found to be downregulated[44]. Many genes involved in amino acid biosynthesis were found to be downregulated in mutant high producing strains[44], however transaminases were found to be upregulated, indicating that conversion of amino acids, rather than synthesis, plays a vital role in the synthesis and secretion of protein[44]. The mutant strains had high requirements for cysteine, specifically, compared to the control strains. Findings were checked with 2 other proteins to prove this was not

a protein specific response[44]. Evidence suggested that higher thiamine synthesis was present in many higher secretory cell lines[44]

Reduced ER stress has regularly been indicated as favourite in the hunt for a high producing strain. As protein production can lead to oxidative stress in the ER, which can be improved by the use of antioxidants[207], it would be sensible to assume higher producing strains of cells have a method to reduce cellular stress and increase functionality of the ER. When high producing strains were analysed it was found that increased expression of HAC1 and ERO1 could be seen[44]. HAC1 is a transcription factor used when the unfolded protein response pathway is activates. As the UPR is used to reduce cellular stress in situations which cause heavy burden on the ER, it may enhance protein secretion when overexpressed by increase of KAR2 expression[204]. PDI1 and EMC1 were also upregulated in some of the increased secretory strains[44]; both genes involves in efficient protein folding within the ER. When Reactive oxidase levels were measured to evaluate oxidative stress, the high production strains showed higher ROS and therefore indicating ER stress was present. Activation of the UPR and enhancing protein folding can reduce this, and strains which were thought to be more efficient in these areas had lower ROS levels[44]. Due to the mutations within the strains creating a more favourable environment for protein production and secretion, the overall ROS/unit of protein was reduced in the optimum mutant strains compared to the wild type[44]. Overexpression of PDI1 has been shown to improve protein production[44,198,200]. Increase in NADPH supply can also reduce oxidative stress via the pentose phosphate pathway [208], and although no significant overproduction of NADPH was expressed, reduced biomass observed from higher producing cell lines may indicate a relocation of NADPH resources from biosynthesis to maintaining redox balances[44].

Due to the many factors implicating cells ability to produce recombinant protein, we wished to look at cells with different native ability to produce and secrete protein and consider what conditions could be used to optimise these. As many studies focus on the effect of a genetic manipulation or condition on one cell line, we wished to explore whether multiple cell lines reacted the same way to changes in growth conditions.

5.2 Results

5.2.1 Testing a wide selection of strains for luciferase production

Strain Number	Strain name	Genotype
181	NCYC3588	Y55 Mata Ura 3
182	NCYC3589	VWOPS83-787.3 Mata Ura 3
183	**NCYC3590**	**SK1 Mata Ura 3**
184	NCYC3591	BC187 Mata Ura 3
185	**NCYC3592**	**YJM978 Mata Ura 3**
186	NCYC3593	YJM981 Mata Ura 3
187	NCYC3594	YJM975 Mata Ura 3
188	NCYC3595	DBVPG1373 Mata Ura 3
189	NCYC3596	DBVPG1106 Mata Ura 3
190	NCYC3597	DPVBG6765 Mata Ura 3
191	NCYC3598	L-1374 Mata Ura 3
192	NCYC3599	L-1528 Mata Ura 3
193	NCYC3582	VWOPS87-292.1 Mata Ura 3
194	NCYC3583	W303 Mata Ura 3
195	NCYC3584	Y59 Mata Ura 3
196	NCYC3585	273619N Mata Ura 3
197	NCYC3586	Y11C17 Mata Ura 3
198	NCYC3587	3221395 Mata Ura 3
199	NCYC3600	DBVPG Mata Ura 3
200	NCYC3601	NCYC110 Mata Ura 3
201	NCYC3602	VWOPS03-567.4 Mata Ura 3
202	NCYC3603	VWOPS05-217.3 Mata Ura 3
203	NCYC3604	VWOPS05-227.2 Mata Ura 3
204	NCYC3605	Y12 Mata Ura 3
205	NCYC3606	YPS606 Mata Ura 3
206	NCYC3607	YPS127 Mata Ura 3
207	**NCYC3629**	**YPS128 Mata Ura 3**
208	NCYC3632	VWOPS05-227.2 Mata Ura 3

We wished to determine the differences in recombinant protein production across a variety of strains of *Saccharomyces cerevisiae*. To do this multiple strains selected from the SGRP collection[280] were transformed with a plasmid to express recombinant *Gaussia* luciferase with a pro alpha sequence secretory tag. Strains were then analysed via assay for luciferase levels both remaining within the cells and secreted into the medium. Initial screens for *Gaussia* luciferase production were carried out on multiple strains, with the aim to select representative strains of high , medium and low levels of luciferase production for further comparative studies. For better comparison of strains the highest observed intracellular and extracellular luciferase signals were set to a value of 100 and other samples normalised based on this value.

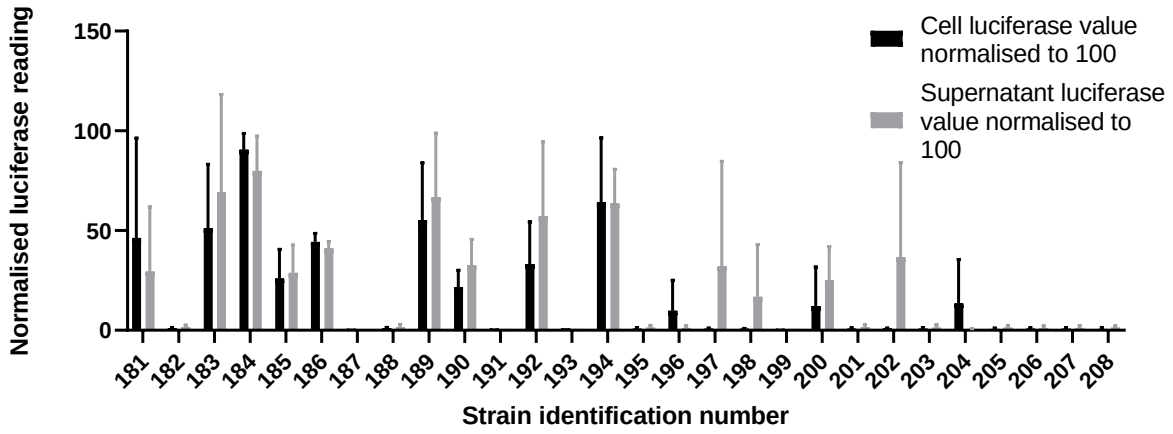

Figure 25: Variety of strains tested for Gaussia luciferase production in the cells or supernatant. Error bars show mean (N=3) +SEM.

Figure 25 shows that a variety of cells were successfully transformed to produce and secrete *Gaussia* luciferase. Luciferase expression levels varied widely between strains, indicating that genome variances between them may be impacting recombinant protein expression and secretion. Three cell lines were chosen that appeared to be typical examples of “high”, “medium” and “low” producing strains.

These strains were labelled in figure 25 as 183, 185 and 207 and represent SK1, YJM978, and YPS128.

5.2.2 choosing indicative strains

It was then decided to consider these cells and their luciferase production levels in further detail. The large selection of strains in figure 25 did not allow for conducting assays with high repeat numbers needed for in-depth statistical analyses. Firstly the cells were compared for *Gaussia* luciferase production as before in with more biological replicates to identify more accurately the difference between the cell types and to enable statistical analyses. The three cell lines were grown overnight in selective media and then assayed for *Gaussia* luciferase production within cells and secretion to further compare these differences.

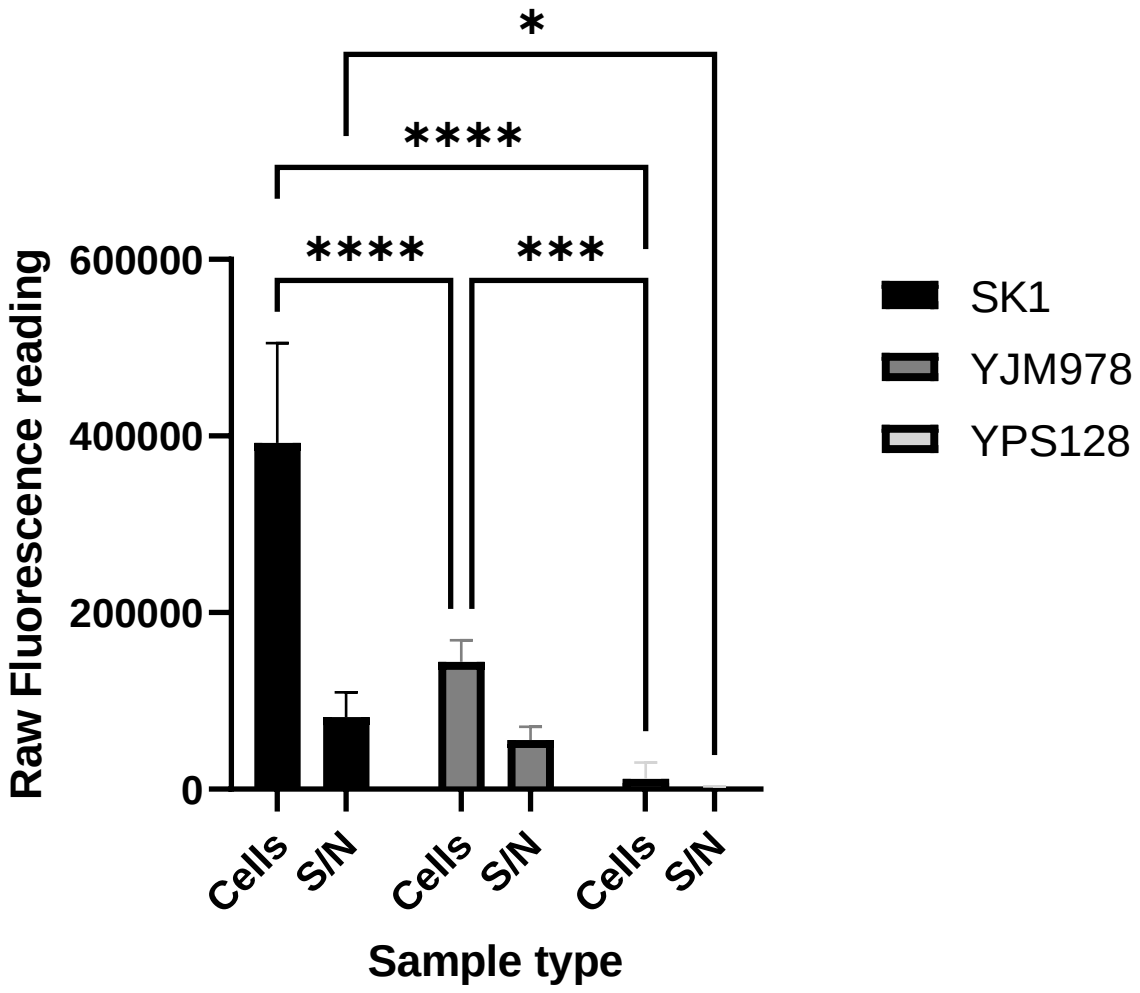

*Figure 26: Gaussia Luciferase production measured within the cell pellet and within the supernatant for 3 different strains.2 way anova plus Tukey test shows significant differences between the strains (n<9). Ns:P > 0.05, *:P ≤ 0.05, **:P ≤ 0.01, ***:P ≤ 0.001, ****:P ≤ 0.0001*

5.2.3 exploration of PDI and ERO1 levels

The further analysis of luciferase production and secretion followed the pattern seen in the larger screening studies and showed significant differences between the strains. Compared to the high producing SK1 the YJM978 and YPS128 show lower luciferase readings in the cells (P= <0.0001) and in the supernatant with a significant

difference seen between SK1 and YPS128 (P=0.0208) With the large variability in strains of the same species, it was hypothesised that expression levels of vital proteins within the cells may vary to cause these changes. Previous research has shown that Ero1 and PDI levels can be implicated in protein production levels[199], It was decided to test whether the cell lines varied in these proteins naturally, as this could begin to explain some of the differences between strains. Western blots were conducted on cell lysates to compare natural levels of PDI and Ero1 within the different strains.

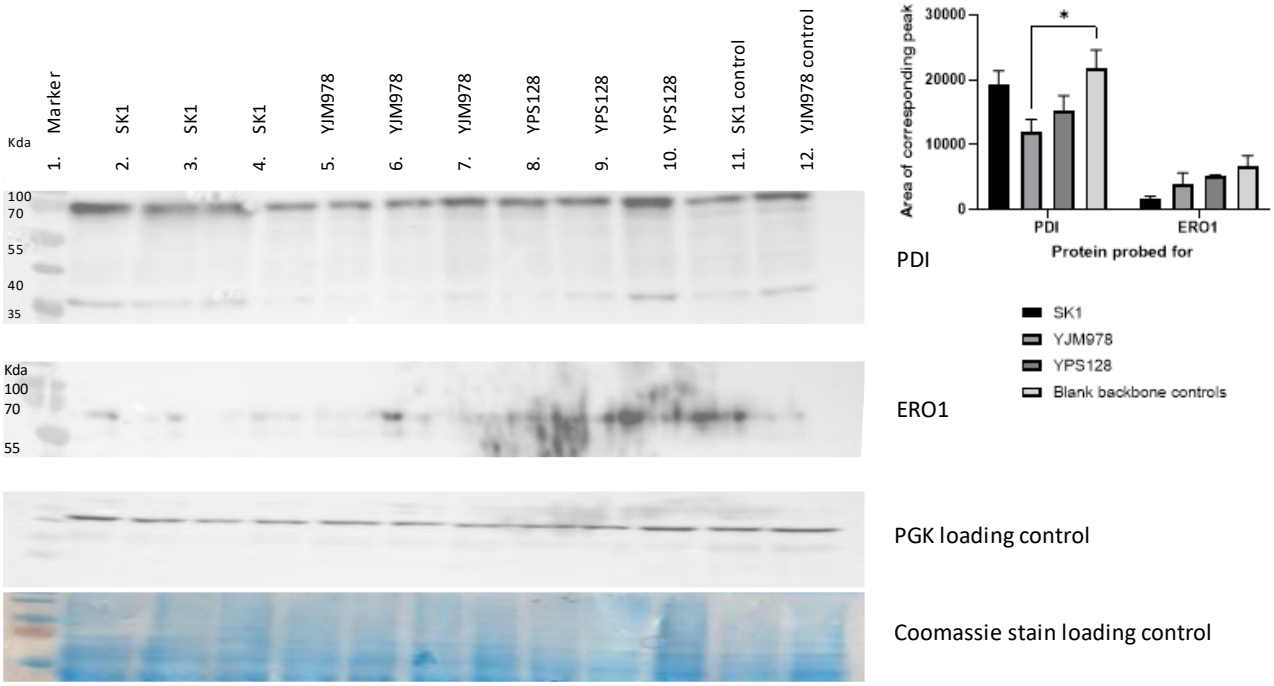

Figure 27: Western blot images with cell lysates ran and stained with antibodies for PDI (top), ERO1 (middle top), PGK (middle bottom) and the SDS gel stained with coomassie as a loading control. Cell lysates of each cell type post transformation with Gaussia luciferase plasmid were ran in biological triplicate with 2 samples containing an empty vector control. Analysis of bands was performed and average peak area plotted. 2 way anova was used to compare between samples.

In the western blot in figure 27 we found no significant difference in the expression levels of either PDI or Ero1 between the three *Gaussia* expressing strains, so it was decided that it was unlikely differences in these proteins were accounting for the differences in *Gaussia* levels previously observed. The bands seen for the Ero1 blot appear to smear which is a frequently observed phenomenon. Ero1 may cause these smeared bands due to its GPI anchor which may interfere with separation on an SDS-gel. Because of the way these bands appear, it made it difficult to draw solid conclusions about the Ero1 levels within the cell types, and band analysis becomes inaccurate.

5.2.4 Luciferase production over time

To explore the temporal dynamics of luciferase expression and ensure that we measured levels at a point in time when differences between strains would be observable A western blot was then conducted to see how luciferase was secreted over time. To do this cells were grown shaking in minimal media, at appropriate time points a cell sample was taken and measured via optical density.

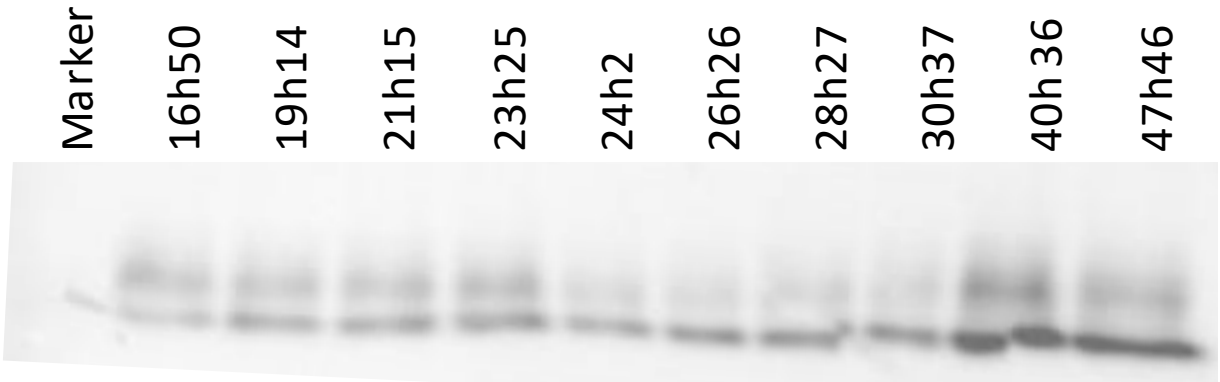

Figure 28: western blot showing secretion of Gaussia luciferase at different time points in SK1. Loading control is absent.

The western blot (figure 28) showing secretion of *Gaussia* luciferase over time shows linear expression levels during the times measured. This implies that cells continue to secrete *Gaussia* luciferase while growing in stationary phase as cells would have reached the stationary growth phase over these time points.

5.2.5 expression in different media

Cell assays had to this point been conducted purely in minimal (sc-ura) medium. This could be reducing the amount of protein the cell is able to create and secrete. It was hypothesized that in richer medium the cells would be able to synthesize and secrete more protein as they would have more available nutrients to meet all of the cells needs and improve cell growth. This was tested by growing the cells up in YPD medium and comparing production and secretion of *Gaussia* luciferase. This was tested for the three example cell lines to determine whether differences between strains remained constant in the different media types.

YPD medium removes the selective pressure upon the cells to keep the plasmid DNA. Growing cells in YPD that express a recombinant protein via plasmid DNA can therefore lead to expulsion of foreign DNA and reduction or cessation of recombinant protein production. To mitigate this issue it was decided to integrate the luciferase DNA into the host genome. This was tried in different ways, from originally a gateway cloning method to create a plasmid capable of inserting the DNA into chromosome 4, followed on by a Gibson assembly method for the same purpose. Both methods failed to produce cells which continued to express the luciferase after removal of the plasmid. Following this it was decided to integrate the DNA. DNA integration was attempted by creation of an integrative plasmid using gateway cloning, creation of an integrative plasmid using Gibson assembly, and insertion at the *SUC2* locus using

CRISPR. Following CRISPR and confirmation of luciferase signal, the CRISPR plasmid required removal. Yeast expressing URA3 can convert 5-FOA to a compound which is toxic to cells and hence growth of cells on FOA containing plates enables selection of cells which have lost the plasmid containing URA3. Although initial luciferase expression results with this looked promising, growth of the cells on FOA plates again saw loss of *Gaussia* luciferase signal. Although the use of plasmid based expression of DNA without selective pressure is non-ideal, it was decided to continue the testing of the strains in this method. In order to minimise plasmid loss, cells were maintained under selective conditions right up the final culture which was used to measure luciferase expression levels, and cells were then grown for 4-6 hours in rich medium before conducting the activity assays. The advantage of this method is that direct comparison between the minimal media and YPD can be seen without any changes in promoter or expression level, however you have to allow for the YPD grown cells to have lost some expression DNA.

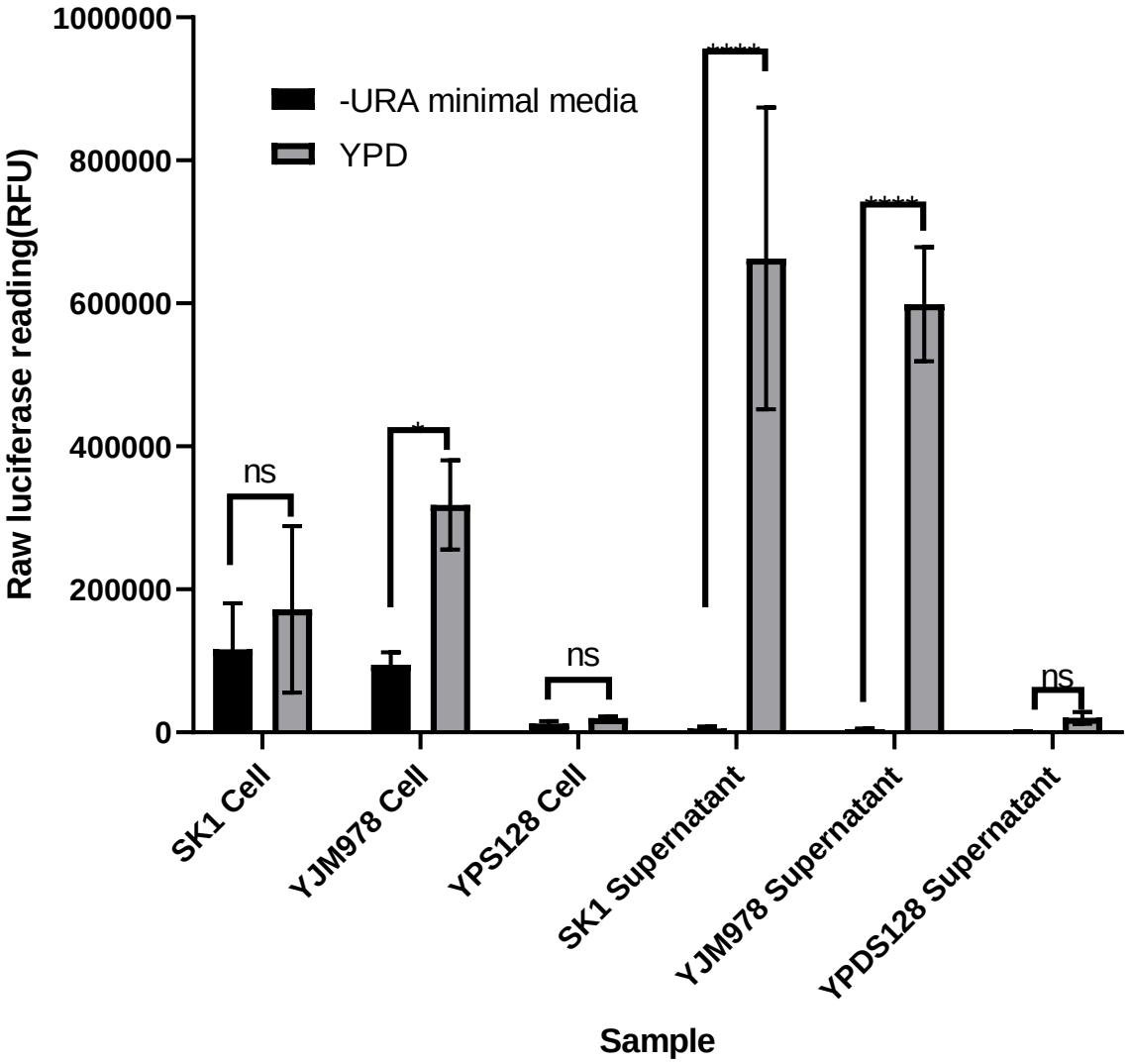

Figure 29: comparison of Gaussia luciferase production and secretion between cell types and media types. Error bars show mean+SEM. N=3. 2 way anova + Tukeys test for multiple comparisons.

YPD grown cells had a significant increase in levels of protein production and secretion in one out of three intracellular and two out of three supernatant samples.. To begin exploring reasons for the higher level of secretion in rich medium, it was then decided to test whether amino acid concentration would have large effects on recombinant protein production.

5.2.6 the effect of changing amino acid concentration

As amino acids are used in building of proteins, it was thought that increased availability of amino acids could lead to an increased production of proteins. To test this amino acids were added into minimal media at 2, 4, 6, 8 and 9 fold levels compared to the usual recipe according to manufacturers instructions. This allowed us to see whether amino acids were acting as a limiting factor where cells were grown in minimal media.

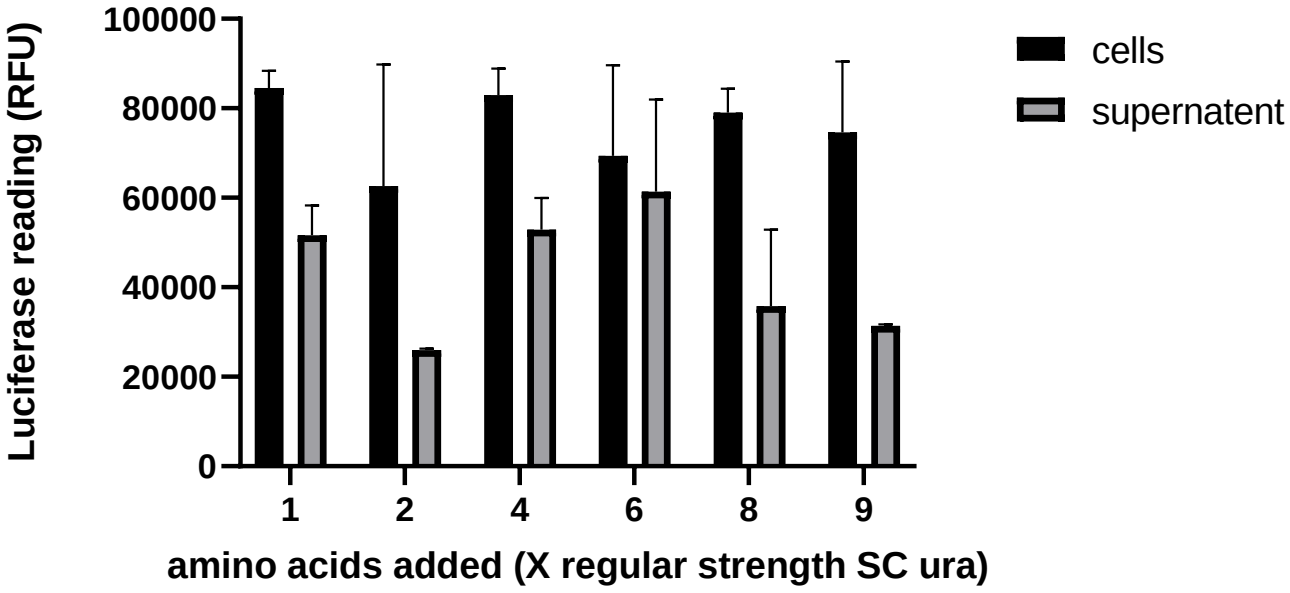

Figure 30: showing the effects of adding increasing concentrations of amino acids upon recombinant protein production. 2 way anova plus Tukeys test showed no significance.

Growth of cells in minimal media with additional amino acids showed no significant changes in either intracellular or secreted *Gaussia* luciferase. So far with the data it was impossible to show what elements of minimal medias were causing the reduced level of recombinant protein production. in addition to amino acids, SC medium contains ammonium sulphate as a nitrogen source and glucose as a carbon source,

and changes in the supply of these nutrients could also underpin the better performance of protein expressing strains in rich medium. Changing 3 different variables. Changing 3 different variables to find an optimum is a non-ideal method to conduct these experiments. This is because the variables may interact with each other, for example in media increasing glucose concentration to an optimum, may change the optimum for peptone concentration. To allow for the systematic and efficient study of the relationship between all the variables (factors) and outputs (responses) Design of Experiments can be used[281].

5.2.7 design of experiments

To be able to consider this in more detail it was decided to study the ideal media composition for recombinant protein production using the design of experiments approach. As we wished to specifically look at secretory protein it was decided to consider 2 different outcomes, firstly the total amount of protein secreted, and secondly the percentage of produced protein which was secreted. This allowed consideration of both what was best for bulk production of protein, as well as what conditions may optimise the yeast secretory pathway.

Using a full factorial design with 3 factors and 1 centre point per block was used to look at how varying the three components in minimal media, yeast nitrogen base (YNB), amino acids – ura (AA) and glucose, could impact upon recombinant protein production. A full factorial design was chosen to ensure we could see any interactions between these components. We chose to vary the concentrations between 0.5X standard up to 5X regular concentration compared to the standard YPD (1% yeast extract, 2% peptone and 2% glucose) and SC media recipes [0.67% Yeast Nitrogen Base without Amino Acids, 2% glucose, amino acids varied as shown

in materials and methods). Below these values it was thought that cell growth would be compromised too much and above these values the media became too viscous.

Given the nutrient requirements of cells, it was expected that there would be some interaction between media components, and so it was decided to choose a full composite design which would allow interactions between factors to be seen. The understanding of both which factors are most important for optimal protein production and how the factors interact would be necessary to design the best possible media conditions.

After inputting the design into Minitab software, the experimental conditions and running order was output. After running the experiments it could then show which factors had the most importance and suggest a best component media to improve luciferase production. The table in figure 31 shows the conditions tested in the experimental design used to optimise minimal media. 2 output parameters were tested for results. The main effects and the interactions can then be seen plotted to indicate how the components are affecting luciferase secretion.

Yeast Nitrogen Base (YNB)	Amino Acids (aa)	**Glucose**	**Luciferase in Supernatant/Luciferase in cell**	**Secreted luciferase normalised**
0.5	5	0.5	0.1	12.0
2.75	2.75	2.75	0.1	10.7
0.5	0.5	5	0.4	1.7
5	5	5	0.7	100.0
5	0.5	0.5	0.1	6.6
5	5	0.5	0.1	15.0
0.5	5	5	0.0	8.6
5	0.5	5	0.2	3.9
2.75	2.75	2.75	0.1	7.5
0.5	0.5	0.5	0.4	1.9
1.412142	1.412142	4.087858	0.2	8.3
2.75	0.5	2.75	0.1	76.9
1.412142	4.087858	4.087858	0.2	11.6
0.5	2.75	2.75	0.1	67.9

2.75	5	2.75	0.4	62.8
2.75	2.75	0.5	0.3	100.0
1.412142	4.087858	1.412142	0.1	23.5
2.75	2.75	2.75	0.1	11.2
4.087858	1.412142	1.412142	0.1	8.8
4.087858	4.087858	1.412142	0.1	10.3
2.75	2.75	2.75	0.3	2.0
2.75	2.75	2.75	0.2	38.9
4.087858	1.412142	4.087858	0.1	17.5
4.087858	4.087858	4.087858	0.6	0.8
1.412142	1.412142	1.412142	1.7	21.4
2.75	2.75	5	0.3	1.7
2.75	2.75	2.75	0.1	8.4
2.75	2.75	2.75	0.1	2.8
2.75	2.75	2.75	0.2	32.4
5	2.75	2.75	0.1	56.5

Figure 31 : Design of experiments tabular design and raw input results to optimise minimal media for secretion of recombinant protein.

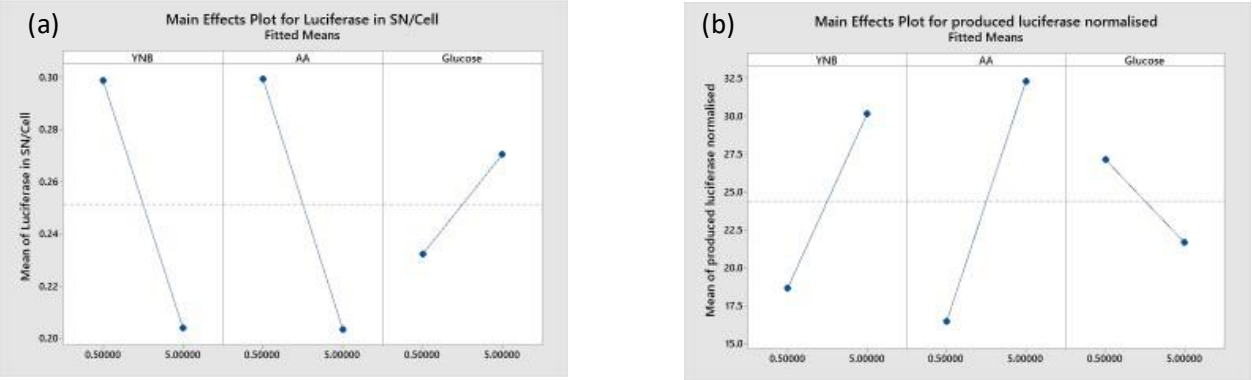

Figure 32: Main effects plot showing the impact of the 3 minimal media components on recombinant protein production based of design of experiment software predictions. a), Main effect Plot for the proportion of secreted luciferase, b) main effect plot for total secreted luciferase levels

The main effects plot (figure 31a) shows lower levels of YNB and AA to increase luciferase secreted whilst higher levels of glucose are needed to increase the ratios of luciferase secreted, however to increase overall production higher levels of yeast nitrogen base and lower glucose in required (figure 31b). This means that the optimal media for secreted luciferase in the conditions tested is 0.5X concentration SC-Ura, but optimal media for overall production of luciferase is 5X concentrated SC-Ura compared to standard recipes. The interactions chart shows strong interactions between all of the components. To consider which components were the most important overall for luciferase production, the Pareto chart was analysed. The Pareto chart shows the factors which have the largest effects.

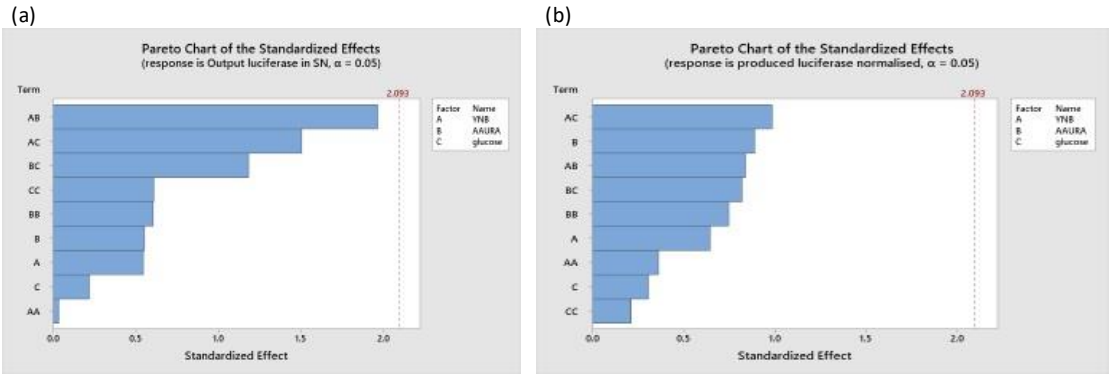

Figure 33: Pareto chart showing the major effects of media components on a) ratio of luciferase which is secreted and b) overall luciferase secretion.

The Pareto chart showed that the largest effects were actually due to interactions of the different components, rather than an individual component acting alone. Which factor has the largest effect depends on whether the percentage of secreted luciferase (figure 33a) or the raw amount of intracellular luciferase produced (figure 33b) were analysed to optimise the percentage of produced luciferase which is secreted (figure 33a) or the raw amount of luciferase produced (fig 33b).

Even with optimising the minimal media, the levels of luciferase produced appeared to be significantly lower than the levels seen in YPD. This indicated that it was an additional component of YPD that was enabling cells grown in YPD to produce higher protein levels. It was considered that the peptone in YPD could be causing the largest increase in recombinant protein levels, as previously published data have shown that peptone can change productivity in CHO cells[282]. To test this it was decided to compare minimal media with added peptone to YPD for *Gaussia* luciferase levels. Peptone was added to a final concentration 2% matching that of the YPD medium to enable comparison.

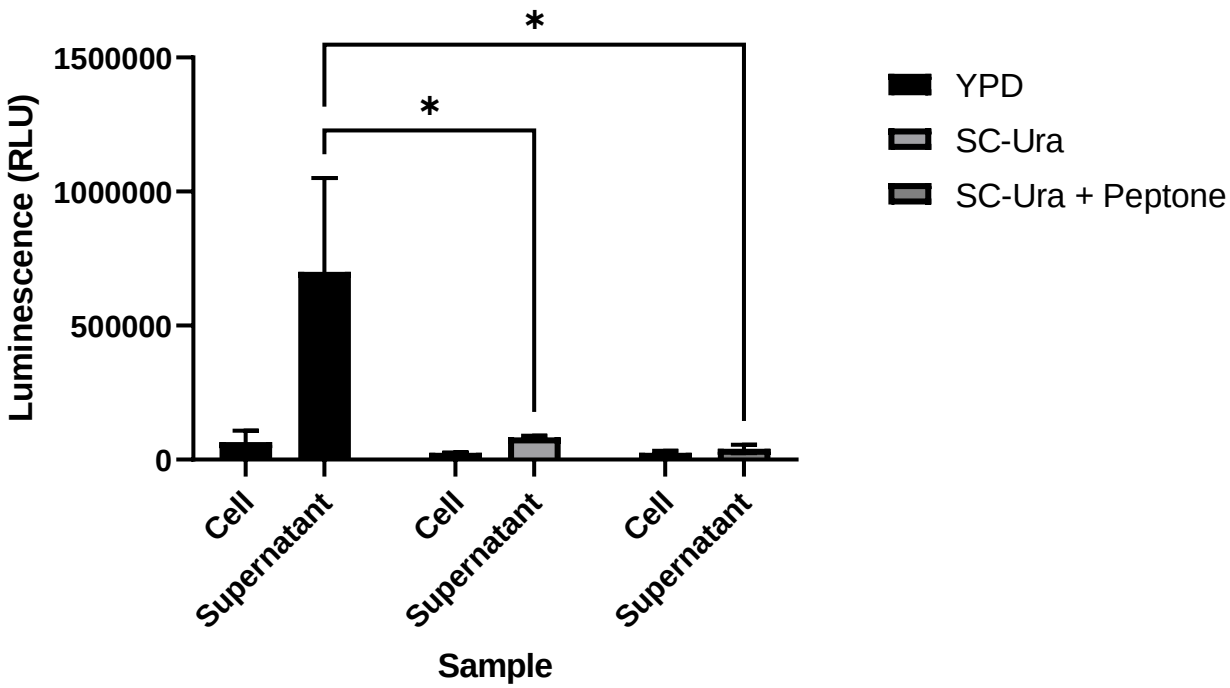

Figure 34: assessing the impact of adding peptones to minimal media upon protein secretion. 2 way anova plus Tukeys multiple comparisons test.

When comparing minimal media to YPD we showed a significant reduction in *Gaussia* luciferase secretion (p= 0.0264), even when peptone levels were matched to the original YPD concentration (P= 0.0181). This showed that peptone alone is not enough to increase recombinant protein to the levels seen in complete media.

To further investigate which properties of YPD were impacting protein secretion, design of experiments was used to find the optimum parameters for *Gaussia* luciferase production and secretion in the high producing cells when grown in YPD. As before, we looked at two possible outcomes leading to the highest overall secreted luciferase as well as amount of total luciferase secreted. The three recipe components of YPD (yeast extract, peptone and glucose) were varied between 0.5X and 2.5X regular concentration (yeast extract between 0.5% and 2.5%, Peptone and glucose between 1% and 4.5%).

StdOrder	RunOrder	PtType	Blocks	yeast extract	peptone	Glucose	Luciferase in Supernatant/Luciferase in cell	Normalised secreted luciferase
3	1	1	1	0.9	2.1	0.9	2.1	61.6
16	2	0	1	1.5	1.5	1.5	3.0	83.8
1	3	1	1	0.9	0.9	0.9	2.2	59.6
2	4	1	1	2.1	0.9	0.9	2.8	79.3
14	5	-1	1	1.5	1.5	2.5	2.5	67.0
18	6	0	1	1.5	1.5	1.5	2.4	67.0
20	7	0	1	1.5	1.5	1.5	2.7	67.3
6	8	1	1	2.1	0.9	2.1	2.6	72.0
9	9	-1	1	0.5	1.5	1.5	1.9	51.8
17	10	0	1	1.5	1.5	1.5	2.7	74.0
12	11	-1	1	1.5	2.5	1.5	2.6	76.1
15	12	0	1	1.5	1.5	1.5	2.7	74.0

19	13	0	1	1.5	1.5	1.5	2.5	65.8
13	14	-1	1	1.5	1.5	0.5	2.6	65.3
8	15	1	1	2.1	2.1	2.1	3.0	87.2
4	16	1	1	2.1	2.1	0.9	3.3	97.4
10	17	-1	1	2.5	1.5	1.5	3.0	100.0
7	18	1	1	0.9	2.1	2.1	2.5	78.5
5	19	1	1	0.9	0.9	2.1	2.1	62.2
11	20	-1	1	1.5	0.5	1.5	2.6	74.4

Figure 35: Design of experiments tabular design and raw input results to optimise YPD media for secretion of recombinant protein. Run order data omitted.

Data were input into minitab software as in table (figure 35). This was then used to analyse the main effects of each components. The main effects plot was then used to view how each factor impacted upon luciferase secretion.

(a) (b)

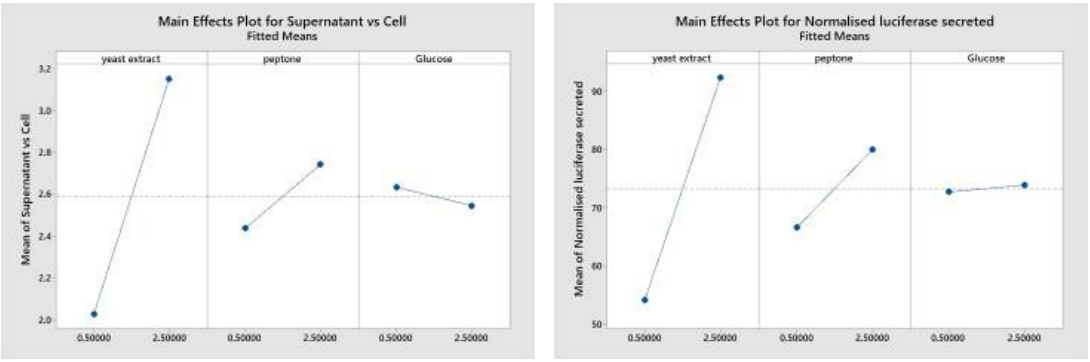

Figure 36: Main effects plot to show the impact of the 3 YPD media components on a) ratio of luciferase which is secreted and b) overall luciferase secretion.

The main effects charts show with both supernatant/cell (fig 36a) and overall secretion (fig 36b) into the supernatant there is a positive correlation with yeast extract and peptone, but the effect of glucose appears moderate and changes in direction between the two parameters. To further clarify which factor was the most important in yeast secretion, the Pareto chart was analysed.

(a) (b)

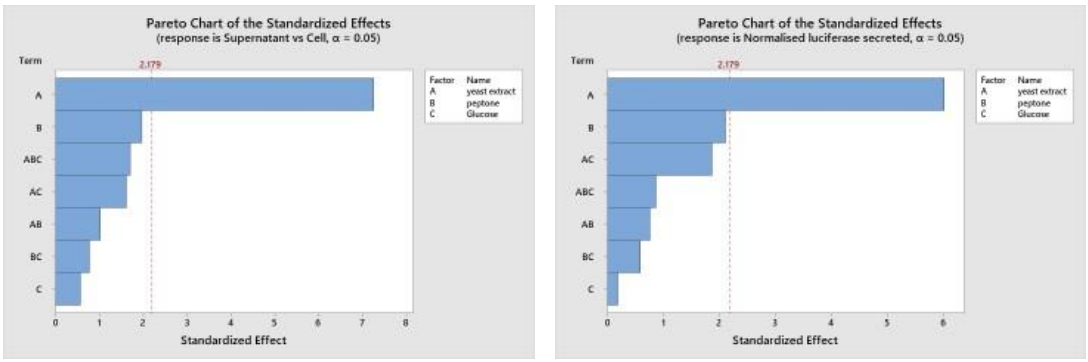

Figure 37: Pareto chart show the most impactful of the 3 YPD media components on a) ratio of produced luciferase which is secreted and b) overall amount of secreted luciferase.

The Pareto chart (figure 37) showed that the yeast extract was the most important effect for recombinant protein secretion. Software was then used to predict the optimum conditions within the design space. This makes the suggested media recipe

for optimised relative secretion 5% peptone, 2.34% Yeast extract and 0.6% glucose and optimised media recipe for overall luciferase production 5% peptone, 2.5% yeast extract and 1.73% glucose.

5.2.8 testing the optimal

It was thought that optimised conditions should show higher secretion compared to original YPD. As a control YPD at half normal concentration (0.5X YPD) and 5 times regular concentration (5X YPD) were also considered to show the optimisation of individua media components performed better than changing the concentration of all ingredients simultaneously. A luciferase assay was conducted to analyse these differences. It was also hypothesised that the responses to the optimised media may vary between cell strains and hence the three example high, medium and low strains were all tested in the optimised and control medias to see whether effects of optimisation were the same for all cell lines.

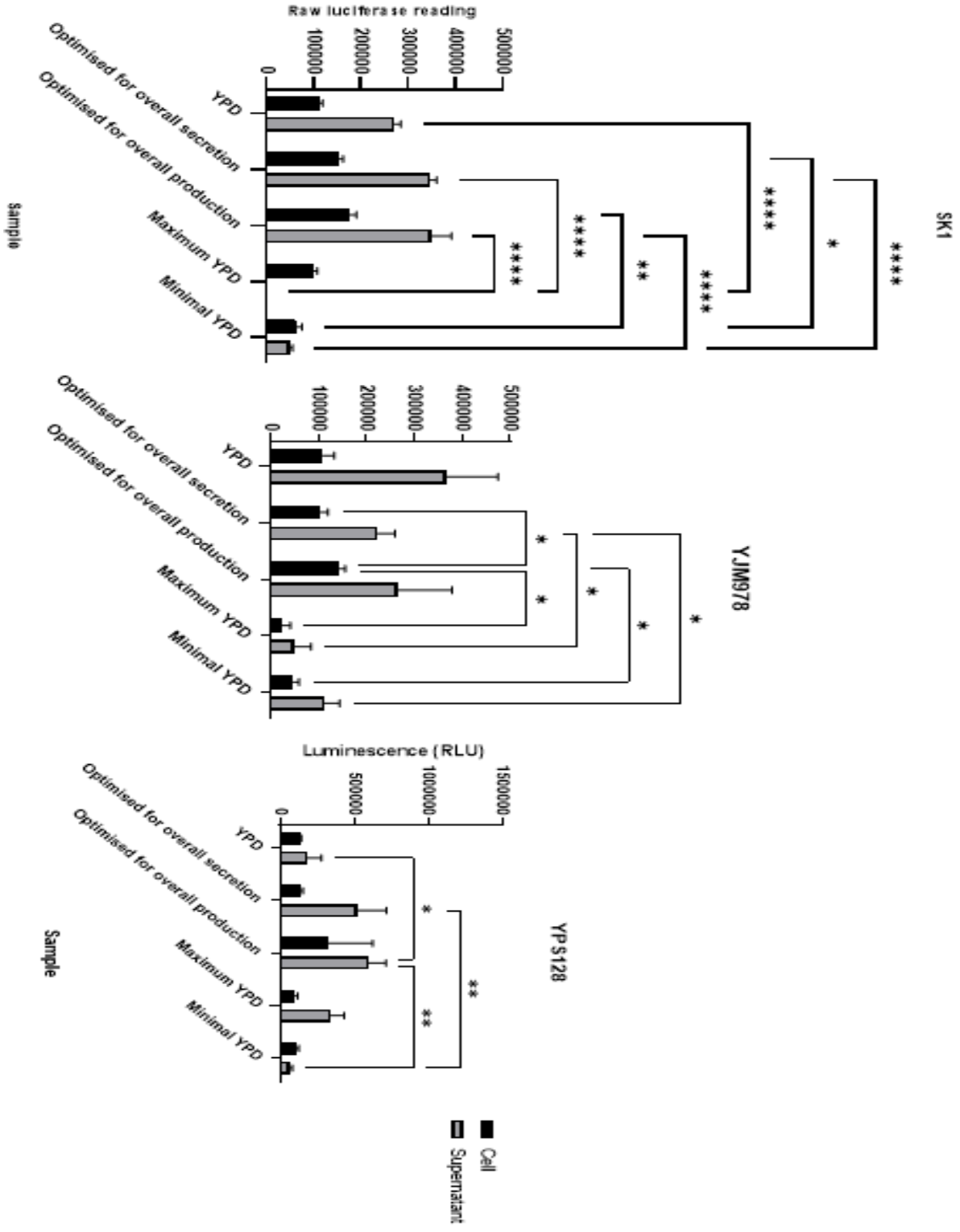

Figure 38: predicted conditions for optimum protein production or secretion in highly producing cells using medias predicted by design of experiments software. 2 way anova plus ad hoc Tukeys test to show statistical significance. N=3

Analysis of the different media showed that higher secretion was seen in the optimised media compared to original YPD in high producing SK1 cells (P= 0.0347), however none of the conditions significantly increase recombinant protein production compared to YPD in the medium YJM978 cells. The low producing YPS128 cells had the strongest increase in secretion in the optimised media for total amount of luciferase (P=0.0214). The YPS128 cells also showed reduction in expression levels in the concentrated YPD compared to the other strains, maybe indicating that further optimisation of these cells could lead to better protein yields. The work using design of experiment software suggests, therefore, that the best conditions for recombinant protein production, are cell line specific.

The media designed by Design of Experiments approach showed generally increased protein production and secretion compared to regular YPD and also compared to maximum ypd (defined here as YPD at 2.5X the regular concentration of components) or minimum YPD (defined here as YPD at 0.5X component concentration).

5.3 Discussion

There is a clear difference in the production and secretion of recombinant proteins between different strains of yeast (figure 25). The differences in production by strain can be 100 fold and this highlights the importance of selecting a good strain at the beginning of a project to produce the most protein. The SK1 strain chosen as a high producer and secretor is a west African strain, with the YJM978 being a European wine strain and the YPS128 being a north American strain[280]. At this time it is not known whether patterns of high protein production are random or whether closely related strains from the same areas show similar profiles. As secretion was still

increasing at 48 hours it is unknown how long levels of secreted protein continue to increase before levelling off and then reducing when degradation occurs.

YPD is clearly a better medium for protein secretion compared to SC-ura, with all strains showing some improvement and SK1 and YJM978 showing hugely increased secretion (figure 29). The increased secretion is thought to be due to extra nutrient availability leading to better growth and healthier cells. It has been shown that culture medium can affect mRNA expression levels in mouse embryo cells[283], but general changes in mRNA expressions due to media changes are not known. Analysing both system wide changes and changes in expression of the mRNA for the recombinant protein would show whether the changes were due to transcriptional or translational differences within the media.

Knowing that the growth in YPD increased secretion greatly, it was interesting that addition of extra amino acids alone (figure 30) did not appear to improve recombinant protein production and secretion. As amino acids are the building blocks for the protein, it was thought that having these in higher excess should increase cellular ability to produce proteins. Recombinant protein production requires extra protein synthesis, and so increasing availability of amino acids could have increased recombinant expression. As the results seen in figure 30 show extra amino acids do not increase recombinant protein production, this indicates that amino acid availability in minimal media is not the limiting factor in protein production, this is consistent with modelling studies[284].

Figure 34 showed that the addition of peptone alone was not enough to increase the protein secreting ability of cells grown in minimal media, and minimal media with additional peptone was not significantly better at allowing protein production and

secretion than minimal media alone. This was particularly interesting as peptone is an efficient source of nutrients for most cells[285]

The difference in expression of proteins between cell strains is unlikely to only be present in the recombinant protein alone. There appears to be a small difference in PDI levels between cell types, with the pattern observed of SK1 being the highest protein producer maintained, however PGK levels shown as a loading control do not vary, indicating that PDI levels are changing, not just total protein loaded. PDI levels being higher in SK1 cells would make sense as we do know that increased PDI levels lead to an increase in recombinant protein expression levels[198,200]. The question remains whether PDI levels appear higher in the SK1 cells because it is a better producer of protein, or whether the higher levels of PDI observed could be one of the factors which make SK1 a good producer of recombinant proteins to begin with. The latter in this situation appears more likely due to the stable expression of PGK across the strains. The second noteworthy phenomenon noted in PDI expression across cell lines is that the control cells (that being those which were not transformed to express recombinant luciferase) showed higher PDI levels combined than each luciferase secreting cell individually. Significance was low when control cells were compared to SK1 (P=0.96), YJM978 (P=0.91) or YPS128 (P=0.25) and so effects could be purely biological variability. A quantitative experiment on PDI levels in different cell types could be interesting to consider how the PDI levels vary and to accurately detect whether the PDI levels do appear to change in cells producing recombinant proteins.

It was very interesting that the suggested optimal conditions within the experimental space for SC-ura were opposites after the design of experiments work (figure 32). As

rP translation increases, the level of secreted protein also increases but less so than levels of intracellular protein - this would indicate that secretion is a bottle neck. It could be that limiting nutrients causes a slower production of proteins which gives the cells more time to produce and secrete the proteins, compared to more nutrients leading to more transcription of mRNA and can cause a backlog of protein production leading to ER stress. Whether there is a difference in protein quality dependent on the speed of production is currently unknown. It would, however, stand to reason that a slower production which did not maximise the capacity of the ER, would be less likely to lead to errors and so could lead to a higher quality product. This of course depends on where the limitation is in protein production. Current beliefs are that bottlenecks occur mostly at the protein folding and translocation stages[286] .

Although increased concentration of media components did not improve the percentage of secretion in SC-ura, it did increase overall secretion by the fact that more overall luciferase was produced in more concentrated media. When YPD is compared, which has much higher nutritional content, the production and secretion of proteins was vastly improved (figure 29). This pattern was see even without integration of stable DNA into the genome, and it could be likely that some plasmid was lost in YPD growth, and therefore even more protein may be produced in a stable expression system. Ability to stably express recombinant DNA has been the subject of much study [287], and stable expression through genomic integration is now a more common method of industrial recombinant protein production.

The comparison of the different cell types with the optimised medias shows that the different cell types react differently to the different nutrient profiles. Although the SK1

strain is improved in luciferase production and secretion (as you would expect) the YJM978 strain is not significantly improved compared to the regular YPD recipe (figure 38). The YPS128 strain which has been the lowest luciferase producer and secretor throughout the project showed the most significant improvements with the optimised media.

It is currently unknown why the strains differ so much in secretion and why they interact differently with the media components. It would be interesting to consider how different cell types grow in different media styles and whether the ER stress is changed in the different strains once they are optimised for higher throughput. Which genome components that differ between the cell strains cause the increased secretion and whether they could be used to further understand the changes in response to medias would be an interesting study, as would the changes in mRNA profiles between the cell strains.

Further work on this project would be to consider how the genetic differences within the strains lead to the changes that we see in recombinant protein production, as well as the study of whether these effects are seen due to increased transcription, translation or a combination of the two.

Chapter 6 Discussion

Discussion

6.1 the effects of PDI

This study aimed to investigate a variety of aspects involved in the biology of recombinant protein production in *S. cerevisiae*. Media composition, strain background and introduction of orthogonal pathways were all explored. During this study we considered the effects of adding human GPx into *Saccharomyces cerevisiae*. The oxidative protein folding pathway in the *Saccharomyces* ER uses Pdi1 and Ero1 as key components [288,289]. Electrons flow from the reduced proteins using the Pdi1 and Ero1 proteins to allow formation of disulfide bonds. The hope at the beginning of the study was that addition of GPx proteins would increase oxidised Pdi1 and therefore increase expression of our *Gaussia* luciferase reporter. Outside research has shown that yeast and human PDI have similar structures[189,290], but that Ero proteins have structural differences [189,291], even whilst performing similar functions. The similarity in PDI protein structure between species led us to hypothesize that GPx proteins may perform similar functions in yeast to those they perform in humans.

Eug1, a paralog of Pdi1 with non-standard CXXS catalytic sites compared to the conserved CXXC motif in eukaryotic Pdi proteins[292]. Deletion of PDI1 in yeast is fatal for cells [289], although viability can be restored by expression of the EUG1 gene which has homology with both yeast and human PDI, although is non essential in itself [293]. Yeast cells are also non viable upon deletion of ERO1[294]. The more complex human oxidative protein folding pathway has a wider variety of ER resident proteins including peroxiredoxin IV which can act in the oxidative-reductive processes with PDI enabling human cells to remain viable even without a functional ERO gene [295],

only showing a slight delay in disulfide bond formation. The fact that yeast cells require a functional Ero gene, but human cells do not despite similar functionality, shows that the additional mechanisms within human cells to reoxidise PDI are adequate to sustain cell viability, however the yeast system cannot support this. Taken together these result indicate that oxidative folding in yeast can be quite malleable, and should therefore be amenable to useful manipulations with the aim of supporting production of disulfide-bond containing recombinant proteins

The formation of disulfide bonds in new proteins is, for both mammals and yeast, essential, yet our research showed that addition of GPx proteins which should increase oxidised Pdi1 and therefore increase *de novo* formation of disulfide bonds, we did not see an increase in recombinant protein production (figure 19). This was unexpected due to the numerous studies which cite increased PDI levels showing increased recombinant protein levels[190,199,200]. It would seem logical that increasing oxidised PDI would increase *de novo* synthesis and therefore protein secretion, but as this did not occur in our study, one has to question what it is that Pdi1 increase is doing within the cells. Interestingly it has been shown that increased Pdi1 levels can increase recombinant protein production, even of proteins which do not contain disulfide bonds[198]. This shows that the increases in recombinant protein production seen in samples with PDI overexpression, are not solely due to increased de novo bond forming activities, but may result from alterations to general ER biology.

To consider what excess PDI could be contributing to the cell, outside of increased capacity for disulfide bonding, previous studies investigated the activity of PDI. It was found that when yeast oxidative protein folding pathways have been analysed by monitoring of disulfide bond formation and oxygen consumption [296], the yeast ER is

theorised to have a capacity of 400,000 disulfide bonds per minute via the PDI oxidative protein folding system if used purely as a *de novo* bond forming protein. In the same study computational estimates showed a maximum requirement of only 269,000 disulfide bonds based on the composition of the endogenous yeast proteome, indicating that the capacity of the ER for disulfide bonding is higher than need. This may imply that the isomerase function is the rate limiter in recombinant protein production, and this could explain why overexpression of PDI1 increases recombinant protein production [190,200] even if there is sufficient Pdi1 available to perform all of the required de novo folding functions within a cell.

If it was primarily the isomerase function of PDI that increased recombinant protein production upon overexpression in yeast, this could begin to explain the results we saw in (figure 19) In our study we saw reduced levels of recombinant protein upon expression of recombinant GPX. If increased oxidation of PDI shifted function to *de novo* bond formation, cells may lose some of the capacity for PDI isomerisation. If the isomerisation is the driving factor of increased recombinant protein expression in excess PDI strains, then increased capacity for disulfide bonding at the expense of isomerase activity, would explain why the GPx samples saw a reduction in protein production, as they may have exhibited a reduction in Pdi1 isomerase activity. It was hoped that confirming the oxidation states of Pdi1 and the interaction between the Pdi1 and GPx proteins could be shown to add to the evidence of this hypothesis. In counter to this explanation, the rate of thiol oxidation in cells has been shown to be linearly dependent on Pdi1p and Ero1p concentrations[296,] which may imply that increased recombinant protein production and secretion in cells over expressing PDI would increase recombinant protein levels due to this increased ability for oxidation. To confirm these findings further work would need to be conducted into the

interaction of GPx with yeast Pdi1, and an accurate and immediate method used to identify the oxidation state, such as immunoblotting or ELISA with antibodies unique to oxidation states.

It is also possible that the GPx proteins simply caused a reduction in recombinant protein production due to the extra pressure upon the cell to produce further recombinant proteins.

Expression of recombinant protein to high levels can cause respiratory and metabolic burden upon the cells. In the case of the cells modified to produce humanised GPx as well as *Gaussia* luciferase, multiple proteins are being produced which could cause further impact. The expression of multiple recombinant proteins to high levels could reduce recombinant protein production and secretion capacity due to oxidative stress[297]. It could be thought that the reduction in luciferase seen upon additional expression of GPx proteins (figure 20) could be due to the metabolic burden upon cells increasing, rather than due to direct activity of the GPx proteins themselves. This is not believed to be the case as the flow cytometry data gathered (figure 21) showed cells in general were not under conditions of high oxidative stress, seen by generally low levels of H_2DCFDA staining. Activation of oxidative stress response genes can increase metabolic rates and reduce the burden on cells of increased ROS accumulation to increase protein capacity[297]. As GPx proteins should sequester some of the hydrogen peroxide produced in recombinant protein production, this could be theorised to have a similar affect on preventing ROS accumulation, however this effect was not shown in increased recombinant protein levels, and hence it appears likely that the effects of GPx proteins themselves are acting to reduce the levels of recombinant protein produced.

This could potentially explain why the GPx cells appeared to show higher oxidative stress than cells with one of the empty vectors in (figure 21), however an alternate empty vector showed increased oxidative stress compared to the GPx transformed cells so these data are still unclear. Genetic manipulation of cells can also have unintended off target affects, and it becomes difficult to predict how these may interact. When expressing an anti-transferrin receptor single-chain antibody in yeast, co overexpression of yeast PDI1 and heavy chain binding protein showed a huge 10 fold increase in gene expression, but this effect was not compounded with the ~7 fold increase in secretion seen by optimisation of gene dosage [298]. The interaction between factors cannot be detected using typical experiments where one factor is changed at a time. This is why in later experiments, when considering media optimisation, multi factor experimental designs were used to fully explore how factors interacted.

6.2 The effects of different proteins and strains

What is clear from data gained in this study as well as the collective literature available on the topic, is that different proteins and different strains can react very differently to changes. Strain variation has been shown to affect secretion of different proteins over an up to 10 fold range, and at the time these changes were not found to coincide with specific genetic markers[299]. This complements our results in figure 25 where individual strains showed strong variability in protein production and secretion levels. It would be interesting in the future to consider whether a similar lack of genetic markers between high producing strains was seen amongst the strains tested in this study also.

There are no comprehensive surveys that as yet have identified the differences between the strains used in this study which could be causing high levels of variability in recombinant protein production and secretion (such as seen in figure 25). It is likely that between strains there will be snps in expected places such as translational or secretory genes, however further study would be needed to confirm this. There are many ways in which saccharomyces cerevisiae have been engineered in the secretory pathway to improve recombinant protein secretion[300], including overexpression of protein folding chaperones, heat shock proteins, PDI, vesicle trafficking proteins and under expression of proteases. It is currently unknown whether the strains used in this study are showing increased protein expressions due to these factors or not. It is important to note here, however, that although we can fully sequence a genome, we cannot claim to fully understand everything about it, with large areas traditionally considered to be "junk DNA" which are now emerging to have more important function than we had previously considered[301,302]. Not only have different strains been shown to vary greatly in recombinant protein production and secretion, how they react to genetic influence is unique also. When testing the effects of OPI1 deletion on antibody secretion in yeast, it was found the effect differed greatly depending on the strain tested [303]. In our study we were only able to test the effects of the GPX proteins upon one strain, but it would be fascinating to know whether all strains showed a similar effect. Particularly strains which undergo high levels of reactive oxidative stress (such as the alpha cox 4 strain used in flow cytometry as a positive control), to see whether increased oxidised PDI relieved some of the strain on the UPR.

6.3 the effects of single deletions

Single gene deletion and addition studies have, like most things, pros and cons. In cases like PDI, overexpression appears almost universally to improve recombinant protein levels, however understanding whether these effects can be seen in all cell lines, and how they may interact with other genes in a cellular system is much too complicated to predict with current understanding and technology. In fact, unpublished results within our laboratory have found cases of recombinant proteins which remain unaffected by excess Pdi1 expression. Given the results found in this study showing the wide variance of a cells reaction to given conditions, it seems unlikely that these cases would be unique. In fact it is likely that other groups have had similar findings, but they remain unpublished. Publication bias is a significant problem in which positive findings are more likely to make it into the literature[304], and in this case could lead to incorrect conclusions such as “increased PDI expression increases recombinant protein expression” which are not wholly true. Computational methods go some way to try and form predictions upon how genes may be affected by eachother, however a computer model can only consider what it is told to, and we are still learning of new genes and functions, and so it is impossible currently to create a perfect model.

6.4 Media Optimisation

With current understanding both in the literature, and with the data shown In this study, the only true way to optimise protein yield is to test each strain and recombinant protein combination at the beginning of the production process. In our system, deletion of the *STE12* gene did not increase production of recombinant luciferase (figures 7 and 13), however we cannot say for certain whether results

would be different with a different protein. As seen in our media experiments, the optimum conditions seem to be tailored to individual strains (figure 38). Other media optimisation studies have looked at different media components completely. For example when design of experiments was employed to improve production of recombinant hirudin in *Saccharomyces cerevisiae* using a response surface design, optimised media was found to increase culture productivity 35% compared to standard medium [305]. In this study, in contrast to ours, 6 components were used (yeast extract, peptone, casamino acids, ammonium sulphate, potassium phosphate and galactose) and the optimum determined. From their original media recipe, yeast extract was increased by 60% to ~15g/L, peptone increased by 33% to ~13g/L, casamino acids increased by 45% to ~7g/L, ammonium sulphate decreased by 25% to ~7g/L, potassium phosphate increased by 24% to ~12g/L and galactose increased by a modest 10% to 33.1 g/L. Previous research has shown SC medium reduces recombinant protein secretion compared to other synthetic minimal media (SD) [306]. It was shown that individual amino acids appeared non toxic when supplemented into SD medias, but that the total composition of amino acids in SC medium appears to have negative effects on the cell. Secretion of phosphatase was 2.7 fold higher in SD compared to SC when compared. Interactions between different amino acids could be explored in depth using design of experiments to further understand this mechanism. The effects of individual amino acids were considered using design of experiments and researchers able to group amino acid supplements into ones which appeared to improve secretion, have negligible effects or reduce secretion of the phosphatase [306].This could explain why in our study (figure 30) supplementation with mixed amino acids only lacking in uracil did not improve secretion. In future considering individual amino acid supplementation may be more beneficial.

Another parameter which appears to have impact on recombinant protein levels is the growth method used. All of our studies were completed on cells grown in small batches shaking. In industry different methods are used to overcome the challenges of scale. For example multiple copies of integrated plasmids leading to a "supersecreting" yeast strain showed secretion of pro-urokinase was growth associated [307], However the method of cell growth appeared to have a large impact upon final protein titre, with nearly 2X the final titer in the perturbed batch fermentation compared to the constant respiratory quotient culture, even with a ~40% lower final cell density [307]. The importance of cell growth upon recombinant protein secretion here appears to be dependant on the method of growth used, and again individual protein expressed. Different proteins appear to be affected by growth rates of cells differently, for example human insulin precursor appears to be produced in line with growth rates, yet amylase appears to give higher yields at lower growth rates [308].

6.5 Genetic mutations

The differences in effects of genetic mutation upon production and secretion of different recombinant proteins at first seemed surprising, however it is in line with the other conclusions of this study in the fact that each individual expression system for each protein needs separate optimisation for highest protein yield. Over expression of heat shock response genes was shown to increase amylase yield between 25 and 70%, invertase 94-118% but only moderately improved secretion of insulin [309]. Alternatively PSE1 and SOD1, two native *Saccharomyces cerevisiae* genes, have been over expressed to improve titres of secreted protein in yeast [310], and increased secreted protein yield of different cellulases. This increased protein activity up to 447% compared to the baseline without any overexpressed protein, however varied

widely between the different cellulases tested. In this study individual transformants were screened for the highest activity, and it is noteworthy that there appears to be a range of activity between different transformants, which was also noticed in the study we conducted. This is perhaps not surprising knowing that genes are constantly and spontaneously mutating within cells [311,312], and especially in haploid cells where these mutations are less likely to be detrimental [311].

Over expression of SEC1 and SLY1, which are involved in the SNARE complex that traffics proteins between different organelles in the protein secretory pathway, has been shown to improve human insulin precursor, invertase and amylase protein secretion in *S. cerevisiae*, however again the level of improvement depends very much on which protein is being produced and which protein over expressed [313]. When these three examples are considered together it becomes very clear that the protein being produced has a huge impact on how favourable genetic manipulations are. This again ties in with the hypothesis that the ste12 deletions initially tested, and the addition of GPx proteins, may have had different effects, or different magnitudes, on different recombinant proteins.

The most important takeaway from this study is the need to optimise systems for individual protein. The table below summarises just some of the different ways researchers have tried to increase recombinant protein production. This illustrates how so many proteins have different optimum conditions, and supports the hypothesis that there are little or no “one size fits all” approaches to optimising secretion and protein production in yeast.

What protein was looked at	How was it analysed	How much improvement could be seen	how many (parent) strains tested	Reference
Human platelet derived growth factor B homodimer	Overexpression PDI	10 fold	1	[199]
Schizosaccharomyces pombe acid phosphatase	overexpression PDI	4 fold	1	[199]
IgG	deletion OPI1	4 fold	67	[303]
Cutinase	media optim	2-3 fold	1	[314]
Hirudin	Design of experiments media optimisation fermentation	35%	1	[305]

Phosphatase	New amino acid supplement formulations	8 fold	1	306
Various	activated heat shock response pathways	0-118%	1	309
3 different cellulase proteins	overexpression of SOD1 and/or PSE1	10-447%.	1	310
human insulin precursor,α-amylase, and endogenous protein invertase	over expression of SEC1 and SLY1	30-62%	1	313
pro-urokinase	constant respiratory quotient growth vs perturbed batch	2 fold	1	307

anti-transferrin receptor single-chain antibody	overexpression of heavy chain binding protein and protein disulfide isomerase	10 fold	1	[298]

6.8 final conclusions

This study has illustrated the lack of predictability in identification of ideal conditions for recombinant protein secretion, and in our study found media optimisation to be the largest positive affect. Media optimisation also varied for the strains we tested, and it would seem likely from data in the literature that this would also be affected by growth type, genetic manipulations and what protein was being produced. The true value of these changes needs to be considered with cost to industry [315], which can also be added as a variable in design of experiments, but also needs to be using scaled up equipment as growing at scale is another element which could change results. To conclude, we have considered 3 different approaches to attempt to optimise protein secretion in S*accharomyces cerevisiae*. Ste12 deletion, and addition of GPx proteins, both did not improve production or secretion of recombinant luciferase in BY4741, however wider literature examples could give hope that these changes would have different effects on other proteins. Industry will require simple universally applicable methodologies to be created to allow for the best optimisations

to be carried out at the beginning of production cycles. The factors which are chosen need to be those which have the greatest effects on the greatest number of proteins to try and reduce resources in the development stage. Optimisation of PDI levels and the pathways surrounding the ER, alongside media optimisation are 2 places that show significant promise.

REFERENCES

1. Tokareva, O., Michalczechen-Lacerda, V. A., Rech, E. L. & Kaplan, D. L. Recombinant DNA production of spider silk proteins. *Microbial Biotechnology* **6**, 651–663 (2013).

2. Hettinga, K. & Bijl, E. Can recombinant milk proteins replace those produced by animals? *Current Opinion in Biotechnology* **75**, 102690 (2022).

3. Recombinant Protein Market - Growth, Trends, Covid-19 Impact, and Forecasts (2021 - 2026). https://www.researchandmarkets.com/reports/4622467/recombinant-protein-market-growth-trends.

4. Tran, K. *et al.* Cell-free production of a therapeutic protein: Expression, purification, and characterization of recombinant streptokinase using a CHO lysate. *Biotechnology and Bioengineering* **115**, 92–102 (2018).

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

6. Naesby, M. *et al.* Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in Saccharomyces cerevisiae. *Microbial Cell Factories* **8**, 45 (2009).

7. Lagassé, H. A. D. *et al.* Recent advances in (therapeutic protein) drug development. *F1000Res* **6**, 113 (2017).

8. Fernandes, J. C. Therapeutic application of antibody fragments in autoimmune diseases: current state and prospects. *Drug Discovery Today* (2018) doi:10.1016/j.drudis.2018.06.003.

9. Lee, A., Sun, S., Sandler, A. & Hoang, T. Recent progress in therapeutic antibodies for cancer immunotherapy. *Current Opinion in Chemical Biology* **44**, 56–65 (2018).

10. Sécher, T. *et al.* Therapeutic antibodies: A new era in the treatment of respiratory diseases? *Pharmacology & Therapeutics* (2018) doi:10.1016/j.pharmthera.2018.05.003.

11. Pelfrene, E., Mura, M., Cavaleiro Sanches, A. & Cavaleri, M. Monoclonal antibodies as anti-infective products: a promising future? *Clinical Microbiology and Infection* (2018) doi:10.1016/j.cmi.2018.04.024.

12. Frau, J., Coghe, G., Lorefice, L., Fenu, G. & Cocco, E. New horizons for multiple sclerosis therapeutics: milestones in the development of ocrelizumab. *Neuropsychiatric Disease and Treatment* **Volume 14**, 1093–1099 (2018).

13. Kinch, M. S. An overview of FDA-approved biologics medicines. *Drug Discovery Today* **20**, 393–398 (2015).

14. Nelson, A. L., Dhimolea, E. & Reichert, J. M. Development trends for human monoclonal antibody therapeutics. *Nature Reviews Drug Discovery* **9**, 767–774 (2010).

15. Monoclonal Antibodies Approved by the EMA and FDA for Therapeutic Use – ACTIP. http://www.actip.org/products/monoclonal-antibodies-approved-by-the-ema-and-fda-for-therapeutic-use/.

16. Enever, C., Batuwangala, T., Plummer, C. & Sepp, A. Next generation immunotherapeutics—honing the magic bullet. *Current Opinion in Biotechnology* **20**, 405–411 (2009).

17. Nylen, A. & Chen, M. T. Production of Full-Length Antibody by Pichia pastoris. *Methods Mol Biol* **1674**, 37–48 (2018).

18. de Ruijter, J. C., Koskela, E. v. & Frey, A. D. Enhancing antibody folding and secretion by tailoring the Saccharomyces cerevisiae endoplasmic reticulum. *Microbial Cell Factories* **15**, 1–18 (2016).

19. Wang, Y. *et al.* Expression of antibody fragments in Saccharomyces cerevisiae strains evolved for enhanced protein secretion. *Microbial Cell Factories* **20**, 1–17 (2021).

20. Liu, X. *et al.* Engineering yeast for the production of breviscapine by genomic analysis and synthetic biology approaches. *Nature Communications* **9**, 448 (2018).

21. Walker, R. S. K. & Pretorius, I. S. Applications of Yeast Synthetic Biology Geared towards the Production of Biopharmaceuticals. *Genes (Basel)* **9**, (2018).

22. Balkani, S., Shamekhi, S., Raoufinia, R., Parvan, R. & Abdolalizadeh, J. Purification and Characterization of Bovine Serum Albumin Using Chromatographic Method. *Advanced Pharmaceutical Bulletin* **6**, 651 (2016).

23. Dell, A., Galadari, A., Sastre, F. & Hitchen, P. Similarities and differences in the glycosylation mechanisms in prokaryotes and eukaryotes. *International Journal of Microbiology* (2010) doi:10.1155/2010/148178.

24. Sewalt, V., Shanahan, D., Gregg, L., la Marta, J. & Carillo, R. The Generally Recognized as Safe (GRAS) process for industrial microbial enzymes. *Industrial Biotechnology* **12**, 295–302 (2016).

25. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Frontiers in Microbiology* **5**, 172 (2014).

26. Khan, K. H. Gene Expression in Mammalian Cells and its Applications. *Advanced Pharmaceutical Bulletin* **3**, 257 (2013).

27. Merkuleva, I. A. *et al.* Comparative Immunogenicity of the Recombinant Receptor-Binding Domain of Protein S SARS-CoV-2 Obtained in Prokaryotic and Mammalian Expression Systems. *Vaccines (Basel)* **10**, (2022).

28. Hammer, S. K. & Avalos, J. L. Harnessing yeast organelles for metabolic engineering. *Nature Chemical Biology* **13**, 823–832 (2017).

29. Hong, K. K. & Nielsen, J. Metabolic engineering of Saccharomyces cerevisiae: A key cell factory platform for future biorefineries. *Cellular and Molecular Life*

Sciences vol. 69 2671–2690 Preprint at https://doi.org/10.1007/s00018-012-0945-1 (2012).

30. Lian, J., Mishra, S. & Zhao, H. Recent advances in metabolic engineering of Saccharomyces cerevisiae: New tools and their applications. *Metabolic Engineering* vol. 50 85–108 Preprint at https://doi.org/10.1016/j.ymben.2018.04.011 (2018).

31. Gündüz Ergün, B., Hüccetoğulları, D., Öztürk, S., Çelik, E. & Çalık, P. Established and Upcoming Yeast Expression Systems. in *Methods in molecular biology (Clifton, N.J.)* vol. 1923 1–74 (2019).

32. Baghban, R. *et al.* Yeast Expression Systems: Overview and Recent Advances. *Molecular Biotechnology* 1–20 (2019) doi:10.1007/s12033-019-00164-8.

33. Chen, X. & Petranovic, D. Role of frameshift ubiquitin B protein in Alzheimer's disease. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine* **8**, 300–313 (2016).

34. Verbandt, S., Cammue, B. P. A. & Thevissen, K. Yeast as a model for the identification of novel survival-promoting compounds applicable to treat degenerative diseases. *Mechanisms of Ageing and Development* **161**, 306–316 (2017).

35. Lasserre, J.-P. *et al.* Yeast as a system for modeling mitochondrial disease mechanisms and discovering therapies. *Disease Models & Mechanisms* **8**, 509–526 (2015).

36. Voisset, C., García-Rodríguez, N., Birkmire, A., Blondel, M. & Wellinger, R. E. Using yeast to model calcium-related diseases: Example of the Hailey–Hailey disease. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**, 2315–2321 (2014).

37. Smith, M. G. & Snyder, M. Yeast as a Model for Human Disease. in *Current Protocols in Human Genetics* vol. Chapter 15 Unit 15.6 (John Wiley & Sons, Inc., 2006).

38. Liu, W. *et al.* From Saccharomyces cerevisiae to human: The important gene co-expression modules. *Biomedical Reports* **7**, 153–158 (2017).

39. Thak, E. J., Yoo, S. J., Moon, H. Y. & Kang, H. A. Yeast synthetic biology for designed cell factories producing secretory recombinant proteins. *FEMS Yeast Res* **20**, (2020).

40. Kalsner, I., Schneider, F. J., Geyer, R., Ahorn, H. & Maurer-Fogy, I. Comparison of the carbohydrate moieties of recombinant soluble Fc epsilon receptor (sFc epsilon RII/sCD23) expressed in Saccharomyces cerevisiae and Chinese hamster ovary cells. Different O-glycosylation sites are used by yeast and mammalian cells. *Glycoconj J* **9**, 209–216 (1992).

41. Martínez-Duncker, I., Díaz-Jímenez, D. F. & Mora-Montes, H. M. Comparative Analysis of Protein Glycosylation Pathways in Humans and the Fungal Pathogen Candida albicans. *International Journal of Microbiology* **2014**, (2014).

42. Vervecken, W. *et al.* In vivo synthesis of mammalian-like, hybrid-type N-glycans in Pichia pastoris. *Applied and Environmental Microbiology* **70**, 2639–2646 (2004).

43. Smekenov, I. *et al.* Heterologous secretory expression of β-glucosidase from Thermoascus aurantiacus in industrial Saccharomyces cerevisiae strains. *Brazilian Journal of Microbiology* **51**, 107–123 (2020).

44. Huang, M., Bao, J., Hallström, B. M., Petranovic, D. & Nielsen, J. Efficient protein production by yeast requires global tuning of metabolism. *Nature Communications* **8**, 1131 (2017).

45. De Pourcq, K., De Schutter, K. & Callewaert, N. Engineering of glycosylation in yeast and other fungi: current state and perspectives. *Applied Microbiology and Biotechnology* **87**, 1617–1631 (2010).

46. Huang, M., Wang, G., Qin, J., Petranovic, D. & Nielsen, J. Engineering the protein secretory pathway of Saccharomyces cerevisiae enables improved protein production. *Proc Natl Acad Sci U S A* **115**, E11025–E11032 (2018).

47. Idiris, A., Tohda, H., Kumagai, H. & Takegawa, K. Engineering of protein secretion in yeast: Strategies and impact on protein production. *Applied Microbiology and Biotechnology* vol. 86 403–417 Preprint at https://doi.org/10.1007/s00253-010-2447-0 (2010).

48. Kim, H., Yoo, S. J. & Kang, H. A. Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS Yeast Research* **15**, n/a-n/a (2014).

49. Nocon, J. *et al.* Model based engineering of Pichia pastoris central metabolism enhances recombinant protein production. *Metabolic Engineering* **24**, 129–138 (2014).

50. Saitua, F., Torres, P., Pérez-Correa, J. R. & Agosin, E. Dynamic genome-scale metabolic modeling of the yeast Pichia pastoris. *BMC Systems Biology* **11**, 27 (2017).

51. Besada-Lombana, P. B. & Da Silva, N. A. Engineering the early secretory pathway for increased protein secretion in Saccharomyces cerevisiae. *Metabolic Engineering* **55**, 142–151 (2019).

52. Piirainen, M. A., Boer, H., de Ruijter, J. C. & Frey, A. D. A dual approach for improving homogeneity of a human-type N-glycan structure in Saccharomyces cerevisiae. *Glycoconjugate Journal* **33**, 189–199 (2016).

53. Bonnet, C. *et al.* PCR on yeast colonies: an improved method for glyco-engineered Saccharomyces cerevisiae. *BMC Research Notes* **6**, 201 (2013).

54. Stovicek, V., Holkenbrink, C. & Borodina, I. CRISPR/Cas system for yeast genome engineering: advances and applications. *FEMS Yeast Research* **17**, (2017).

55. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–41 (2009).

56. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563–7 (1996).

57. Giaever, G. & Nislow, C. The Yeast Deletion Collection: A Decade of Functional Genomics. *Genetics* **197**, 451–465 (2014).

58. Thak, E. J., Kim, J., Lee, D. J., Kim, J. Y. & Kang, H. A. Structural analysis of N-/O-glycans assembled on proteins in yeasts. *Journal of Microbiology* **56**, 11–23 (2018).

59. Gündüz Ergün, B., Hüccetoğulları, D., Öztürk, S., Çelik, E. & Çalık, P. Established and upcoming yeast expression systems. in *Methods in Molecular Biology* vol. 1923 1–74 (Humana Press Inc., 2019).

60. Manfrão-Netto, J. H. C., Gomes, A. M. V. & Parachin, N. S. Advances in Using Hansenula polymorpha as Chassis for Recombinant Protein Production. *Frontiers in Bioengineering and Biotechnology* **7**, (2019).

61. Yoo, S. J., Moon, H. Y. & Kang, H. A. Screening and selection of production strains: Secretory protein expression and analysis in Hansenula polymorpha. in *Methods in Molecular Biology* vol. 1923 133–151 (Humana Press Inc., 2019).

62. Delic, M. *et al.* The secretory pathway: exploring yeast diversity. doi:10.1111/1574-6976.12020.

63. Nevalainen, H. & Peterson, R. Making recombinant proteins in filamentous fungi- Are we expecting too much? *Frontiers in Microbiology* **5**, 75 (2014).

64. Ward, O. P. Production of recombinant proteins by filamentous fungi. *Biotechnology Advances* **30**, 1119–1139 (2012).

65. Deshpande, N., Wilkins, M. R., Packer, N. & Nevalainen, H. Protein glycosylation pathways in filamentous fungi. *Glycobiology* **18**, 626–637 (2008).

66. Awan, A. R. *et al.* Biosynthesis of the antibiotic nonribosomal peptide penicillin in baker's yeast. *Nature Communications* **8**, 15202 (2017).

67. Felnagle, E. A. *et al.* Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. *Molecular Pharmaceutics* **5**, 191–211 (2008).

68. Li, S., Li, Y. & Smolke, C. D. Strategies for microbial synthesis of high-value phytochemicals. *Nature Chemistry* **10**, 395–404 (2018).

69. Dai, Z., Liu, Y., Guo, J., Huang, L. & Zhang, X. Yeast synthetic biology for high-value metabolites. *FEMS Yeast Research* **15**, n/a-n/a (2014).

70. Chen, B. *et al.* Synthetic biology toolkits and applications in Saccharomyces cerevisiae. *Biotechnology Advances* **36**, 1870–1881 (2018).

71. Wang, G., Huang, M. & Nielsen, J. Exploring the potential of Saccharomyces cerevisiae for biopharmaceutical protein production. *Current Opinion in Biotechnology* vol. 48 77–84 Preprint at https://doi.org/10.1016/j.copbio.2017.03.017 (2017).

72. Chevance, F. F. V., Le Guyon, S. & Hughes, K. T. The Effects of Codon Context on In Vivo Translation Speed. *PLoS Genetics* **10**, e1004392 (2014).

73. Brule, C. E. & Grayhack, E. J. Synonymous Codons: Choose Wisely for Expression. *Trends in Genetics* vol. 33 283–297 Preprint at https://doi.org/10.1016/j.tig.2017.02.001 (2017).

74. Chung, B. K. S. & Lee, D. Y. Computational codon optimization of synthetic gene for protein expression. *BMC Systems Biology* **6**, 134 (2012).

75. Ahn, J. *et al.* Codon optimization of Saccharomyces cerevisiae mating factor alpha prepro-leader to improve recombinant protein production in Pichia pastoris. *Biotechnology Letters* **38**, 2137–2143 (2016).

76. Zhao, H., Blazanovic, K., Choi, Y., Bailey-Kellogg, C. & Griswold, K. E. Gene and protein sequence optimization for high-level production of fully active and aglycosylated lysostaphin in Pichia pastoris. *Applied and Environmental Microbiology* **80**, 2746–2753 (2014).

77. Johnston, C., Martin, B., Fichant, G., Polard, P. & Claverys, J. P. Bacterial transformation: Distribution, shared mechanisms and divergent control. *Nature Reviews Microbiology* vol. 12 181–196 Preprint at https://doi.org/10.1038/nrmicro3199 (2014).

78. Yoshida, N. & Sato, M. Plasmid uptake by bacteria: A comparison of methods and efficiencies. *Applied Microbiology and Biotechnology* vol. 83 791–798 Preprint at https://doi.org/10.1007/s00253-009-2042-4 (2009).

79. Kawai, S., Hashimoto, W. & Murata, K. Transformation of Saccharomyces cerevisiae and other fungi: methods and possible underlying mechanism. *Bioeng Bugs* **1**, 395–403 (2010).

80. Oliveira, P. H., Touchon, M. & Rocha, E. P. C. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res* **42**, 10618–31 (2014).

81. Vasu, K. & Nagaraja, V. Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. *Microbiology and Molecular Biology Reviews* **77**, 53–72 (2013).

82. Yang, X., Xu, M. & Yang, S. T. Restriction modification system analysis and development of in vivo methylation for the transformation of Clostridium cellulovorans. *Applied Microbiology and Biotechnology* **100**, 2289–2299 (2016).

83. Zhang, G. *et al.* A Mimicking-of-DNA-Methylation-Patterns Pipeline for Overcoming the Restriction Barrier of Bacteria. *PLoS Genetics* **8**, e1002987 (2012).

84. Sfeir, A. & Symington, L. S. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends in Biochemical Sciences* vol. 40 701–714 Preprint at https://doi.org/10.1016/j.tibs.2015.08.006 (2015).

85. Nakade, S. *et al.* Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nature Communications* **5**, 1–8 (2014).

86. Li, L., Liu, X., Wei, K., Lu, Y. & Jiang, W. Synthetic biology approaches for chromosomal integration of genes and pathways in industrial microbial

systems. *Biotechnology Advances* vol. 37 730–745 Preprint at https://doi.org/10.1016/j.biotechadv.2019.04.002 (2019).

87. Rothstein, R. Targeting, Disruption, Replacement, and Allele Rescue: Integrative DNA Transformation in Yeast. *Methods in Enzymology* **194**, 281–301 (1991).

88. Warren P. Voth, J. D. R. J. M. S. D. J. S. Yeast vectors for integration at the HO locus. *Nucleic Acids Research* vol. 29 e59 Preprint at (2001).

89. Çelik, E. & P Çalık. Production of recombinant proteins by yeast cells. *Avances Biotechnology* (2012).

90. Gnügge, R. & Rudolf, F. *Saccharomyces cerevisiae* Shuttle vectors. *Yeast* **34**, 205–221 (2017).

91. Kotula, L. & Curtis, P. J. Evaluation of Foreign Gene Codon Optimization in Yeast: Expression of a Mouse IG Kappa Chain. *Bio/Technology* **9**, 1386–1389 (1991).

92. Mauro, V. P. Codon Optimization in the Production of Recombinant Biotherapeutics: Potential Risks and Considerations. *BioDrugs* **32**, 69–81 (2018).

93. Li, L., Liu, X., Wei, K., Lu, Y. & Jiang, W. Synthetic biology approaches for chromosomal integration of genes and pathways in industrial microbial systems. *Biotechnology Advances* vol. 37 730–745 Preprint at https://doi.org/10.1016/j.biotechadv.2019.04.002 (2019).

94. Tyo, K. E. J., Ajikumar, P. K. & Stephanopoulos, G. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nature Biotechnology* **27**, 760–765 (2009).

95. Xiao, Y., Bowen, C. H., Liu, D. & Zhang, F. Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis. *Nature Chemical Biology* **12**, 339–344 (2016).

96. Taipakova, S. M., Smekenov, I. T., Saparbaev, M. K. & Bissenbaev, A. K. Characterization of Aspergillus niger endo-1,4-β-glucanase ENG1 secreted from Saccharomyces cerevisiae using different expression vectors. *Genetics and Molecular Research* **14**, 6439–6452 (2015).

97. Chen, X. & Zhang, J. The Genomic Landscape of Position Effects on Protein Expression Level and Noise in Yeast. *Cell Systems* **2**, 347–354 (2016).

98. Wu, X. Le *et al.* Genome-wide landscape of position effects on heterogeneous gene expression in Saccharomyces cerevisiae. *Biotechnology for Biofuels* **10**, 189 (2017).

99. Schwartz, C., Shabbir-Hussain, M., Frogue, K., Blenner, M. & Wheeldon, I. Standardized Markerless Gene Integration for Pathway Engineering in Yarrowia lipolytica. *ACS Synthetic Biology* **6**, 402–409 (2017).

100. Moon, H. Y. *et al.* A new set of rDNA-NTS-based multiple integrative cassettes for the development of antibiotic-marker-free recombinant yeasts. *Journal of Biotechnology* **233**, 190–199 (2016).

101. Shi, S., Liang, Y., Zhang, M. M., Ang, E. L. & Zhao, H. A highly efficient single-step, markerless strategy for multi-copy chromosomal integration of large biochemical pathways in Saccharomyces cerevisiae. *Metabolic Engineering* **33**, 19–27 (2016).

102. Aw, R. & Polizzi, K. M. Can too many copies spoil the broth? *Microbial Cell Factories* vol. 12 128 Preprint at https://doi.org/10.1186/1475-2859-12-128 (2013).

103. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLoS ONE* **3**, e3647 (2008).

104. Klein, J. *et al.* Yeast Synthetic Biology Platform Generates Novel Chemical Structures as Scaffolds for Drug Discovery. *ACS Synthetic Biology* **3**, 314–323 (2014).

105. Lau, Y. H., Giessen, T. W., Altenburg, W. J. & Silver, P. A. Prokaryotic nanocompartments form synthetic organelles in a eukaryote. *Nature Communications* **9**, 1311 (2018).

106. Weigand, J. E. & Suess, B. Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. *Nucleic Acids Res* **35**, 4179–85 (2007).

107. Pham, H. L. *et al.* Engineering a riboswitch-based genetic platform for the self-directed evolution of acid-tolerant phenotypes. *Nature Communications* **8**, 411 (2017).

108. Kennedy, A. B., Vowles, J. V., d'Espaux, L. & Smolke, C. D. Protein-responsive ribozyme switches in eukaryotic cells. *Nucleic Acids Research* **42**, 12306–12321 (2014).

109. Cataldo, V. F., Salgado, V., Saa, P. A. & Agosin, E. Genomic integration of unclonable gene expression cassettes in Saccharomyces cerevisiae using rapid cloning-free workflows. *Microbiologyopen* **9**, e978 (2020).

110. Zhou, T. C., Kim, B. G. & Zhong, J. J. Enhanced production of validamycin A in Streptomyces hygroscopicus 5008 by engineering validamycin biosynthetic gene cluster. *Applied Microbiology and Biotechnology* **98**, 7911–7922 (2014).

111. Elliott, K. T., Cuff, L. E. & Neidle, E. L. Copy number change: evolving views on gene amplification. *Future Microbiology* **8**, 887–899 (2013).

112. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **31**, 233–239 (2013).

113. DiCarlo, J., Norville, J., Mali, P., … X. R.-N. acids & 2013, undefined. Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. *academic.oup.com*.

114. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science (1979)* **339**, 823–826 (2013).

115. Tarasava, K., Oh, E. J., Eckert, C. A. & Gill, R. T. CRISPR-Enabled Tools for Engineering Microbial Genomes and Phenotypes. *Biotechnology Journal* vol. 13 Preprint at https://doi.org/10.1002/biot.201700586 (2018).

116. Ronda, C. *et al.* CrEdit: CRISPR mediated multi-loci gene integration in Saccharomyces cerevisiae. *Microbial Cell Factories* **14**, 1–11 (2015).

117. Mitsui, R., Yamada, R. & Ogino, H. CRISPR system in the yeast Saccharomyces cerevisiae and its application in the bioproduction of useful chemicals. *World Journal of Microbiology and Biotechnology* vol. 35 111 Preprint at https://doi.org/10.1007/s11274-019-2688-8 (2019).

118. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology* **13**, 722–736 (2015).

119. Jakočinas, T. *et al.* Multiplex metabolic pathway engineering using CRISPR/Cas9 in Saccharomyces cerevisiae. *Metabolic Engineering* **28**, 213–222 (2015).

120. Verwaal, R., Buiting-Wiessenhaan, N., Dalhuijsen, S. & Roubos, J. A. CRISPR/Cpf1 enables fast and simple genome editing of *Saccharomyces cerevisiae*. *Yeast* **35**, 201–211 (2018).

121. Si, T. *et al.* Automated multiplex genome-scale engineering in yeast. *Nature Communications* **8**, 15187 (2017).

122. Lian, J., Hamedirad, M., Hu, S. & Zhao, H. Combinatorial metabolic engineering using an orthogonal tri-functional CRISPR system. *Nature Communications* **8**, 1–9 (2017).

123. Li, T. *et al.* CRISPR-Cpf1-mediated genome editing and gene regulation in human cells. *Biotechnology Advances* vol. 37 21–27 Preprint at https://doi.org/10.1016/j.biotechadv.2018.10.013 (2019).

124. Nakade, S., Yamamoto, T. & Sakuma, T. Cas9, Cpf1 and C2c1/2/3—What's next? *Bioengineered* vol. 8 265–273 Preprint at https://doi.org/10.1080/21655979.2017.1282018 (2017).

125. Easmin, F. *et al.* gRNA-transient expression system for simplified gRNA delivery in CRISPR/Cas9 genome editing. *Journal of Bioscience and Bioengineering* **128**, 373–378 (2019).

126. Mans, R. *et al.* CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae. *FEMS Yeast Research* **15**, (2015).

127. Horwitz, A. A. *et al.* Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. *Cell Systems* **1**, 88–96 (2015).

128. Hou, S., Qin, Q. & Dai, J. Wicket: A Versatile Tool for the Integration and Optimization of Exogenous Pathways in Saccharomyces cerevisiae. *ACS Synthetic Biology* **7**, 782–788 (2018).

129. Bourgeois, L., Pyne, M. E. & Martin, V. J. J. A Highly Characterized Synthetic Landing Pad System for Precise Multicopy Gene Integration in Yeast. *ACS Synthetic Biology* **7**, 2675–2685 (2018).

130. Shao, Z., Zhao, H. & Zhao, H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res* **37**, e16 (2009).

131. Sasaki, Y., Mitsui, R., Yamada, R. & Ogino, H. Secretory overexpression of the endoglucanase by Saccharomyces cerevisiae via CRISPR-δ-integration and

multiple promoter shuffling. *Enzyme and Microbial Technology* **121**, 17–22 (2019).

132. Gilbert, L. A. *et al.* XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442 (2013).

133. Zalatan, J. G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).

134. Smith, J. D. *et al.* Quantitative CRISPR interference screens in yeast identify chemical-genetic interactions and new rules for guide RNA design. *Genome Biology* **17**, 45 (2016).

135. Ni, J., Zhang, G., Qin, L., Li, J. & Li, C. Simultaneously down-regulation of multiplex branch pathways using CRISPRi and fermentation optimization for enhancing β-amyrin production in Saccharomyces cerevisiae. *Synthetic and Systems Biotechnology* **4**, 79–85 (2019).

136. Raschmanová, H., Weninger, A., Glieder, A., Kovar, K. & Vogl, T. Implementing CRISPR-Cas technologies in conventional and non-conventional yeasts: Current state and future prospects. *Biotechnology Advances* vol. 36 641–665 Preprint at https://doi.org/10.1016/j.biotechadv.2018.01.006 (2018).

137. Chen, X. & Zhang, J. The Genomic Landscape of Position Effects on Protein Expression Level and Noise in Yeast. *Cell Systems* **2**, 347–354 (2016).

138. Babarinde, I. A. & Saitou, N. Genomic Locations of Conserved Noncoding Sequences and Their Proximal Protein-Coding Genes in Mammalian Expression Dynamics. *Mol Biol Evol* **33**, 1807–17 (2016).

139. Sekinger, E. A., Moqtaderi, Z. & Struhl, K. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Molecular Cell* **18**, 735–748 (2005).

140. Lam, F. H., Steger, D. J. & O'Shea, E. K. Chromatin decouples promoter threshold from dynamic range. *Nature* **453**, 246–250 (2008).

141. Eaton, M. L., Galani, K., Kang, S., Bell, S. P. & MacAlpine, D. M. Conserved nucleosome positioning defines replication origins. *Genes and Development* **24**, 748–753 (2010).

142. Aslankoohi, E. *et al.* Nucleosomes affect local transformation efficiency. *Nucleic Acids Res* **40**, 9506–12 (2012).

143. Chen, M., Licon, K., Otsuka, R., Pillus, L. & Ideker, T. Decoupling Epigenetic and Genetic Effects through Systematic Analysis of Gene Position. *Cell Reports* **3**, 128–137 (2013).

144. Yoo, B. C., Yadav, N. S., Orozco, E. M. & Sakai, H. Cas9/gRNA-mediated genome editing of yeast mitochondria and Chlamydomonas chloroplasts. *PeerJ* **2020**, e8362 (2020).

145. Fox, T. D., Sanford, J. C. & Mcmullin, T. W. *Plasmids can stably transform yeast mitochondria lacking endogenous mtDNA (Saccharomyces cerevisiae/oxil/rhol-/high-velocity microprojectile bombardment)*. *Proc. Nati. Acad. Sci. USA* vol. 85 (1988).

146. Almaraz-Delgado, A. L., Flores-Uribe, J., Pérez-España, V. H., Salgado-Manjarrez, E. & Badillo-Corona, J. A. Production of therapeutic proteins in the chloroplast of Chlamydomonas reinhardtii. *AMB Express* **4**, 1–9 (2014).

147. Avalos, J. L., Fink, G. R. & Stephanopoulos, G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nature Biotechnology* **31**, 335–341 (2013).

148. Zimmermann, R., Eyrisch, S., Ahmad, M. & Helms, V. Protein translocation across the ER membrane. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1808**, 912–924 (2011).

149. Van Puyenbroeck, V. & Vermeire, K. Inhibitors of protein translocation across membranes of the secretory pathway: novel antimicrobial and anticancer agents. *Cell Mol Life Sci* **75**, 1541–1558 (2018).

150. Wilkinson, B. M., Tyson, J. R. & Stirling, C. J. Ssh1p Determines the Translocation and Dislocation Capacities of the Yeast Endoplasmic Reticulum. *Developmental Cell* **1**, 401–409 (2001).

151. Rapoport, T. A. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature* **450**, 663–669 (2007).

152. Boisramé, A., Kabani, M., Beckerich, J. M., Hartmann, E. & Gaillardin, C. Interaction of Kar2p and Sls1p is required for efficient co-translational translocation of secreted proteins in the yeast Yarrowia lipolytica. *J Biol Chem* **273**, 30903–8 (1998).

153. de Keyzer, J., Steel, G. J., Hale, S. J., Humphries, D. & Stirling, C. J. Nucleotide Binding by Lhs1p Is Essential for Its Nucleotide Exchange Activity and for Function *in Vivo*. *Journal of Biological Chemistry* **284**, 31564–31571 (2009).

154. Kampinga, H. H. & Craig, E. A. The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nature Reviews Molecular Cell Biology* **11**, 579–592 (2010).

155. Payne, T. *et al.* Transcript-specific translational regulation in the unfolded protein response of Saccharomyces cerevisiae. *FEBS Letters* **582**, 503–509 (2008).

156. Kim, H., Yoo, S. J. & Kang, H. A. Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS Yeast Research* vol. 15 1–16 Preprint at https://doi.org/10.1111/1567-1364.12195 (2015).

157. Van Zyl, J. H. D., Den Haan, R. & Van Zyl, W. H. Overexpression of native Saccharomyces cerevisiae ER-to-Golgi SNARE genes increased heterologous cellulase secretion. *Applied Microbiology and Biotechnology* **100**, 505–518 (2016).

158. Bao, J., Huang, M., Petranovic, D. & Nielsen, J. Moderate Expression of *SEC16* Increases Protein Secretion by Saccharomyces cerevisiae. *Applied and Environmental Microbiology* **83**, (2017).

159. Huang, M. *et al.* Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. *Proceedings of the National Academy of Sciences* **112**, E4689–E4696 (2015).

160. Van Zyl, W. H., Bloom, M. & Viktor, M. J. Engineering yeasts for raw starch conversion. *Applied Microbiology and Biotechnology* vol. 95 1377–1388 Preprint at https://doi.org/10.1007/s00253-012-4248-0 (2012).

161. Sakwa, L., Cripwell, R. A., Rose, S. H. & Viljoen-Bloom, M. Consolidated bioprocessing of raw starch with Saccharomyces cerevisiae strains expressing fungal alpha-amylase and glucoamylase combinations. *FEMS Yeast Research* **18**, (2018).

162. Cripwell, R. A., Rose, S. H., Favaro, L. & Van Zyl, W. H. Construction of industrial Saccharomyces cerevisiae strains for the efficient consolidated bioprocessing of raw starch. *Biotechnology for Biofuels* **12**, 201 (2019).

163. Favaro, L., Jansen, T. & van Zyl, W. H. Exploring industrial and natural Saccharomyces cerevisiae strains for the bio-based economy from biomass: the case of bioethanol. *Critical Reviews in Biotechnology* vol. 39 800–816 Preprint at https://doi.org/10.1080/07388551.2019.1619157 (2019).

164. Corte, L. *et al.* Delta-Integration of Single Gene Shapes the Whole Metabolomic Short-Term Response to Ethanol of Recombinant Saccharomyces cerevisiae Strains. *Metabolites* **10**, 140 (2020).

165. Marsalek, L., Puxbaum, V., Buchetics, M., Mattanovich, D. & Gasser, B. Disruption of vacuolar protein sorting components of the HOPS complex leads

to enhanced secretion of recombinant proteins in Pichia pastoris. *Microbial Cell Factories* **18**, 119 (2019).

166. Marsalek, L. *et al.* Disruption of genes involved in CORVET complex leads to enhanced secretion of heterologous carboxylesterase only in protease deficient Pichia pastoris. *Biotechnology Journal* **12**, (2017).

167. Galanie, S., Thodey, K., Trenchard, I. J., Filsinger Interrante, M. & Smolke, C. D. Complete biosynthesis of opioids in yeast. *Science (1979)* **349**, 1095–1100 (2015).

168. Reider Apel, A. *et al.* A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **45**, 496–508 (2017).

169. Xu, Y. *et al.* Diversity-oriented combinatorial biosynthesis of benzenediol lactone scaffolds by subunit shuffling of fungal polyketide synthases. *Proceedings of the National Academy of Sciences* **111**, 12354–12359 (2014).

170. Nakamura, Y., Ishii, J. & Kondo, A. Construction of a yeast-based signaling biosensor for human angiotensin II type 1 receptor via functional coupling between Asn295-mutated receptor and Gpa1/G_{i3} chimeric Gα. *Biotechnology and Bioengineering* **111**, 2220–2228 (2014).

171. Quinby, G. E., Dean, J. P. & Deschenes, R. J. Expression of MFA1 and STE6 is sufficient for mating type-independent secretion of yeast a-factor, but not mating competence. *Curr Genet* **35**, 1–7 (1999).

172. Xie, Y., Han, X., biology, Y. M.-C. protocols in molecular & 2018, undefined. An Effective Recombinant Protein Expression and Purification System in Saccharomyces cerevisiae. *Wiley Online Library*.

173. Charron, M., Dubin, R., cellular, C. M.-M. and & 1986, undefined. Structural and functional analysis of the MAL1 locus of Saccharomyces cerevisiae. *Am Soc Microbiol*.

174. Gagiano, M., Bauer, F., research, I. P.-F. yeast & 2002, undefined. The sensing of nutritional status and the relationship to filamentous growth in Saccharomyces cerevisiae. *academic.oup.com*.

175. Lin-Cereghino, G. P. *et al.* The effect of α-mating factor secretion signal mutations on recombinant protein expression in Pichia pastoris. *Gene* **519**, 311–317 (2013).

176. Bae, J. H. *et al.* An Efficient Genome-Wide Fusion Partner Screening System for Secretion of Recombinant Proteins in Yeast. *Scientific Reports* **5**, (2015).

177. Smeekens, J. M., Xiao, H. & Wu, R. Global Analysis of Secreted Proteins and Glycoproteins in *Saccharomyces cerevisiae*. *Journal of Proteome Research* **16**, 1039–1049 (2017).

178. Makridakis, M. & Vlahou, A. Secretome proteomics for discovery of cancer biomarkers. *Journal of Proteomics* **73**, 2291–2305 (2010).

179. Pavlou, M. P. & Diamandis, E. P. The cancer cell secretome: A good source for discovering biomarkers? *Journal of Proteomics* **73**, 1896–1906 (2010).

180. Pandey, A. *et al.* Analysis of receptor signaling pathways by mass spectrometry: identification of vav-2 as a substrate of the epidermal and platelet-derived growth factor receptors. *Proc Natl Acad Sci U S A* **97**, 179–84 (2000).

181. Zhu, Z. & Desaire, H. Carbohydrates on Proteins: Site-Specific Glycosylation Analysis by Mass Spectrometry. *Annual Review of Analytical Chemistry* **8**, 463–483 (2015).

182. Spiro, R. G. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* **12**, 43R-56R (2002).

183. Huang, M., Wang, G., Qin, J., Petranovic, D. & Nielsen, J. Engineering the protein secretory pathway of Saccharomyces cerevisiae enables improved protein production. *Proc Natl Acad Sci U S A* **115**, E11025–E11032 (2018).

184. Harmsen, M. M., Bruyne, M. I., Raué, H. A. & Maat, J. Overexpression of binding protein and disruption of the PMR1 gene synergistically stimulate secretion of bovine prochymosin but not plant thaumatin in yeast. *Appl Microbiol Biotechnol* **46**, 365–70 (1996).

185. de Ruijter, J. C., Koskela, E. V. & Frey, A. D. Enhancing antibody folding and secretion by tailoring the Saccharomyces cerevisiae endoplasmic reticulum. *Microbial Cell Factories* **15**, 87 (2016).

186. Tang, H. *et al.* Engineering protein folding and translocation improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering* **112**, 1872–1882 (2015).

187. Huang, M. *et al.* Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. *Proceedings of the National Academy of Sciences* **112**, E4689–E4696 (2015).

188. Gouka, R. J., Punt, P. J., Hessing, J. G., van den Hondel, C. A. & Archer, D. B. Analysis of heterologous protein production in defined recombinant Aspergillus awamori strains. *Appl Environ Microbiol* **62**, 1951–7 (1996).

189. Inaba, K. *et al.* Crystal structures of human Ero1α reveal the mechanisms of regulated and targeted oxidation of PDI. *The EMBO Journal* **29**, 3330–3343 (2010).

190. Chen, F., Hu, M., Jiang, X., Tao, Y. & Huang, J. Enhancement of Coprinus cinereus peroxidase in Pichia pastoris by co-expression chaperone PDI and Ero1. *Sheng Wu Gong Cheng Xue Bao* **31**, 1682–9 (2015).

191. Freedman, R. B. *et al.* 'Something in the way she moves': The functional significance of flexibility in the multiple roles of protein disulfide isomerase (PDI). *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1865**, 1383–1394 (2017).

192. Römer, R. A. *et al.* The flexibility and dynamics of protein disulfide isomerase. *Proteins: Structure, Function, and Bioinformatics* **84**, 1776–1785 (2016).

193. Tang, X. & Wang, Y. Pds1/Esp1-dependent and -independent sister chromatid separation in mutants defective for protein phosphatase 2A. *Proceedings of the National Academy of Sciences* **103**, 16290–16295 (2006).

194. Sevier, C. S. & Kaiser, C. A. Ero1 and redox homeostasis in the endoplasmic reticulum. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1783**, 549–556 (2008).

195. Margittai, É. *et al.* Production of H_2O_2 in the Endoplasmic Reticulum Promotes *In Vivo* Disulfide Bond Formation. *Antioxidants & Redox Signaling* **16**, 1088–1099 (2012).

196. Kim, S., Sideris, D. P., Sevier, C. S. & Kaiser, C. A. Balanced Ero1 activation and inactivation establishes ER redox homeostasis. *The Journal of Cell Biology* **196**, 713–725 (2012).

197. Two endoplasmic reticulum PDI peroxidases increase the efficiency of the use of peroxide during disulfide bond formation. | BioGRID. https://thebiogrid.org/187876/publication/two-endoplasmic-reticulum-pdi-peroxidases-increase-the-efficiency-of-the-use-of-peroxide-during-disulfide-bond-formation.html.

198. Smith, J. D., Tang, B. C. & Robinson, A. S. Protein disulfide isomerase, but not binding protein, overexpression enhances secretion of a non-disulfide-bonded protein in yeast. *Biotechnology and Bioengineering* **85**, 340–350 (2004).

199. Robinson, A. S., Hines, V. & Wittrup, K. D. Protein Disulfide Isomerase Overexpression Increases Secretion of Foreign Proteins in Saccharomyces cerevisiae. *Bio/Technology 1994 12:4* **12**, 381–384 (1994).

200. Powers, S. L. & Robinson, A. S. PDI Improves Secretion of Redox-Inactive β-Glucosidase. *Biotechnology Progress* **23**, 364–369 (2007).

201. Kulp, M. S., Frickel, E. M., Ellgaard, L. & Weissman, J. S. Domain architecture of protein-disulfide isomerase facilitates its dual role as an oxidase and an isomerase in Ero1p-mediated disulfide formation. *J Biol Chem* **281**, 876–884 (2006).

202. Appenzeller-Herzog, C. & Ellgaard, L. In vivo reduction-oxidation state of protein disulfide isomerase: the two active sites independently occur in the reduced and oxidized forms. *Antioxid Redox Signal* **10**, 55–64 (2008).

203. Čiplys, E., Sasnauskas, K. & Slibinskas, R. Overexpression of human calnexin in yeast improves measles surface glycoprotein solubility. *FEMS Yeast Research* **11**, 514–523 (2011).

204. Valkonen, M., Penttilä, M. & Saloheimo, M. Effects of inactivation and constitutive expression of the unfolded- protein response pathway on protein production in the yeast Saccharomyces cerevisiae. *Appl Environ Microbiol* **69**, 2065–72 (2003).

205. Gregoire, S., Irwin, J. & Kwon, I. Techniques for Monitoring Protein Misfolding and Aggregation in Vitro and in Living Cells. *Korean J Chem Eng* **29**, 693–702 (2012).

206. Cabantous, S., Rogers, Y., Terwilliger, T. C. & Waldo, G. S. New molecular reporters for rapid protein folding assays. *PLoS One* **3**, e2387 (2008).

207. Malhotra, J. D. *et al.* Antioxidants reduce endoplasmic reticulum stress and improve protein secretion. *Proceedings of the National Academy of Sciences* **105**, 18525–18530 (2008).

208. Nocon, J. *et al.* Increasing pentose phosphate pathway flux enhances recombinant protein production in Pichia pastoris. *Applied Microbiology and Biotechnology* **100**, 5955–5963 (2016).

209. Malhotra, J. D. & Kaufman, R. J. Endoplasmic reticulum stress and oxidative stress: a vicious cycle or a double-edged sword? *Antioxid Redox Signal* **9**, 2277–2293 (2007).

210. Buday, K. & Conrad, M. Emerging roles for non-selenium containing ER-resident glutathione peroxidases in cell signaling and disease. *Biol Chem* **402**, 271–287 (2020).

211. Navarro-Yepes, J. *et al.* Oxidative stress, redox signaling, and autophagy: cell death versus survival. *Antioxid Redox Signal* **21**, 66–85 (2014).

212. Poli, G., Leonarduzzi, G., Biasi, F. & Chiarpotto, E. Oxidative stress and cell signalling. *Curr Med Chem* **11**, 1163–1182 (2004).

213. Mills, G. C. Hemoglobin Catabolism I. Glutathione Peroxidase, an erythrocyte enxyme which protects hemoglobin from oxidative breakdown. (1957) doi:10.1016/S0021-9258(18)70608-X.

214. Bosello-Travain, V. *et al.* Protein disulfide isomerase and glutathione are alternative substrates in the one Cys catalytic cycle of glutathione peroxidase 7. *Biochim Biophys Acta* **1830**, 3846–3857 (2013).

215. Tosatto, S. C. E. *et al.* The catalytic site of glutathione peroxidases. *Antioxid Redox Signal* **10**, 1515–1525 (2008).

216. Ramming, T., Hansen, H. G., Nagata, K., Ellgaard, L. & Appenzeller-Herzog, C. GPx8 peroxidase prevents leakage of H2O2 from the endoplasmic reticulum. *Free Radical Biology and Medicine* **70**, 106–116 (2014).

217. Wang, L., Zhang, L., Niu, Y., Sitia, R. & Wang, C. C. Glutathione Peroxidase 7 Utilizes Hydrogen Peroxide Generated by Ero1α to Promote Oxidative Protein Folding. *Antioxidants & Redox Signaling* **20**, 545 (2014).

218. Wei, P. C. *et al.* Loss of the oxidative stress sensor NPGPx compromises GRP78 chaperone activity and induces systemic disease. *Mol Cell* **48**, 747–759 (2012).

219. Schröder, M. & Kaufman, R. J. THE MAMMALIAN UNFOLDED PROTEIN RESPONSE. *Annual Review of Biochemistry* **74**, 739–789 (2005).

220. Corbett, E. F. *et al.* Ca^{2+} Regulation of Interactions between Endoplasmic Reticulum Chaperones. *Journal of Biological Chemistry* **274**, 6203–6211 (1999).

221. Tomlin, G. C., Wixon, J. L., Bolotin-Fukuhara, M. & Oliver, S. G. A new family of yeast vectors and S288C-derived strains for the systematic analysis of gene function. *Yeast* **18**, 563–575 (2001).

222. Jossé, L., Mark Smales, C. & Tuite, M. F. Transient expression of human TorsinA enhances secretion of two functionally distinct proteins in cultured chinese hamster ovary (CHO) cells. *Biotechnology and Bioengineering* **105**, 556–566 (2010).

223. Chu, D. *et al.* Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J* **33**, 21–34 (2014).

224. Han, G. S., O'Hara, L., Carman, G. M. & Siniossoglou, S. An Unconventional Diacylglycerol Kinase That Regulates Phospholipid Synthesis and Nuclear Membrane Growth. *The Journal of Biological Chemistry* **283**, 20433 (2008).

225. Laughery, M. F. *et al.* New vectors for simple and streamlined CRISPR-Cas9 genome editing in Saccharomyces cerevisiae. *Yeast* **32**, 711–720 (2015).

226. Koontz, L. TCA Precipitation. in *Methods in enzymology* vol. 541 3–10 (2014).

227. Dong, W.-H., Wang, T.-Y., Wang, F. & Zhang, J.-H. Simple, Time-Saving Dye Staining of Proteins for Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis Using Coomassie Blue. *PLoS ONE* **6**, e22394 (2011).

228. Wu, D. & Yotnda, P. Production and Detection of Reactive Oxygen Species (ROS) in Cancers. *Journal of Visualized Experiments : JoVE* (2011) doi:10.3791/3357.

229. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research* **46**, W537–W544 (2018).

230. Blankenberg, D. *et al.* Dissemination of scientific software with Galaxy ToolShed. *Genome Biology* **15**, 1–3 (2014).

231. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology 2019 37:8* **37**, 907–915 (2019).

232. Cherry, J. M. *et al.* Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res* **40**, D700-5 (2012).

233. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–50 (2005).

234. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genetics* **34**, 267–273 (2003).

235. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13**, 2498–2504 (2003).

236. Errede, B. & Ammerer, G. STE12, a protein involved in cell-type-specific transcription and signal transduction in yeast, is part of protein-DNA complexes. *Genes Dev* **3**, 1349–61 (1989).

237. Gavrias, V., Andrianopoulos, A., Gimeno, C. J. & Timberlake, W. E. Saccharomyces cerevisiae TEC1 is required for pseudohyphal growth. *Mol Microbiol* **19**, 1255–63 (1996).

238. Pi, H., Chien, C. T. & Fields, S. Transcriptional activation upon pheromone stimulation mediated by a small domain of Saccharomyces cerevisiae Ste12p. *Mol Cell Biol* **17**, 6410–8 (1997).

239. Lo, W. S., Raitses, E. I. & Dranginis, A. M. Development of pseudohyphae by embedded haploid and diploid yeast. *Curr Genet* **32**, 197–202 (1997).

240. Yuan, Y. L. & Fields, S. Properties of the DNA-binding domain of the Saccharomyces cerevisiae STE12 protein. *Mol Cell Biol* **11**, 5910–8 (1991).

241. Cook, J. G., Bardwell, L., Kron, S. J. & Thorner, J. Two novel targets of the MAP kinase Kss1 are negative regulators of invasive growth in the yeast Saccharomyces cerevisiae. *Genes Dev* **10**, 2831–48 (1996).

242. Olson, K. A. *et al.* Two regulators of Ste12p inhibit pheromone-responsive transcription by separate mechanisms. *Mol Cell Biol* **20**, 4199–209 (2000).

243. Bardwell, L. *et al.* Repression of yeast Ste12 transcription factor by direct binding of unphosphorylated Kss1 MAPK and its regulation by the Ste7 MEK. *Genes Dev* **12**, 2887–98 (1998).

244. Wong Sak Hoi, J. & Dumas, B. Ste12 and Ste12-Like Proteins, Fungal Transcription Factors Regulating Development and Pathogenicity. *Eukaryotic Cell* **9**, 480–485 (2010).

245. Bao, M. Z., Schwartz, M. A., Cantin, G. T., Yates, J. R. & Madhani, H. D. Pheromone-Dependent Destruction of the Tec1 Transcription Factor Is Required for MAP Kinase Signaling Specificity in Yeast. *Cell* **119**, 991–1000 (2004).

246. Chang, Y. C. *et al.* Regulatory roles for the homeodomain and C2H2 zinc finger regions of Cryptococcus neoformans Ste12αp. *Molecular Microbiology* **53**, 1385–1396 (2004).

247. Park, G., Bruno, K. S., Staiger, C. J., Talbot, N. J. & Xu, J.-R. Independent genetic mechanisms mediate turgor generation and penetration peg formation during plant infection in the rice blast fungus. *Molecular Microbiology* **53**, 1695–1707 (2004).

248. Yoshida, S. *et al.* A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science (1979)* **351**, 1196–1199 (2016).

249. Chu, D. *et al.* Translation elongation can control translation initiation on eukaryotic mRNAs. *EMBO J* **33**, 21–34 (2014).

250. Lorenz, W. W., McCann, R. O., Longiaru, M. & Cormier, M. J. Isolation and expression of a cDNA encoding Renilla reniformis luciferase. *Proceedings of the National Academy of Sciences* **88**, 4438–4442 (1991).

251. Verhaegent, M. & Christopoulos, T. K. Recombinant Gaussia Luciferase. Overexpression, Purification, and Analytical Application of a Bioluminescent Reporter for DNA Hybridization. *Analytical Chemistry* **74**, 4378–4385 (2002).

252. Tannous, B. A., Kim, D.-E., Fernandez, J. L., Weissleder, R. & Breakefield, X. O. Codon-optimized Gaussia luciferase cDNA for mammalian gene expression in culture and in vivo. *Mol Ther* **11**, 435–43 (2005).

253. Szent-Gyorgyi, C., Ballou, B. T., Dagnal, E. & Bryan, B. Cloning and characterization of new bioluminescent proteins. *https://doi.org/10.1117/12.351015* **3600**, 4–11 (1999).

254. Curwin, A. J., von Blume, J. & Malhotra, V. Cofilin-mediated sorting and export of specific cargo from the Golgi apparatus in yeast. *Molecular Biology of the Cell* **23**, 2327–2338 (2012).

255. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research* **46**, W537–W544 (2018).

256. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545–15550 (2005).

257. Rispail, N. & Pietro, A. di. The homeodomain transcription factor Ste12. *http://www.tandfonline.com/action/authorSubmission?journalCode=kcib20&page=instructions* **3**, 327–332 (2010).

258. Chan, C. Y. *et al.* Purinosome formation as a function of the cell cycle. *Proceedings of the National Academy of Sciences* **112**, 1368–1373 (2015).

259. Yin, J. *et al.* Potential Mechanisms Connecting Purine Metabolism and Cancer Therapy. *Front Immunol* **9**, 1697 (2018).

260. Moffatt, B. A. & Ashihara, H. Purine and pyrimidine nucleotide synthesis and metabolism. *Arabidopsis Book* **1**, e0018 (2002).

261. Samant, S. *et al.* Nucleotide Biosynthesis Is Critical for Growth of Bacteria in Human Blood. *PLoS Pathogens* **4**, e37 (2008).

262. Control of Macromolecular Synthesis. *The Yale Journal of Biology and Medicine* **39**, 136 (1966).

263. Lempiäinen, H. & Shore, D. Growth control and ribosome biogenesis. *Curr Opin Cell Biol* **21**, 855–863 (2009).

264. Vadivel Gnanasundram, S., Fåhraeus, R., Vadivel Gnanasundram, S. & Fåhraeus, R. Translation Stress Regulates Ribosome Synthesis and Cell Proliferation. *International Journal of Molecular Sciences* **19**, 3757 (2018).

265. Shah, P., Ding, Y., Niemczyk, M., Kudla, G. & Plotkin, J. B. Rate-limiting steps in yeast protein translation. *Cell* **153**, 1589–601 (2013).

266. Ellgaard, L. & Ruddock, L. W. The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO Rep* **6**, 28–32 (2005).

267. Nørgaard, P. *et al.* Functional Differences in Yeast Protein Disulfide Isomerases. *The Journal of Cell Biology* **152**, 553 (2001).

268. Martínez, J. L., Meza, E., Petranovic, D. & Nielsen, J. The impact of respiration and oxidative stress response on recombinant α-amylase production by Saccharomyces cerevisiae. *Metabolic Engineering Communications* **3**, 205–210 (2016).

269. Mehmeti, I., Lortz, S., Avezov, E., Jörns, A. & Lenzen, S. ER-resident antioxidative GPx7 and GPx8 enzyme isoforms protect insulin-secreting INS-

1E β-cells against lipotoxicity by improving the ER antioxidative capacity. *Free Radic Biol Med* **112**, 121–130 (2017).

270. Barrero, J. J., Casler, J. C., Valero, F., Ferrer, P. & Glick, B. S. An improved secretion signal enhances the secretion of model proteins from Pichia pastoris. *Microbial Cell Factories* **17**, 1–13 (2018).

271. Wu, N. *et al.* Solution structure of Gaussia Luciferase with five disulfide bonds and identification of a putative coelenterazine binding cavity by heteronuclear NMR. *Scientific Reports 2020 10:1* **10**, 1–11 (2020).

272. Young, B. P., Craven, R. A., Reid, P. J., Willer, M. & Stirling, C. J. Sec63p and Kar2p are recquired for the translocation of SRP-dependent precursors into the yeast endoplasmic reticulum in vivo. *EMBO Journal* **20**, 262–271 (2001).

273. Tarassov, K. *et al.* An in vivo map of the yeast protein interactome. *Science (1979)* **320**, 1465–1470 (2008).

274. Chen, Y. I., Wei, P. C., Hsu, J. L., Su, F. Y. & Lee, W. H. NPGPx (GPx7): a novel oxidative stress sensor/transmitter with multiple roles in redox homeostasis. *American Journal of Translational Research* **8**, 1626 (2016).

275. Gomes, A. M. V., Carmo, T. S., Carvalho, L. S., Bahia, F. M. & Parachin, N. S. Comparison of Yeasts as Hosts for Recombinant Protein Production. *Microorganisms* **6**, (2018).

276. Wang, L., Zhang, L., Niu, Y., Sitia, R. & Wang, C. C. Glutathione peroxidase 7 utilizes hydrogen peroxide generated by Ero1α to promote oxidative protein folding. *Antioxid Redox Signal* **20**, 545–556 (2014).

277. Joshi, S. & Yu, D. Immunofluorescence. *Basic Science Methods for Clinical Researchers* 135–150 (2017) doi:10.1016/B978-0-12-803077-6.00008-4.

278. Hackstadt, T. Steric hindrance of antibody binding to surface proteins of Coxiella burnetti by phase I lipopolysaccharide. *Infection and Immunity* **56**, 802 (1988).

279. Kanemura, S. *et al.* Characterization of the endoplasmic reticulum–resident peroxidases GPx7 and GPx8 shows the higher oxidative activity of GPx7 and its linkage to oxidative protein folding. *Journal of Biological Chemistry* **295**, 12772–12785 (2020).

280. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337 (2009).

281. Draper, N. R. & Pukelsheim, F. An overview of design of experiments. *Statistical Papers 1996 37:1* **37**, 1–32 (1996).

282. Fatemeh Davami, F. E. L. N. F. B. and F. M. Effects of Peptone Supplementation in Different Culture Media on Growth, Metabolic Pathway and Productivity of CHO DG44 Cells; a New Insight into Amino Acid Profiles. *Iran Biomed J* 194–205 (2015).

283. Pfeifer, N. *et al.* Selection of the in vitro culture media influences mRNA expression of Hedgehog genes, Il-6, and important genes regarding reactive oxygen species in single murine preimplantation embryos. *The Scientific World Journal* **2012**, (2012).

284. Chu, D., Barnes, D. J. & von der Haar, T. The role of tRNA and ribosome competition in coupling the expression of different mRNAs in Saccharomyces cerevisiae. *Nucleic Acids Res* **39**, 6705–6714 (2011).

285. Davami, F., Eghbalpour, F., Nematollahi, L., Barkhordari, F. & Mahboudi, F. Effects of peptone supplementation in different culture media on growth, metabolic pathway and productivity of CHO DG44 cells; a new insight into amino acid profiles. *Iranian Biomedical Journal* **19**, 194–205 (2015).

286. Ashe, M. P. & Bill, R. M. Mapping the yeast host cell response to recombinant membrane protein production: Relieving the biological bottlenecks. *Biotechnology Journal* vol. 6 707–714 Preprint at https://doi.org/10.1002/biot.201000333 (2011).

287. Zhang, Z., Moo-Young, M. & Chisti, Y. *PLASMID STABILITY IN RECOMBINANT SA CCHAROMYCES CEREVISIAE*. *Biotechnology Advances* vol. 14 (1996).

288. Frand, A. R. & Kaiser, C. A. The ERO1 gene of yeast is required for oxidation of protein dithiols in the endoplasmic reticulum. *Mol Cell* **1**, 161–170 (1998).

289. Farquhar, R. *et al.* Protein disulfide isomerase is essential for viability in Saccharomyces cerevisiae. *Gene* **108**, 81–89 (1991).

290. Tian, G., Xiang, S., Noiva, R., Lennarz, W. J. & Schindelin, H. The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell* **124**, 61–73 (2006).

291. Gross, E., Kastner, D. B., Kaiser, C. A. & Fass, D. Structure of Ero1p, Source of Disulfide Bonds for Oxidative Protein Folding in the Cell. *Cell* **117**, 601–610 (2004).

292. Nørgaard, P. & Winther, J. R. Mutation of yeast Eug1p CXXS active sites to CXXC results in a dramatic increase in protein disulphide isomerase activity. *Biochemical Journal* **358**, 269–274 (2001).

293. Tachibana, C. & Stevens, T. H. The yeast EUG1 gene encodes an endoplasmic reticulum protein that is functionally related to protein disulfide isomerase. *Molecular and Cellular Biology* **12**, 4601 (1992).

294. Frand, A. R. & Kaiser, C. A. The ERO1 Gene of Yeast Is Required for Oxidation of Protein Dithiols in the Endoplasmic Reticulum. *Molecular Cell* **1**, 161–170 (1998).

295. Zito, E. *et al.* Oxidative protein folding by an endoplasmic reticulum-localized peroxiredoxin. *Mol Cell* **40**, 787–797 (2010).

296. Beal, D. M. *et al.* Quantitative Analyses of the Yeast Oxidative Protein Folding Pathway In Vitro and In Vivo. *Antioxidants & Redox Signaling* **31**, 261 (2019).

297. Martínez, J. L., Meza, E., Petranovic, D. & Nielsen, J. The impact of respiration and oxidative stress response on recombinant α-amylase production by Saccharomyces cerevisiae. *Metabolic Engineering Communications* **3**, 205–210 (2016).

298. Hackel, B. J., Huang, D., Bubolz, J. C., Wang, X. X. & Shusta, E. v. Production of soluble and active transferrin receptor-targeting single-chain antibody using Saccharomyces cerevisiae. *Pharmaceutical Research* **23**, 790–797 (2006).

299. SCHULTZ, L. D. *et al.* Using molecular genetics to improve the production of recombinant proteins by the yeast Saccharomyces cerevisiae. *Ann N Y Acad Sci* **721**, 148–157 (1994).

300. Madhavan, A. *et al.* Customized yeast cell factories for biopharmaceuticals: from cell engineering to process scale up. *Microbial Cell Factories 2021 20:1* **20**, 1–17 (2021).

301. Vernikos, G., Medini, D., Riley, D. R. & Tettelin, H. Ten years of pan-genome analyses. doi:10.1016/j.mib.2014.11.016.

302. Li, G., Ji, B. & Nielsen, J. The pan-genome of Saccharomyces cerevisiae. *FEMS Yeast Res* **19**, (2019).

303. de Ruijter, J. C., Koskela, E. v. & Frey, A. D. Enhancing antibody folding and secretion by tailoring the Saccharomyces cerevisiae endoplasmic reticulum. *Microbial Cell Factories* **15**, 1–18 (2016).

304. Marks-Anglin, A. & Chen, Y. A historical review of publication bias. *Research Synthesis Methods* **11**, 725–742 (2020).

305. Rao, K. J., Kim, C. H. & Rhee, S. K. Statistical optimization of medium for the production of recombinant hirudin from Saccharomyces cerevisiae using response surface methodology. *Process Biochemistry* **35**, 639–647 (2000).

306. Wittrup, K. D. & Benig, V. Optimization of amino acid supplements for heterologous protein secretion in Saccharomyces cerevisiae. *Biotechnology Techniques 1994 8:3* **8**, 161–166 (1994).

307. Turner, B. G., Avgerinos, G. C., Melnick, L. M. & Moir, D. T. Optimization of pro-urokinase secretion from recombinant Saccharomyces cerevisiae. *Biotechnology and Bioengineering* **37**, 869–875 (1991).

308. Liu, Z., Hou, J., Martínez, J. L., Petranovic, D. & Nielsen, J. Correlation of cell growth and heterologous protein production by Saccharomyces cerevisiae. *Applied Microbiology and Biotechnology* **97**, 8955–8962 (2013).

309. Hou, J., Österlund, T., Liu, Z., Petranovic, D. & Nielsen, J. Heat shock response improves heterologous protein secretion in Saccharomyces cerevisiae. *Applied Microbiology and Biotechnology* **97**, 3559–3568 (2013).

310. Kroukamp, H., den Haan, R., van Wyk, N. & van Zyl, W. H. Overexpression of native PSE1 and SOD1 in Saccharomyces cerevisiae improved heterologous cellulase secretion. *Applied Energy* **102**, 150–156 (2013).

311. Sharp, N. P., Sandell, L., James, C. G. & Otto, S. P. The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. *Proc Natl Acad Sci U S A* **115**, E5046–E5055 (2018).

312. Lynch, M. *et al.* A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci U S A* **105**, 9272–9277 (2008).

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

314. Calado, C. R. C., Taipa, M. Â., Cabral, J. M. S. & Fonseca, L. P. Optimisation of culture conditions and characterisation of cutinase produced by recombinant Saccharomyces cerevisiae. *Enzyme and Microbial Technology* **31**, 161–170 (2002).

315. Nurmianto, L., Greenfield, P. F. & Lee, P. L. Optimal feeding policy for recombinant protein production via the yeast saccharomyces cerevisiae in fed-batch culture. *Process Biochemistry* **29**, 55–68 (1994).