Exploring the dual roles of tRNAs in *Komagataella phaffii* production hosts

Nisreen Chahid
PhD

University of York
Department of Biology

October 2024

Abstract

tRNAs play an essential role in protein synthesis by decoding mRNA and delivering amino acids to ribosomes during translation. Beyond this, tRNA genes (tDNAs) contribute to genome organisation by acting as chromatin barriers by maintaining the boundaries between active and repressive chromatin using a property known as barrier activity. Exploiting these dual roles of tRNAs could be a promising approach to enhancing recombinant protein production. Here, we explore tRNA-based strategies in the yeast *Komagataella phaffii*, an emerging production platform.

Our findings show that *K. phaffii* tDNAs exhibit barrier activity in the baker's yeast, *S. cerevisiae*, outperforming native barriers in a mating screen. Notably, the *AOX1* locus, a popular integration site in *K. phaffii* expression systems, lacks protection from active barrier tDNAs, suggesting potential susceptibility to chromatin-based repression. Introducing a *K. phaffii* tDNA expectedly increased chromatin accessibility at this locus but paradoxically suppressed transgene transcription, highlighting complex roles of tDNAs in regulating expression.

Assessing downstream applications of tRNA-based interventions, this thesis identifies an anticodon missing from the *K. phaffii* repertoire, for which the cognate codon is highly prevalent amongst endogenous proteins, thereby implying a significant demand on wobble decoding in *K. phaffii* cells. Wobble decoding, a mechanism that allows a single tRNA to recognise multiple codons through flexible base-pairing at the third codon position, is found not to be a bottleneck in the translation of recombinant proteins, nor does it impede translation of endogenous proteins. Yet, introduction of an exogenous *S. cerevisiae* tRNA gene carrying the missing anticodon was proven to have potential in being an effective intervention for recombinant protein production, suggesting a bottleneck in tRNA supply.

In summary, this thesis highlights the potential of leveraging the diverse and complex roles of tRNAs to enhance recombinant protein production.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This work was conducted as part of a collaboration between the Industrial Biotechnology Innovation Centre (IBioIC), Fujifilm Diosynth Biotechnologies UK (FDBK) and the University of York. Thank you for the providing the resources and funding that were essential for the successful completion of this research.

Thanks to my supervisors for their support and contributions to this project. Another big thanks to Chris MacDonald for being a kind and supportive TAP member.

Big thanks to everyone in the J0 Cell Biology lab, past and present, for making lab life a little bit more fun; especially Josie Ayre for big laughs and therapeutic coffee breaks, Sanjana for always making me smile, Savvas for the motivational speeches, Katherine for being an amazing role model, Becky, Sienna and Jodie for being by my side from the beginning and Saffron for bringing energy to that final push.

To my PhD family, my little big sister, Javeria Mehboob, and my lab mum, Ginevra Camboni. I love you both so much, thank you for your unconditional love and care.

To my dad and brother, Said Chahid and Faysal Chahid. Thank you for always being there for me. You both are what drives me to work harder and build up our lives together.

To my little family at home; my love and my person, Jamal Nawaz, and my girls, Nori and Nova. Thank you for being my source of comfort, joy and healing. I can't wait to finally completely enjoy the life we've been building together.

Author's declaration

I declare that this thesis is a presentation of original work and I am the sole author.

This work has not previously been presented for a degree or other qualification at this University or elsewhere.

All sources are acknowledged as references.

Publications derived from this work include:

Sizer RE, Chahid N, Butterfield SP, Donze D, Bryant NJ, White RJ. TFIIIC-based chromatin insulators through eukaryotic evolution. Gene. 2022 Aug 15;835:146533. doi: 10.1016/j.gene.2022.146533. Epub 2022 May 24. PMID: 35623477.

1. Introduction

Despite diverging in evolution millions of years ago, cells across all three domains of life; archaea, bacteria and eukarya, share the same remarkable ability to produce proteins (1). The fundamental process of encoding the information required to make these proteins into polymers, which are then decoded through a series of chemical reactions, is simply described in the Central Dogma of Molecular Biology (2) (Figure 1.1). In short, the recipe for a single protein is stored in genes, comprised of DNA. These genes are then transcribed into RNA which, in turn, is translated into polypeptide chains composed of amino acids. Polypeptides are then folded and modified into functional proteins. It is this conserved phenomenon that has empowered scientists, through the feat of genetic engineering, to manipulate cells to produce proteins derived from entirely foreign organisms.

Over decades of scientific research, we have studied and determined the functional activity of millions of proteins across different species and characterised their contributions to survival. Many have garnered attention for their applications, including in the pharmaceutical industry for treating human diseases (3). Others are vital for the chemical industry, to catalyse chemical reactions which convert raw materials (oil, natural gas, air, water, metals, and minerals) into products and materials essential for everyday life as we know it (4). Extracting these proteins from their native source and producing them in another context classifies them as recombinant proteins. In order to use these proteins, we have developed the ability to exploit the highly conserved protein production process and engineer host cells to mass-produce proteins of interest, known as recombinant protein production (RPP).

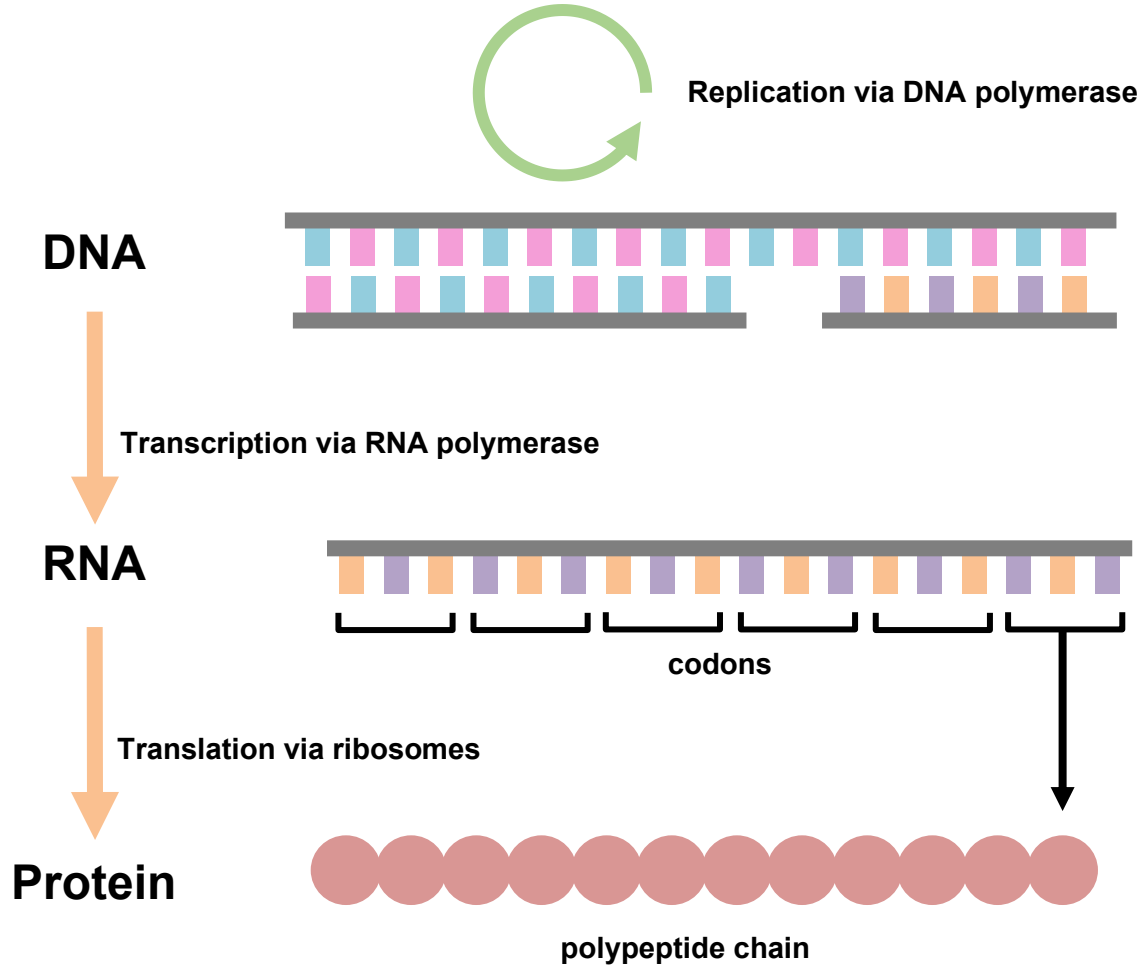


Figure 1.1: Central Dogma of Molecular Biology.

Schematic illustrating the flow of genetic information during the production of a protein. DNA is passed to its progeny during cell division through DNA replication, which is conducted by DNA polymerases. Genes encoded within DNA are transcribed via RNA polymerases. The resulting messenger RNA (mRNA) is then exported from the nucleus to the cytoplasm, where it is translated by ribosomes. Ribosomes can be either free in the cytosol or bound to the endoplasmic reticulum (ER), depending on the destination of the protein being synthesized. Every set of three nucleotides within mRNA (codons), is read using transfer RNAs (tRNAs) and assigned the appropriate amino acid. The final polypeptide chain is assembled from these amino acids.

Contract development and manufacturing organisations (CDMOs) are companies devoted to developing cell factories optimised for producing proteins at high quality and high yields. Their processes can be divided into distinct stages: upstream and downstream processing. Upstream processing concerns itself with the engineering, selection and cultivation of production host cells. Downstream processes focus on harvesting the protein and purifying, concentrating and formulating the final product. While these stages are distinct from one another in their goals, they both require optimisation to achieve high yields, purity, and quality of the recombinant protein.

The work described in this thesis is the product of a collaboration with the University of York, the Industrial Biotechnology Innovation Centre (IBioIC) and Fujifilm Diosynth Biotechnologies (FDB). FDB is a CDMO based in Billingham, UK which specialises in the large-scale production of recombinant proteins. One of their strengths lies in their versatile array of expression platforms, spanning both eukaryotic and prokaryotic origins. Depending on the protein of interest, these platforms offer a range of distinct benefits and drawbacks. Building upon their well-established portfolios of *Escherichia coli* and Chinese Hamster Ovary cell platforms, this thesis investigates a rising new production host, *Komagataella phaffii* (also known as *Pichia pastoris*).

1.1. Recombinant protein production

In order to produce a recombinant protein, a gene encoding that protein, known as the transgene, is required. Typically this is an exogenous gene which is introduced to the new host species for expression (5). The gene, along with carefully selected genetic elements such as a high-powered promoter and efficient terminator, are typically stored within circular DNA molecules called plasmids (6). These plasmids are introduced to the host cell and retained transiently, or it is linearised and integrated into the genome to produce stable clones (6).

The transgene can be drawn directly from the target organism's genome but is more often synthesised artificially (7). In preparation for synthesis, the gene is often modified by making synonymous codon changes in order to better fit the host organism's codon usage preferences and tRNA availability, thereby improving translation efficiency (5). This comes into effect once the gene is transcribed into a messenger molecule, mRNA, which is exported out of the nucleus where it can be translated into protein by host ribosomes (Figure 1.1).

Protein translation occurs just outside the nucleus, at the ribosome. In eukaryotes, the ribosome is a macromolecular complex composed of a small (40S) and large (60S) subunit. Depending on the protein being synthesised, it will be translated at ribosomes found free in the cytosol or tethered to the endoplasmic reticulum (ER) membrane (8). The mRNA molecule is decoded during translation, whereby tRNA molecules charged with amino acids bind to the mRNA, allowing their amino acids to form peptide bonds with the previous amino acid, forming polypeptide chains (9).

Either co-translationally or post-translationally depending on the protein being produced, the polypeptide chain is translocated into the ER where it leaves the cytosolic environment and enters the secretory pathway (10). Co-translational targeting is facilitated an N-terminal signal peptide on the secretory protein, which is recognised by the signal recognition particle (SRP). SRP then directs the ribosome-mRNA complex to the ER membrane, ensuring that nascent polypeptides enter the ER lumen efficiently for proper folding and modification (10).

Within the ER are an abundance of enzymes poised to fold, modify and monitor its progression as the peptide travels. Proteins that fail to fold correctly in the ER are targeted for degradation via the ER-associated degradation (ERAD) pathway, a quality control system that retrotranslocates misfolded proteins into the cytosol where they are ubiquitinated and degraded by the proteasome, thereby maintaining ER homeostasis (11). Correctly-folded

proteins then enter the Golgi for glycosylation and further modification, such as trimming (12), before being exported out of the cell via exocytosis and secreted into the supernatant (13). It is advantageous for recombinant proteins to be secreted into the media as it simplifies downstream purification and processing (14).

This broad summary of the protein production process barely touches the surface of how many different regulatory processes are involved in modulating production. Regulatory mechanisms affecting these processes often link the cell to its environment, optimising processes for survival. As such, these regulatory mechanisms vary greatly depending on the organism, as they are adapted for their specific niche. For example, since bacteria lack membrane-bound organelles, disulfide bond formation is compartmentalised to the periplasm rather than ER, which impacts the regulation of protein misfolding and aggregation (15). Some species have evolved highly efficient protein secretion machinery for essential cell-to-cell signalling. Prolific secretors are ideal production hosts as they avoid the need for additional cell lysis and fractionation steps prior to purification of the final product (16, 17).

Synthetic biology has been used to optimise some of these steps in order to increase yield, generate specific protein modifications, increase reproducibility and facilitate downstream steps such as purification. Production hosts undergo constant development, allowing us to achieve unprecedented levels of control (18). The most studied species have the benefit of there being a deeper understanding of how protein production is regulated and resulted in more diverse and targeted approaches to enhancing protein production. Lesser studied organisms have attracted broader approaches, including large-scale mutagenic screens and more recently AI-mediated modelling to identify the biggest bottlenecks in production and guide further development (19-21). The current climate for recombinant protein production sees a future where a diverse range of host species can be harnessed for their unique strengths and selected accordingly, based on the specific requirements of the recombinant protein of interest. However, achieving this will require additional investment into advancing our understanding of the cell biology of unconventional hosts, elevating it to the level of well-established hosts. This can equip scientists with the knowledge and tools to develop targeted approaches toward enhancing production in unconventional hosts and thereby unlocking their full potential.

1.2. Hosts for recombinant protein production

1.2.1 Prokaryotic production hosts

The first recombinant proteins were produced in the most highly studied organism, *Escherichia coli*. In 1977, the gene encoding somatostatin, a mammalian peptide hormone, was synthesised and cloned into a bacterial expression vector, and was successfully expressed under the control of the lac operon (22). Soon after this proof-of-concept study, human insulin was also produced in *E. coli*, becoming the first ever commercially produced recombinant protein (23). Prior to this, insulin was harvested from pig and cow pancreas, which not raised ethical concerns, but posed risks for patients by triggering adverse immune responses (23). Human insulin produced in *E. coli* was identical to insulin produced in patients' bodies and therefore became a much safer and more effective treatment for diabetes (23). This also paved the way for the production of many other valuable proteins in *E. coli* and other bacterial species, including human growth hormone (24), human interleukin-7 (25) and bone morphogenetic protein 2 (BMP2) (26).

Bacterial production platforms, particularly *E. coli,* offer numerous advantages that make them indispensable in biomanufacturing. Notably, their rapid growth rates are unparalleled, as with a doubling time of 20 minutes under optimal growth conditions, maximum titres can be achieved in as little as one day (27). Rapid growth coupled with its media being inexpensive and readily accessible makes bacterial hosts an incredibly cost-effective option (27). In addition to this, *E. coli* is a well-established model organism in research, meaning it has been extensively studied. This results in work with this species being supported by abundant resources, allowing for easy manipulation of these cells. For example, transformation procedures have been developed to be as short as 5 minutes (27). Given these advantages, it is no surprise that *E. coli* continues to be a popular choice for production host in biotechnology.

However, despite these advantages, bacterial systems have limitations that restrict their versatility in producing complex proteins, particularly for therapeutic applications. One of the major challenges lies in their inability to perform certain post-translational modifications (PTMs), which are critical for protein folding, stability, and function. PTMs are biochemical changes that proteins undergo after translation. These modifications, such as glycosylation, phosphorylation, methylation, and acetylation, play essential roles in determining protein folding, stability, activity, and interactions. The precise pattern of PTMs varies between

species and cell types, influencing protein function and compatibility, especially when producing recombinant proteins for therapeutic applications (28).

For example, glycosylation patterns in prokaryotic systems differ significantly from those in mammalian cells, often leading to improperly folded or non-functional proteins. This can negatively impact protein stability, activity, and immunogenicity, especially for therapeutic applications. Consequently, heavy glycoengineering is required to achieve the production of human-like glycosylated proteins (28). In addition, the small size of the cells imposes a physical limitation, as the expression of larger proteins can pose a significant metabolic burden, often leading to the formation of aggregates, or inclusion bodies (29).

1.2.2 Eukaryotic production hosts

With the increase in demand for therapeutic proteins (biotherapeutics) came a need for a production host with greater capabilities than *E. coli*. The vast majority of currently licensed biotherapeutic products are produced in eukaryotic expression systems (30). This is because these systems can produce PTMs which most closely resemble that of human patients (31-33). Insect and plant production hosts have emerged as a promising alternative host for their ability to carry out post-translational modifications similar to those in humans (34, 35). The first protein produced in insects was the human β-interferon in 1983 (36) and the first plant-made pharmaceutical (PMP) was the human growth hormone, which was produced in tobacco plant cells in 1986 (37). However, these systems also have several constraints, such as limitations in product yields, incompatible glycosylation patterns and challenges with downstream processing which can affect the efficiency and efficacy of the final product (38, 39).

The challenge of producing proteins compatible for therapeutic applications led efforts towards the development of human-based production platforms. Human embryonic kidney cells (HEK cells) have been immortalised and adapted for use in cell biology studies, and many of its derivatives being approved biopharmaceutical production (40, 41). For example, the HEK293 cell line was developed with the ability to generate stable clones and grow in suspension (42, 43). Despite being able to produce excellent quality of protein, due to its human origin, HEK cells can host a plethora of human pathogenic viruses, posing a significant risk of contamination to the patient (44). As a result of this and the often lacking yields seen in human production hosts, non-human recombinant protein platforms are generally preferred for therapeutic protein production.

Since the production of the first recombinant protein, a tissue plasminogen activator, produced in Chinese Hamster Ovary (CHO) cells was approved in 1986, CHO cells have become the main host for therapeutic proteins (45, 46). Like human cells, CHO cells have the appropriate machinery for the folding and PTMs required to produce effective biotherapeutics, so has attracted keen attention for further development. From this, further attractive features were established such as having been adapted to growth suspension, allowing for easily scaled-up production and increased surface area for protein secretion (47). They also now use a more chemically defined media which enables more reproducibility and increased safety compared to human-derived cell lines (48).

Despite the numerous advantages of CHO cells, there are factors which limit their capacity and versatility of production. For example, despite being proficient at conducting most human-like glycosylation patterns, CHO cells are still unable to produce some important human glycosylation types (specifically, α-2,6-sialylation (49) and α-1,3/4-fucoslyation (50)). These differences in glycosylation pathways can lead to the production of glycans that are not expressed in humans, namely α-gal and NGNA, which can evoke immunogenic responses in humans, causing unwanted side effects (32). They are also relatively expensive compared to its microbial counterparts, requiring longer cultivation times and extensive screening as some clones suffer from dramatic loss of productivity due to genetic instability (51). Consequently, a more affordable and reliable production host with the same capabilities is needed.

1.3. Yeast production hosts

The diverse but opposing advantages and disadvantages of mammalian and bacterial expression platforms highlight their distinct use cases; with mammalian production platforms being adept at producing complex therapeutic proteins at relatively high cost and low yield, and bacterial systems being capable of cost-effectively producing high yields of simple proteins. Offering the possibility of producing a wide range of proteins effectively and affordably, the microbial and eukaryotic yeast expression platforms have the potential to offer the best of both worlds (52).

Saccharomyces cerevisiae is a well-studied species of yeast that has been used as a model organism for research for centuries (53). As a result of this, many tools have been developed for the genetic manipulation and analysis of this organism, allowing for the development of the first *S. cerevisiae* expression system in the 1980s (54). This pioneering work brought forward the ability to affordably produce safe and effective biologics such as the first vaccine effective against human viral infections, hepatitis B, which was produced in *S. cerevisiae* in 1984 (55).Whilst *S. cerevisiae* is the most well-established yeast expression platform, the variety of proteins it can produce is limited by issues with plasmid instability, low protein yields and hyperglycosylation (52, 56). Simultaneously, research in other species of yeast revealed further advantageous features. *Kluyveromyces lactis* has reduced ethanol production (produced as a byproduct of fermentation) which negates certain safety issues. *Yarrowia lipolytica* and *Hansenula polymorpha* can better perform complex, human-like PTMs and *Komagataella phaffii,* also known as *Pichia pastoris,* exhibits extremely high cell densities and enhanced secretion capabilities, resulting in high yields and reduced purification requirements (52, 57-59).

Komagataella phaffii in particular has garnered attention, displaying immense potential for the establishment of powerful *K. phaffii* expression platforms for biotherapeutic production. With its early adoption into industrial labs, *K. phaffii* has a proven history of success in producing high yields of recombinant proteins, especially those that are difficult to express in other systems. Naturally a versatile and efficient host, *K. phaffii* has reams of potential to unlock with further innovation, but progress is stunted by the lack of deep understanding of its cell biological processes. Advancements have relied on findings drawn from related yeast species, *S. cerevisiae* and *S. pombe*, which continuously fall short when addressing the unique complexities of *K. phaffii*, such that progress has been slow and imprecise. This thesis aims

to deepen our understanding of this understudied, but promising species and explore novel approaches to enhancing the productivity of *K. phaffii* expression platforms.

1.4. Discovery of Komagataella phaffii

From a French chestnut tree in 1919 and a Californian black oak tree in 1954, isolates of a new yeast species were discovered, and were given the name *Pichia pastoris* (60, 61). In 1969, they were discovered to be methylotrophic, meaning they were capable of utilising methanol as a carbon source (62). Twenty six years later, in 1995, ribosomal RNA sequencing results revealed that the previously indistinguishable isolates from France and the US were, in fact, different species of yeast (63). All *P. pastoris* strains were moved to a new genus, named *Komagataella* after the Japanese scientist Kazuo Komagata, who pioneered a significant body of work investigating methanol-assimilating yeasts (63). A decade later and they were separated into two species; *Komagataella phaffii* and *Komagataella pastoris*, which genomically share 90% identity, with two reciprocal translocations (64). Despite their strong genetic similarity, most of the early work and industrial development was done using *K. phaffii*, leading to its dominance in the field (14).

In the 1970s, Phillips Petroleum Company developed a keen interest in methylotrophic yeasts as they had a vast supply of cheap methane gas, produced as a by-product of their oil refinement process (65). Methane gas can be easily oxidised to methanol, so they saw an opportunity to capitalise on this resource by using their waste product as microbial feed. They explored the public domain of available methylotrophic yeasts and selected *K. phaffii,* known as *Pichia pastoris* at the time, to grow on the synthesized methanol and produce a single cell protein source for animal feed using fermentation. They patented this, restricting global development of this process, but in the early 80's, Phillips Petroleum Company contracted with the Salk Institute Biotechnology/Industrial Associates (SIBIA) to develop the organism for recombinant protein production (65).

Auxotrophy refers to the inability of an organism to synthesize a specific compound required for its growth, necessitating supplementation of that compound in the growth medium. This property is often exploited in genetic engineering to select for successful transformants. In this context, auxotrophic strains were generated, such as the GS115 strain, which is a *his4* auxotrophic mutant, and the X33 strain that is a *HIS4* complemented strain deriving from GS115. These were the first tools for easy genetic manipulation of *K. phaffii* cells (65).

In 1993, Phillips Petroleum sold its patent position to Research Corporation Technologies, who commercially distributed *K. phaffii* strains via Invitrogen (currently a brand of Thermo Fisher Scientific) (65). From this, thousands of labs have obtained *K. phaffii* isolates and begun developing expertise in cultivating and manipulating the species. Now, *K. phaffii* has taken up the intermediate-complexity position in most protein expression lab's toolkit, as it combines the easy cultivation and fast growth of a microbe with the presence of a eukaryotic secretory system, giving it the ability to perform complex post-translational modifications such as N-glycosylation as well as a strong capacity for the formation and isomerization of disulphide bonds (14).

1.5. Study of *Komagataella phaffii*

Due to having genetic and physiological similarities, *K. phaffii* is often compared to *S. cerevisiae* as a model yeast system, especially in the context of protein expression and fermentation processes. *S. cerevisiae* was the first eukaryotic organism to have its genome sequenced (66). However, since the two species diverged 250 million years ago, *K. phaffii* appears to have evolved less rapidly than *S. cerevisiae*, resulting in it retaining more ancient characteristics (67). For example, despite having similar genome sizes, *K. phaffii* having 9.4 Mbp and *S. cerevisiae* having 12.5 Mbp, the organisations of their genomes are vastly different. For example, *S. cerevisiae* distributes its genomic content into sixteen smaller chromosomes, whereas *K. phaffii* instead has four larger chromosomes (66, 68). In contrast, *S. pombe* has an even larger genome of 13.8 Mbp but only 3 chromosomes (69).

Whilst *K. phaffii,* retains ancient features, it is also more reminiscent of higher organisms. For example, *K. phaffii* has large modular centromeres like human centromeres, whereas *S. cerevisiae* has smaller ~125 bp centromeres (70-72). Also, like mammalian cells, *K. phaffii* has an advanced secretory pathway with a stacked Golgi, which unlike the model yeast *S. cerevisiae* which has a dispersed Golgi (73-76). These differences, to name a few, emphasise that, although comparisons between yeast species can provide valuable insights, they may also be misleading, warranting careful interpretation.

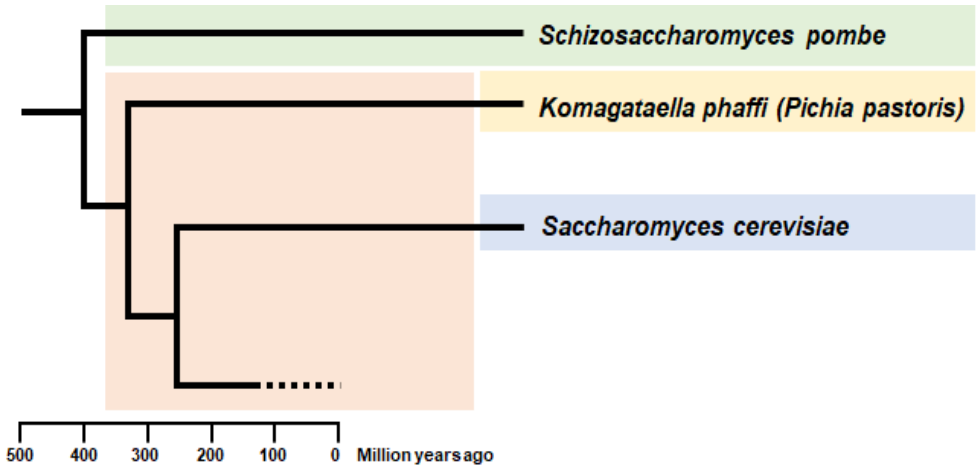


Figure 1.2: Phylogenetic tree illustrating the evolutionary relationships between *S. pombe, K. phaffii* and *S. cerevisiae.*

S. pombe belongs to the Taphrinomycotina subdivision (green) *S. cerevisiae* and *K. phaffii* are both derived from the Saccharomycotina subdivision (orange), but *S. cerevisiae* belongs to the Saccharomycetaceae family (blue) and *P. pastoris* is a methylotrophic yeast (yellow) (67). Times of divergence described by (77).

1.6. Cultivation of *Komagataella phaffii*

1.6.1 The alcohol oxidase I (*AOX1*) promoter system

As a methylotrophic yeast, *K. phaffii* can be cultivated in media containing methanol as the sole carbon source (14). However, it can also be grown on other carbon sources, including glucose, glycerol and sorbitol (78). *K. phaffii* expression systems typically couple recombinant protein production with carbon source metabolism, with transgenes largely being expressed under one of two widely used promoters; the alcohol oxidase I, *AOX1*, promoter (pAOX1) and the glyceraldehyde-3-phosphate dehydrogenase, GAP, promoter (pGAP) (14, 79).

Both pAOX1 and pGAP promoters drive expression of genes which encode enzymes involved in the metabolism of methanol and glycerol respectively (14, 80, 81). They are strong promoters and therefore are effective systems for high yield recombinant protein production but have differing advantages. pAOX1 is an inducible promoter, which is advantageous as it allows for biomass accumulation and protein production phases to be isolated, meaning cellular resources can be devoted to either task without impeding on the other (14). This is achieved by inoculating cultures containing glycerol and once stationary phase is met and glycerol is depleted, transgene expression can be induced by the addition of methanol. It also means that production of recombinant proteins which are toxic to *K. phaffii* cannot be produced, unlike in the pAOX1 system where good yields can still be achieved.

On the other hand, pGAP is a constitutive promoter and therefore does not require unfavourable carbon sources such as methanol (79). Methanol metabolism produces toxic by-products, formaldehyde and hydrogen peroxide (H_2O_2), which causes oxidative stress and elicits the undesirable proteolytic degradation of some important recombinant proteins (82, 83). Methanol is also highly inflammable which poses considerable risk when handling it at large industrial scales (84). Despite these drawbacks, the advantages of the *AOX1* system make it the most widely used within *K. phaffii* production platforms, and it is therefore the chosen system to study in this thesis.

In *K. phaffii*, alcohol oxidase is encoded by two genes; *AOX1* and *AOX2* (85). These allow for three different expression phenotypes. The Mut+ phenotype is where both genes are active, and methanol utilisation is normal. The MutS phenotype is where the *AOX1* gene is not expressed and methanol utilisation is slow. Finally, the Mut- phenotype is where neither *AOX1* or *AOX2* are expressed and cells cannot grow on methanol as a single carbon source (85, 86). MutS strains are produced when the *AOX1* gene is replaced by the transgene, leaving only the *AOX2* gene intact, whereas Mut+ strains are produced when the *AOX1* expression remains intact after integration of the transgene at the *AOX1* locus. Mut+ and MutS strains are the more commonly used strains in recombinant production, with Mut+ having the capability to reach high growth rates and high cell densities and MutS strains producing less heat in the bioreactor and having lower oxygen requirements (87, 88).

This study utilises Mut+ strains that express transgenes under the control of the pAOX1 promoter due to the promoter's tight regulation and strong induction in response to methanol. This design enables controlled, high-level recombinant protein expression while maintaining the organism's native methanol utilisation capacity. The Mut^+ background was chosen to maximise biomass accumulation and protein yield, taking advantage of the strain's fast growth and high cell density potential, which are critical for the objectives of this study.

1.6.2 Feeding strategies

Biologic production in *K. phaffii* involves cultivating cells in bioreactors, which are large vessels that create a controlled environment optimised for protein production (81, 89). Variables such as pH, temperature and cell density are constantly measured and tightly controlled to facilitate the growth and viability of the cells (90). The culture is also consistently agitated using an impeller to minimise the presence of microenvironments where cells could experience transient anoxia, nutrient starvation, and hypoxia (89). In a typical run where the transgene is expressed under pAOX1, cultivation begins with a 'batch' phase, whereby the carbon source (glycerol) is added all at once to the media, either at the same time or shortly after inoculation (81). The vessel is then left until the glycerol has depleted, measured by a spike in dissolved oxygen (DO). Following this is a 'fed-batch' phase whereby methanol is added intermittently (81). The routines of this can vary run to run. Conversely, 'continuous' feeding strategies can be used whereby the carbon source is continuously fed into the bioreactor, typically this is glycerol when the GAP system is used agitation (81). These conditions are simulated during small-scale experiments by cultivating *K. phaffii* in shake flasks where, following a 72-hour batch phase, methanol is added every 24 hours (Figure 1.3).

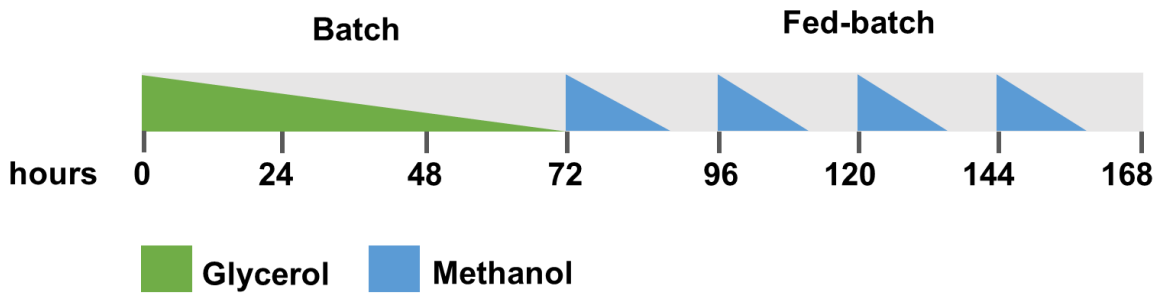


Figure 1.3: Schematic representing carbon sources throughout small-scale cultivation of *K. phaffii*.

1.7. Challenges with Komagataella phaffii

While its capabilities make *K. phaffii* a promising host for recombinant protein production, it is still largely understudied, with significant knowledge gaps restricting innovation. Growing our understanding of the basic cell biology of *K. phaffii* has empowered further advancements in alleviating the most significant bottlenecks in the protein production pipeline.

Discovery that *K. phaffii* has an advanced secretory pathway with a stacked Golgi instead of a dispersed Golgi like *S. cerevisiae* (73-76) opened up the possibility of *K. phaffii* being adept at human-like processing of protein. New strains have been engineered to humanise *K. phaffii* glycosylation patterns, enabling the production of therapeutics while taking advantage of the species' other desirable traits (73, 91, 92). Specifically, the issue to address was that *K. phaffii* glycosylation is of the high-mannose type, with hypermannosylation causing a short *in vivo* half-life to the protein and reducing its efficiency (93, 94). In some cases, it can even cause immunogenic responses, causing unwanted side effects of the biologic (95). On the other hand, human glycosylation is more complex; built on a core with only three mannose sugars, extended with GlcNAc, galactose and sialyic acid (96).

Glycosylation pathways in mammals and yeast are highly conserved, with early processes being identical, including the assembly of the core oligosaccharide, the site-specific transfer of the core to the protein, and trimming of the oligosaccharide by glucosidases (97). The processes diverge at the point of transferring the protein to the Golgi apparatus, where mammalian cells further trim the oligosaccharides via mannosidases, whereas in *K. phaffii,* mannosyltransferases act on the oligosaccharides and add further mannose sugars (97).

Successful attempts to humanise *K. phaffii* glycosylation patterns have deleted the genes involved with hypermannosylation (98). In parallel, heterologous expression of mammalian glycosylation genes has been implemented, including those encoding N-acetylglucosaminyltransferases, galactosyltransferases, sialyltransferases, and nucleotide sugar transporters. These enable stepwise extension of the glycan structure to mimic human N-glycans composed of GlcNAc, galactose, and sialic acid residues (91, 92, 94, 96). Therefore, basic cell biological research into *K. phaffii* cell morphology and endogenous glycosylation patterns unlocked new capabilities of producing human-like therapeutics quickly and more cost-effectively.

Furthermore, it has been found that the *K. phaffii* secretes few endogenous proteins, which is a beneficial trait as it facilitates the purification of recombinant proteins (83). However, this also highlights that secretory capacity can be a limiting factor in recombinant protein production. Efforts to improve recombinant protein production in *K. phaffii* have focused on alleviating potential bottlenecks in the secretory pathway. For example, one strategy involved innovation of the signal peptide, the sorting signal to direct proteins from cytosol to extracellular matrix (99). The α-mating factor (MF) secretion signal from *S. cerevisiae* is commonly used in *K. phaffii* hosts and has been optimised to improve the secretion of recombinant proteins (100). Several endogenous signal peptides were reported to yield much more efficient secretion than the *S. cerevisiae* α-MF (101, 102). However, many approaches towards innovation in this area are tested on reporter proteins such as GFP and prove ineffective, especially in difficult-to-produce proteins. This highlights how often blanket approaches towards improving recombinant protein production can be limited in its efficacy due to the lack of consideration towards product-specific bottlenecks (103, 104). It is also important to note that secretory efficiency is also dependent on other cellular factors such as ER folding capacity, vesicle formation, and trafficking machinery, which may also require engineering to achieve optimal secretion.

Limitations in secretory capacity have also been implicated in causing product yield heterogeneity (105). This is when genetically identical strains can produce protein yields which vary significantly between batches. The exact drivers of this are unclear but it has been hypothesised that epigenetic changes and responses to varying environmental conditions could cause changes in secretory capacity, which result in unpredictable yields (105).

In summary, the key issues surrounding *K. phaffii* expression platforms include a limited depth of fundamental knowledge of the species, resulting in unidentified bottlenecks that hinder efficient production and prevent effective tailoring of the system to the diversity of products that it has the potential to produce. Overcoming such challenges requires innovative approaches that question assumptions based on related species and production hosts, alongside exploring novel strategies to improve production efficiencies. This relies on advancing our current understanding of *K. phaffii*'s cellular processes, which will allow for more effective tailoring of the organism for varied production needs.

1.8. Roles of tRNAs in recombinant protein production

This thesis explores novel avenues towards improving *K. phaffii* expression platforms using the diverse roles of tRNAs. The canonical role of transfer RNAs, or tRNAs, in protein production involves reading the codons on mRNAs and carrying the appropriate amino acid to be joined in the polypeptide chain (106). However, these small molecules and the genes which encode them (tDNAs) are involved in numerous important functions within the cell. This work aims to pioneer new lines of research into tRNAs and tDNAs, and investigates their potential to be a powerful intervention through simultaneously addressing multiple bottlenecks in the production process.

1.8.1 tRNAs in epigenetic regulation

Epigenetic regulation of gene transcription involves changes to the packing of DNA which affects its accessibility to transcription factors and therefore transcriptional activity. A role for tDNAs in epigenetic regulation was implicated when a tRNA-Threonine gene in the silent HMR (homothallic mating right) locus was identified to affect mating type switching in *Saccharomyces cerevisiae* (107). The well-characterised HMR locus can be found downstream of the mating type-determining MAT locus on chromosome 3, where it stores a silenced copy of the MATa allele for mating-type switching in haploid yeast cells (107). The MATa gene is epigenetically silenced until mating-type switching is enabled when the MATa allele from the HMR locus, or conversely the MATα allele from the HML locus, is copied into the MAT locus (108).

Transcriptional repression, or silencing, of the MATa allele is mediated by the HMR-E silencer (109). This is achieved by the recruitment of silent information regulator (Sir) proteins SIR2, SIR3 and SIR4. These bind to the tails of nucleosomes and form a multimeric compound which causes the condensation of the chromatin into transcriptionally silent heterochromatin (110, 111). The issue with this, however, is that propagation of Sir protein activity can cause the spread of heterochromatin to neighbouring genes, but it was found that a genetic element was blocking this spread and insulating neighbouring genes from unintended epigenetic regulation. (107). This element was found to be a tRNA-Threonine gene, or tDNA, and this phenomenon was described as barrier activity (107).

Since then, barrier activity has been observed in other species, including the yeast *Schizosaccharomyces pombe,* mammals and fruit flies (112-114). This unique and powerful

property has been harnessed for industrial applications in mammalian expression platforms, which use barrier elements to improve transgene expression, resulting in increased titre (115). Interestingly, barrier activity is conserved in a pair of human tDNAs, which can act as a barrier against yeast heterochromatin as effectively as endogenous yeast tDNA (116), opening up the possibility of extracting the most powerful tDNAs and using them in an array of recombinant production hosts. Thus far, this avenue has not yet been explored in *K. phaffii*, so this thesis explores whether *K. phaffii* has retained this epigenetic phenomenon and whether it is a beneficial approach towards improving titre.

1.8.2 tRNAs in protein translation

Protein translation occurs at the ribosome, a complex macromolecular machine composed of two subunits, the small 40S and large 60S subunits in eukaryotes, which together decode mRNA sequences into polypeptides. tRNAs recognise codons on the mRNA through their anticodon loops and deliver specific amino acids to the ribosome's A (aminoacyl), P (peptidyl), and E (exit) sites, facilitating sequential peptide bond formation and elongation of the nascent polypeptide chain (106). The efficiency and fidelity of translation are influenced by codon-anticodon interactions and the availability of corresponding tRNAs.

This canonical role of tRNAs in facilitating the translation of mRNA molecules into polypeptide chains has also been exploited for improving the capabilities of production hosts. Routinely, transgenes are adapted to a new system by changing its codons to suit the anticodon portfolio available in the system. It aims to improve translation rates but changing the mRNA structure can have deleterious effects on mRNA transcription and stability. To negate the need for codon optimisation, innovations like the *E. coli* Rosetta strains have been developed to improve the production of eukaryotic proteins in prokaryotes (117). These strains express tRNAs carrying cognate anticodons for rare codons in *E. coli*, in order to compensate for the lack of endogenous tRNAs capable of decoding them. Proteins produced in Rosetta strains have demonstrated increased yields compared to *E. coli* strains without supplemented tRNAs (117). This suggests that increasing the diversity and supply of tRNAs can be an effective intervention for improving recombinant protein production.

Mammalian proteins are routinely codon optimised for production in *K. phaffii*, assuming that rare codons are rate-limiting for protein synthesis, due to a limited supply of anticodons for rare codons in *K. phaffii.* Rare codons are replaced with frequently used ones which align with *K. phaffii* codon usage patterns, assuming that this will increase the rate of translation and improve protein production (5). No consideration is typically taken to assess the effect of codon

changes on mRNA structure and translation efficiencies. This thesis interrogates these assumptions by investigating the two types of decoding that occurs in abundant and rare codons. While most abundant codons are decoded by Watson-Crick base-pairing, a minority rely instead on wobble base-pairing, which involves pairing of only the first two bases of the codon triplet and the last two of the anticodon (118). Wobble-base pairing confers the ability of the tRNA anticodon to recognise more than one mRNA codon, meaning codons can be translated when there is no cognate anticodon available (118). This is advantageous in instances where tRNAs are in limited supply (119). However, this form of decoding is slower and more error-prone than Watson-Crick base-pairing (118). Changing translation rates with synonymous codon changes has been shown to affect protein structure (120). Therefore, there could be a trade-off between increasing yields and decreasing product stability when increasing translation rates.

This thesis investigates some of the assumptions involved in codon-optimising genes for production in *K. phaffii*, to determine if non-optimal codons are limiting for recombinant protein production and explore whether this can be addressed by supplementing tRNAs carrying anticodons missing from the *K. phaffii* portfolio.

1.9. Recombinant proteins used in this study

To explore the versatility of the findings of this study, it is essential to test a variety of proteins with differing characteristics and applications. For this purpose, trypsinogen and PHA-L serve as representative examples of the diverse client proteins FDB may produce. By assessing the production of these proteins under different conditions, we aim to elucidate the effects of introducing tRNA gene to various stages in the protein production pipeline. From this, the capacity for the production host to accommodate the complex demands of different proteins can be explored, providing insights into its broader applicability in biotechnological and pharmaceutical contexts.

1.9.1 Trypsinogen

Trypsinogen is the zymogen, or inactive precursor, of the serine protease enzyme trypsin (121, 122). The physiological role of trypsin is protein digestion in the small intestine, but interest in commercially producing this enzyme stems from its other interesting applications (122). For example, trypsin is commonly used in cell culture laboratories to dissociate adherent cells from the surface of culture vessels (123). By breaking down proteins in the extracellular matrix, trypsin allows cells to detach, which is crucial for passaging (subculturing) cells, harvesting cells for analysis, or preparing them for cryopreservation (123). It is also used widely in proteomics to digest proteins into peptides, which can then be analysed by mass spectrometry (124). The predictable cleavage pattern of trypsin-generated peptides is advantageous for determining the amino acid sequence of proteins (122). Trypsin is also used in the production of recombinant human insulin as cleaves proinsulin at specific sites to produce active insulin for the treatment of diabetes (125, 126).

Despite trypsin being the useful product, its precursor is recombinantly produced instead. This is because the trypsin protease is highly active, meaning it readily digests proteins, including itself and other cellular proteins, including the host cell's machinery(127, 128). This results in severely compromised yields and cell viability. It also presents challenges during purification due to autolysis, or self-digestion (129, 130). Producing the inactive precursor means that it can be converted into trypsin in a controlled manner and has enhanced stability in storage.

K. phaffii has proven to be a successful host for producing trypsinogen, being highly adept at forming the necessary six disulphide bridges for proper protein folding (127, 128). Disulphide bridge formation is necessary to produce many different biologics, especially monoclonal antibodies (mAbs) (131). Proteins like trypsinogen are typically insoluble when produced in *E. coli*, producing inclusion bodies, so *K. phaffii* is often chosen as the more suitable host (127, 132). Trypsinogen is also unglycosylated, so represents a significant proportion of products produced in *K. phaffii*. Risks associated with protease production is also mitigated through the *AOX1* system, where cellular growth and production priorities are separated in the cultivation run.

Trypsinogen production in *K. phaffii* has scope for improvement, with multicopy integrations of the gene increasing production, suggesting that increases in transcriptional activity will be reflected in recombinant trypsinogen production (133). This means that experiments aiming to increase transcriptional activity of the trypsinogen transgene can be tested. Furthermore, this gene is routinely codon-optimised for production in *K. phaffii*, so there is scope to test the role this has in improving translation rates. Ultimately, trypsinogen serves as a faithful representative of typical *K. phaffii* products.

1.9.2 Phytohemagglutinin-L (PHA-L)

Phytohemagglutinin (PHA) is a lectin, which is a carbohydrate-binding protein, derived from the common bean, Phaseolus vulgaris (134). It consists of two closely related proteins: PHA-L and PHA-E, which are responsible for binding leukocytes (white blood cells) and erythrocytes (red blood cells) respectively (135). PHA-L is widely used as a neuronal tracer in neuroscience due to its ability to bind to specific carbohydrates and trace neuronal pathways (136). Producing PHA-L in large quantities for research applications is challenging using natural sources, as it is difficult to purify, so it is instead produced recombinantly in protein production hosts (137).

PHA-L is N-glycosylated at two different sites; a high-mannose type sugar attached at Asn-12, and a complex type sugar at Asn-60 (138). This poses a challenge when producing it in prokaryotic hosts. *E. coli* is incapable of glycosylating eukaryotic proteins, so cannot produce recombinant lectins with the correct functional properties (139-141). Instead, they typically form insoluble inclusion bodies, impeding purification (139-141). *S. cerevisiae* is also not a satisfactory host for PHA-L production as the protein is not correctly processed due to inefficient cleavage of the signal peptide, with approximately half containing an uncleaved signal peptide (142). As a result of this, the protein largely accumulates in the vacuole, with only about 1% being secreted (143). On the other hand, *K. phaffii* has been demonstrated as a suitable host for the recombinant expression of PHA-L, offering several advantages in terms of yield and protein quality (143). In this work, PHA-L serves as a beneficial model for glycoprotein production in *K. phaffii* expression systems.

1.10. Aims and Objectives

K. phaffii is a promising host for the production of a wide variety of recombinant proteins. Innovation in the development of this host requires increasing our knowledge of the protein production pipeline in this host, which will allow us to discover bottlenecks and develop targeted approaches to alleviate them. In this vein, the work presented in this thesis aimed to:

Chapter 3:

- To investigate the role of epigenetic regulation in regulating transgene expression in *K. phaffii* production hosts.
- To explore the possibility of using tRNA gene (tDNA) barriers to improve recombinant protein production in *K. phaffii*

Chapter 4:

- To investigate the role of wobble-mediated translation of the recombinant protein in affecting yields in *K. phaffii* production hosts.
- To explore the possibility of using an exogenous tRNA to improve recombinant protein production in *K. phaffii*

Chapter 5:

- To investigate the global effects of introducing tRNAs on recombinant protein production

2. Materials and Methods

2.1. Media and reagent recipes

Table 1: Buffers and solutions

Reagent	Composition	Preparation
PBS	137mM NaCl, 2.7mM KCl, 10mM Na_2HPO_4, 1.8mM KH_2PO_4	Stored at RT
Lithium acetate buffer	100mM Lithium acetate, 10mM DTT, 0.6M Sorbitol in 10mM Tris HCl pH 7.4	Filter sterilised Stored at 4°C
Li-TE sorbitol	100mL Lithium acetate, 10mM Tris-HCl pH 7.5, 1.2 M sorbitol, 1mM EDTA and 200µM calcium chloride	Filter sterilised Stored at 4°C
Transformation solution	0.6M Sorbitol, 0.1M LiAc, 10mM Tris-HCl	Filter sterilised Stored at RT
Recovery solution	50% YPD, 50% 1 M Sorbitol	Filter sterilised Stored at RT
YNB solution (10X)	34g/L YNB, 100g/L Ammonium sulphate	Filter sterilised Stored at 4°C
Potassium Phosphate Buffer (1 M, pH 6.5)	0.698M Potassium Phosphate Monobasic, 0.3014M Potassium phosphate dibasic	Autoclaved Stored at 4°C
Biotin solution	0.02% Biotin	
Pichia trace elements	0.9 g/l $CaSO_4.2H_2O$ 11.17 g/l $MgSO_4.7H_2O$ 20.0 g/l $ZnSO_4.7H_2O$ 3.0 g/l $MnSO_4$ 5 ml/l $H2SO_4$	Autoclaved Stored at 4°C
Pichia iron and copper solution	6.5 g/l $FeSO_4.7H_2O$ 6.0 g/l $CuSO_4.7H_2O$	Autoclaved Stored at 4°C

RT = room temperature, filter sterilisation was conducted to 0.2 nm

Table 2: Yeast media reagents

Reagent	Composition	Preparation
YPD	1% w/v yeast extract, 2% w/v peptone, 2% w/v d-glucose	Autoclaved Stored at RT
YMD -URA	0.7% YNB, 2% d-glucose + 770mg/L -URA	Autoclaved Stored at RT
YMD -URA -HIS	0.7% YNB, 2% d-glucose + 740mg/L -URA -HIS	Autoclaved Stored at RT
Agar plates	Components of liquid media + 2% w/v agar	Autoclaved Stored at RT
BMM	1% w/v yeast extract, 2% w/v peptone, 0.1 M phosplate buffer, 1X YNB solution, 5% methanol, 0.2% Biotin solution, 0.1% Pichia trace elements, 0.1% Pichia iron and copper solution	Autoclaved Stored at RT
BMGLY	1% w/v yeast extract, 2% w/v peptone, 0.1 M phosplate buffer, 1X YNB solution, 10% glycerol, 0.2% Biotin solution, 0.1% Pichia trace elements, 0.1% Pichia iron and copper solution	Autoclaved Stored at RT
FAIRE fixation solution	10% formaldehyde, 4% methanol in 1xPBS	Prepared fresh

RT = room temperature

Table 3: Reagents for protein analysis

Reagent	Composition	Preparation
TWIRL buffer	50mM Tris.HCl (pH 6.8), 5% SDS, 10% glycerol, 8M urea, 10% (v/v) 2-mercaptoethanol is added fresh before use	Filter sterilised Stored at RT
SDS-PAGE sample buffer	150mM Tris pH 6.8, 8% w/v SDS, 166mM DTT	Stored at RT
Fairbanks A	0.05% Coomassie, 10% Acetic acid, 25% Isopropanol	Stored at RT
Fairbanks B	0.005% Coomassie, 10% Acetic acid, 10% isopropanol	Stored at RT
Fairbanks C	0.002% Coomassie, 10% Acetic acid	Stored at RT
Fairbanks D	10% Acetic acid	Stored at RT
Running Buffer	25mM Tris, 192mM Glycine, 0.1% SDS	Stored at RT
Semi-Dry transfer buffer	25mM Tris, 192mM Glycine, 0.1% SDS, Methanol	Stored at RT
Resolving gel buffer	1.5 M Tris HCl pH 8.8, 0.4% SDS	Stored at RT
Stacking gel buffer	0.5 M Tris HCl pH 6.8, 0.4% SDS	Stored at RT

RT = room temperature, filter sterilisation was conducted to 0.2 nm

Table 4: Recipe for SDS-PAGE gel preparation

	15 % resolving	4 % stacking
dH2O	3ml	2.7ml
Resolving/stacking gel buffer	3ml	1.125ml
Protogel (30% Acrylamide)	6ml	0.6ml
10 % APS *	100µl	40µl
TEMED	20µl	10µl

Table 5: Plasmids used and constructed for *S. cerevisiae* mating assay for barrier activity

Plasmid	Features	Origin
pDD371	a1 gene downstream of the E silencer *Negative control as it contains no barrier elements. Putative barrier elements can be cloned into the linker using NotI and BamHI. Retains mating when transformed into DDY171.* Backbone is vector is pRS406 (196) Selection: URA3	Donze lab – Referred to as pRO363 (196)
pDD442	*S. cerevisiae* tDNA ($tRNA^{Thr}$) downstream of HMR with about 100 bp flanking sequences on both sides *Positive control as it contains the $tRNA^{Thr}$ gene, shown to be a strong barrier in S. cerevisiae.* Backbone is vector is pRS406 (196) Selection: URA3	Donze lab - Referred to as pRO466 (196)
pDD371 + tRNA-threonine candidates	*K. phaffii* $tRNA^{Thr}$ genes downstream of the HMR silencer, upstream of the a1 mating type gene, with 250bp flanking regions. Backbone is vector is pRS406 (196) Selection: URA3	This study
pDD371 + orphan A and B boxes	*K. phaffii* centromeric putative ETC sites genes downstream of the HMR silencer, upstream of the a1 mating type gene, with 250bp flanking regions Backbone is vector is pRS406 (196) Selection: URA3	This study

Table 6: *K. phaffii* integrative plasmids used and constructed in Chapter 4

Plasmid name	Linearisation site	Description	Origin
Barrier plasmid	*NdeI*	Codon-optimised Trypsinogen gene under *AOX1* promoter with 1000bp homologous recombination site and the tRNA-ch1.tRNA31 and ch1.tRNA32 gene pair upstream of the *AOX1* locus	FDB
pT1	*PmeI*	Codon-optimised Trypsinogen gene under *AOX1* promoter	This study
pT1 + tRNA	*PmeI*	pT1 with *S. cerevisiae* tRNA-Leu-GAG	This study
pT2	*PmeI*	pT1 with half of all leucine codons changed to CUC	This study
pT2 + tRNA	*PmeI*	pT2 with *S. cerevisiae* tRNA-Leu-GAG	This study
pT3	*PmeI*	pT1 with all leucine codons changed to CUC	This study
pT3 + tRNA	*PmeI*	pT3 with *S. cerevisiae* tRNA-Leu-GAG	This study
pP1	*PmeI*	Codon optimised PHA-L gene under *AOX1* promoter	FDB
pP1 + tRNA	*PmeI*	pP1 with *S. cerevisiae* tRNA-Leu-GAG	FDB
pP2	*PmeI*	pP1 with half of all leucine codons changed to CUC	FDB
pP2 + tRNA	*PmeI*	pP2 with *S. cerevisiae* tRNA-Leu-GAG	FDB
pP3	*PmeI*	pP1 with all leucine codons changed to CUC	FDB
pP3 + tRNA	*PmeI*	pP3 with *S. cerevisiae* tRNA-Leu-GAG	FDB

Table 7: *S. cerevisiae* strains used and constructed in this study

Strain name	Species	Genotype	Origin/Source
DDY7	*S. cerevisiae*	MATa can1 his4-519 leu2-3,112 trp1 ura3-52	David Donze
DDY171	*S. cerevisiae*	MATα ADE2-1 his3-11 leu2-3,112 LYS2 trp1-1 ura3-1 hmr::bgl-bclΔ	David Donze
DDY171 + $tRNA^{Thr}$	*S. cerevisiae*	DDY171 transformed with 'pDD371 + tRNA-threonine candidates'	This study
BY4741	*S. cerevisiae*	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	MacDonald lab

Table 8: *K. phaffii* strains used and constructed in this study

Strain name	Species	Genotype	Transformation protocol	Origin/Source
CLD392	*K. phaffii*	Wildtype strain CBS7435	N/A	FDBK
Barrier strain	*K. phaffii*	CLD392 expression Barrier plasmid	2	FDBK
T1	*K. phaffii*	CLD392 expressing pT1	1	FDBK
T1 + tRNA	*K. phaffii*	CLD392 expressing pT1 + tRNA	1	This study
T2	*K. phaffii*	CLD392 expressing pT2	1	This study
T2 + tRNA	*K. phaffii*	CLD392 expressing pT2 + tRNA	1	This study
T3	*K. phaffii*	CLD392 expressing pT3	1	This study
T3 + tRNA	*K. phaffii*	CLD392 expressing pT3 + tRNA	1	This study
P1	*K. phaffii*	CLD392 expressing pP1	2	This study
P1 + tRNA	*K. phaffii*	CLD392 expressing pP1 + tRNA	2	This study
P2	*K. phaffii*	CLD392 expressing pP2	2	This study
P2 + tRNA	*K. phaffii*	CLD392 expressing pP2 + tRNA	2	This study
P3	*K. phaffii*	CLD392 expressing pP3	2	This study
P3 + tRNA	*K. phaffii*	CLD392 expressing pP3 + tRNA	2	This study

2.2. Plasmid construction

Plasmids used and constructed in this thesis are listed in Table 5 and Table 6. For *S. cerevisiae* mating assay plasmids, pDD371 and pDD442 were generous gifts from David Donze (Louisiana State University). *K. phaffii* barrier candidates were cloned into the NotI site using Gibson Assembly. *K. phaffii* barrier plasmid was synthesised by Invitrogen GeneArt Synthesis services. *S. cerevisiae* tRNA-Leu-GAG was cloned into PciI site of T1, also known as pAVE703 (FDB pPICZ variant), using Gibson Assembly. Transgenes were subcloned from Invitrogen GeneArt carrier vectors into the pAVE703 vectors with and without the *S. cerevisiae* tRNA gene using restriction ligation cloning and the NotI and EcoRI restriction sites.

Inserts which were not subcloned were generated by high-fidelity PCR from genomic DNA using high fidelity Q5 Hotstart Mastermix, following manufacturer's protocol. Primers were designed using Benchling and obtained from IDT technologies and Merck. Inserts were gel-purified using the Qiagen QIAquick PCR Purification Kit. Vectors were purified using gel electrophoresis and the Thermo Scientific GeneJET Gel Extraction Kit. Where linearised by one restriction enzyme, vector was treated with AP phosphatase (Thermofisher) before gel purification. Inserts and vectors were assembled using the NEBuilder® HiFi DNA Assembly Master Mix, following manufacturer's protocol. Assemblies were obtained upon 1 hour incubation of a 15µL reaction mixture at 50°C. Successful clones were selected for by transformation into competent DH5a *E. coli* cells (NEB® 5-alpha Competent *E. coli*) following the NEB High Efficiency Transformation Protocol. Transformed *E. coli* cells were plated onto LB Ampicillin or LB Zeocin plates for selection. Successful clones were confirmed using restriction digestion and colony PCR using PCRBIO Ultra polymerase, following manufacturer's instructions.

2.3. Genomic DNA isolation

Genomic DNA was extracted from overnight cultures, prepared by inoculating 5ml YPD with either wildtype *S. cerevisiae* cells (BY4142) (Table 7) or wildtype *K. phaffii* (CLD392) (Table 8) and incubating cultures at 30°C (shaking, 200 rpm) overnight. Cells were harvested and spheroplasts were generated by a further overnight incubation of cells with 3µl β-mercaptoethanol and 10µl zymolyase (10mg/ml) . gDNA was released and purified by three cycles of adding 500µl Phenol:Chloroform:Isoamyl alcohol (25:24:1), vortexing and centrifugation at 15000rpm for 5 minutes to fractionate the aqueous phase. gDNA was then concentrated using isopropanol precipitation, washed 3 times with 70% ethanol and resuspended in 100µl dH_2O.

2.1. *S. cerevisiae* transformations

Competent cells were prepared from overnight cultures, obtained by inoculating 5ml YPD with wildtype *S. cerevisiae* cells (BY4142) (Table 7). 3ml overnight culture was added to 47ml YPD in a sterile 50ml flask and grown at 30°C (shaking, 200 rpm) until OD_{600} 1-1.5 was achieved. Cells were pelleted and washed in 5ml Li-TE sorbitol. Cells were pelleted again, resuspended in 1ml Li-TE sorbitol and incubated at 30°C (shaking, 200 rpm) for 1 hour. Cells were incubated on ice for 30 minutes prior to transformation.

100µl competent cells were incubated with 150µl 70% PEG-3350, 5µl salmon sperm and 500ng plasmid DNA at 30°C (shaking, 200 rpm) for 45 minutes. Cells were heat shocked at 42°C for 20 minutes and pelleted before being resuspended in 100µl dH2O and plated on auxotrophic plates (-URA or -URA -HIS) for selection.

2.2. *S. cerevisiae* mating assay for barrier activity

Mating assays were performed to test *K. phaffii* tDNAs and other genetic elements for barrier activity in *S. cerevisiae* as described by Donze et al in 1999 (6). DDY171 (α-mating strain) (Table 7) was transformed with plasmids with and without a putative barrier between the silencer and a1 gene. Four colonies of each transformant were selected for mating with a-expressing strain, DDY7 (Table 7) (Figure 2.1). DDY cells were prepared for mating by being grown at 30°C (shaking, 200rpm) overnight in YPD. Cultures were washed by pelleting and resuspension in equal volumes of sterile dH_2O to remove residual nutrients. Mating lawns were prepared by spreading 300µl DDY7 cells on YMD -URA -HIS plates. Mating lawn was allowed to dry before colonies were spotted directly onto the mating lawn to prevent spreading. Colonies were prepared using overnight cultures diluted to an OD_{600} of 0.2. Plates containing spotted mating lawns were incubated at 30°C (stationary) for 2 days. The mating potential of cells was monitored by the -HIS selection as the *his3-11* and *his4-519* mutations in DDY7 and DDY171 would be cancelled out in their progeny, enabling growth on media lacking histidine. The schematic in Figure 2.1 outlines the procedure and concepts underpinning this technique.

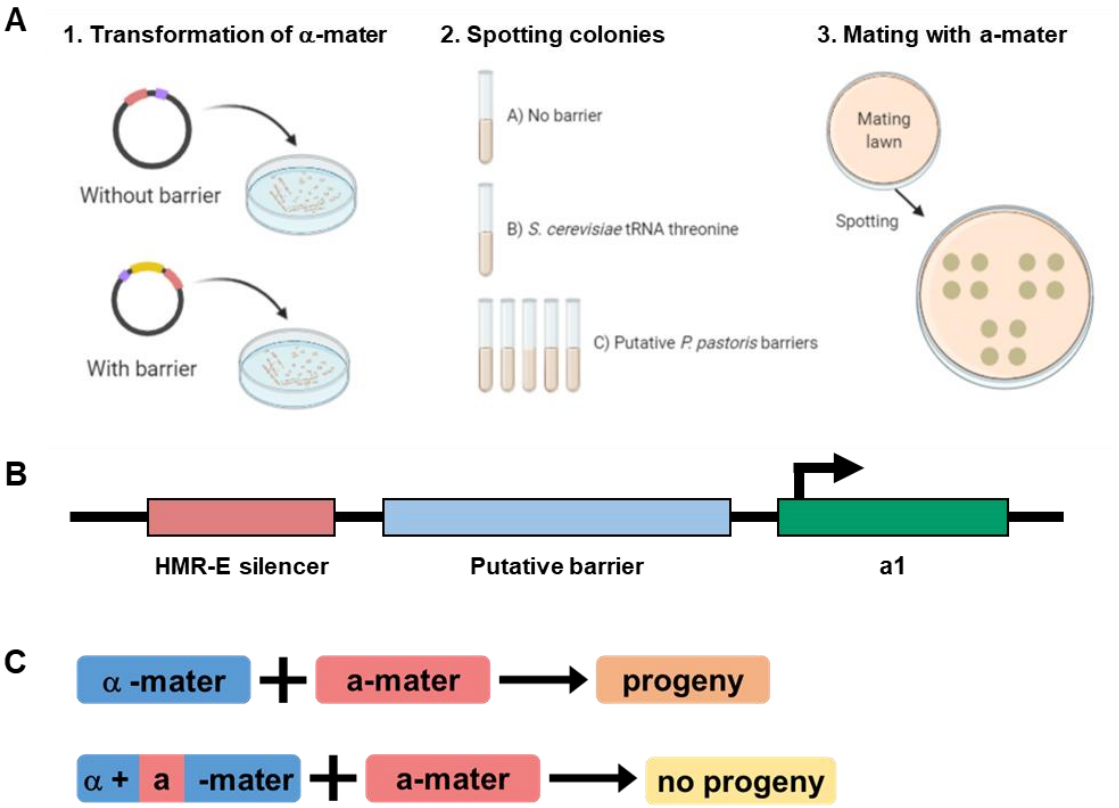


Figure 2.1: Schematic portraying the adapted barrier assay protocol used in this study.

(A) Schematic of barrier assay protocol. Strain DDY171 was transformed with either pDD371, containing no barrier, or a plasmid containing a putative barrier. Colonies of the transformants were grown overnight in YMD-URA media at 30°C before being spotted directly onto the mating lawn using overnight liquid cultures diluted to an OD_{600} of 0.2. (B) Map of the mating-type region in the barrier assay. Potential barriers can be tested for their ability to prevent silencing of the a1 gene by being placed between the HMR-E silencer region and the a1 gene. (C) Schematic displaying the outcomes of positive and negative mating phenotypes. Mating phenotype is determined by the expression of the a1 gene in alpha mating cells. Barrier activity is assessed through expression of the a1 gene.

2.3. *K. phaffii* transformations

2.3.1 Preparation of linear DNA

Overnight cultures of E. coli transformed with the plasmid of interest (Table 6) were grown overnight at 37°C (shaking, 200rpm). For Transformation protocol 1, 5µg DNA is required so DNA was extracted using Qiagen Midi/Maxi prep kits. Plasmid concentrations were determined using nanodrop. Plasmids were linearised according to the enzyme indicated in Table 6. Digestion reaction mixtures were conducted according to NEB protocols and incubated overnight at 37°C (stationary). Linearised plasmid was checked for complete digestion using gel electrophoresis. For Transformation protocol 1, DNA was then precipitated using 0.7x isopropanol and centrifugation before being washed with 70% ethanol and resuspended in 50µl dH_2O. For transformation protocol 2, DNA was purified by gel extraction using the Thermo Scientific GeneJET Gel Extraction Kit.

2.3.2 Transformation protocol 1

Wild-type *K. phaffii* (CLD392) strains were streaked out on YPD plates and incubated at 30°C (stationary) for 3 days before competent cell preparation. Overnight cultures were prepared by inoculating a single colony into 5 ml YPD and incubating overnight at 30°C (shaking, 200 rpm). An aliquot of overnight culture was used to inoculate 500ml YPD in a 2L shake flask. The following formula was used to calculate the volume of overnight culture to use in order to achieve OD_{600} of 0.8 at a convenient time:

$$\text{Volume of overnight culture (µl)} = \frac{0.8 \times 500}{e^{0.347 \times \text{OD600 of overnight culture}}}$$

Culture was grown overnight at 30°C (shaking, 200rpm). Once the correct optical density was achieved, the culture was pelleted and resuspended in 150ml Li-Ac before incubating at room temperature for 30 minutes. Cells were pelleted again and washed with decreasing volumes of 1M Sorbitol before finally resuspending in 2ml 1M Sorbitol before electroporation. 80µl cells was incubated in an Eppendorf with 10µl DNA for 5 minutes on ice before being transferred to 0.2mm gap electroporation cuvettes (BIORAD). Cells were electroporated using a BIORAD MicroPulser Electroporator using the PPIC setting before 1ml ice cold 1M Sorbitol was instantly added to the cells. Cells underwent recovery in a 15ml falcon tube at 30°C (stationary) for 1 hour, followed by 1ml YPD addition to the mixture and a further 1-hour incubation. Cells were plated on YPDS plates containing 100µg/ml to 1.2mg/ml Zeocin for selection.

2.3.3 Transformation protocol 2

Wild-type *K. phaffii* (CLD392) strains were streaked out on YPD plates and incubated at 30°C (stationary) for 3 days before competent cell preparation. Overnight cultures were prepared by inoculating a single colony into 5 ml YPD and incubating overnight at 30°C (shaking, 200 rpm). An aliquot of overnight culture was used to inoculate 100ml YPD in a 250ml shake flask. The following formula was used to calculate the volume of overnight culture to use in order to achieve OD_{600} of 1 at a convenient time:

$$\text{Volume of overnight culture (µl)} = \frac{1 \times 100}{e^{0.347 \times \text{OD600 of overnight culture}}}$$

Culture was grown overnight at 30°C, (shaking, 200rpm). Once the correct optical density was achieved, the culture was pelleted and resuspended in 25ml transformation solution. 250µl DTT was added to the cells before incubating at room temperature for 30 minutes. Cells were pelleted again and washed three times with 25ml ice cold 1M Sorbitol, with 5-minute incubations between washes. Cells were pelleted and finally resuspended in 500µl 1M Sorbitol before electroporation.

80 µl cells was incubated in an Eppendorf with 10µl DNA for 5 minutes on ice before being transferred to 0.2mm gap electroporation cuvettes (BIORAD). Cells were electroporated using a BIORAD MicroPulser Electroporator using the PPIC setting before 1ml recovery solution was instantly added to the cells. Cells underwent recovery in a 2ml falcon tubes at 30°C (shaking) for 4 hours. Cells were plated on YPDS plates containing 100µg/ml Zeocin for selection.

2.3.4 Colony PCR in *K. phaffii*

Genomic DNA was isolated according to *K. phaffii* gDNA extraction methods or prepared crudely by pelleting cells, creating a cell suspension in dH_2O, boiling cells for 5 minutes at 95°C and vortexing with glass beads. gDNA was used as a template with the relevant primers and PCR was carried out using PCRBIO Ultra polymerase following manufacturer's instructions.

2.4. *K. phaffii* cultivation and sampling

Single colonies of *K. phaffii* strains (Table 8) were used to inoculate 50 ml BMGLY (Table 2) in 125ml conical flasks, which were incubated at 30°C in an orbital shaker at 210 +/- 10rpm for 72 hours, before the temperature was lowered to 28°C for the rest of the experiment. Methanol inductions involved adding 5% BMM (Table 2) such that the final methanol concentration in culture was 0.5%. Cultures were fed methanol every 24 hours and, where indicated, samples were taken 1 hour, 4 hours and/or 7 hours post-induction for RNA, chromatin, extracellular protein or intracellular protein analysis.

Cell pellets for RNA and intracellular protein were obtained by aliquoting 1ml culture into an Eppendorf and centrifugation at max speed for 5 minutes. Supernatants were collected, combined with SDS-PAGE sample buffer and stored at -20°C in preparation for secreted protein analysis. For RNA, cell pellets were resuspended in 1ml Trizol and stored at -80°C. For intracellular protein, cell pellets were washed once with sterile dH_2O and resuspended in 1ml freshly prepared TWIRL buffer and stored at -20°C. For chromatin, 1ml cells were added to 9ml fixation solution, incubated at 30°C (shaking, 200rpm) for 30 mins. 500µl cell fixation mixture was removed and replaced with 500µl 2.5M glycine before incubation at room temperature on a rocker for 5 minutes. Cells were washed three times in 1ml PBS (Table 1) before fixed pellets were stored at -80°C in preparation for chromatin extraction.

2.1. Reverse-transcriptase quantitative PCR (RT-qPCR)

2.1.1 RNA isolation

Cells in 1ml in TRIzol® reagent were lysed by vortexing with glass beads for 10 minutes before incubating for 5 minutes to permit complete dissociation of the nucleic acid-protein complexes. 200µl chloroform was added and tubes were inverted for 15 seconds before incubation for 5 minutes. Aqueous phase was separated by centrifugation for 5 mins at 16000 × g at 4°C. Aqueous phase was transferred to a new tube and was washed with another 200µl chloroform. 10µg of RNase-free glycogen (ThermoFisher Scientific) was added as a carrier to the aqueous phase. RNA was precipitated out of the aqueous phase using isopropanol and washed with 75% ethanol three times before being resuspended in 40µl dH_2O. 5µl DNAse I and 5µl DNAse I buffer (ThermoFisher Scientific) was added to the RNA, which was then incubated at 37°C for 30 minutes. RNA was stored at -80°C.

2.1.2 cDNA synthesis and qPCR

cDNA library was synthesised using PCR Biosystems Reverse Transcriptase kit, following the manufacturer's protocol. cDNA samples were analysed by quantitative PCR (qPCR) using Luna® Universal qPCR Master Mix, following the manufacturer's protocol. Data was analysed using the delta-delta Ct method, normalising to ACT1 or TAF10 as indicated.

2.2. Formaldehyde Assisted Isolation of Regulatory Element quantitative PCR (FAIRE-qPCR)

Buffers are described in Table 1. Formaldehyde-fixed cells were thawed and resuspended in 1mL lysis buffer (2% (v/v) Triton X-100, 1% (w/v) SDS, 100mM NaCl, 10mM Tris-Cl, 1mM EDTA) and transferred to 2 mL screw cap tubes containing acid-washed glass beads. Cells were vortexed at 4°C for 5 cycles (2 seconds on, 30 seconds off). The lysate was transferred to a 15 ml tube. The glass beads were washed with an additional 1ml lysis buffer, which was then added to the same conical tube. The lysate was sonicated in a Bioruptor® Plus for 10 cycles (30s on/30s off) at high power setting. The sonicated lysate was clarified by centrifugation at 10,000 × g for 5 minutes and around 150µL of this clarified lysate was removed for preparation of input control DNA. The remaining lysate was aliquot into fresh 1.5mL tubes.

1 volume of phenol/chloroform/isoamyl alcohol (ThermoFisher Scientific, Vienna, Austria) was added, vortexed vigorously and centrifuged at full speed for 10 minutes. The top layer was transferred to a fresh tube. This was repeated once more, after which 200µL of chloroform/isoamyl alcohol (ThermoFisher Scientific, Vienna, Austria) was added to each tube to remove any remnant phenol. The tubes were vortexed, centrifuged and the aqueous layer transferred to a fresh tube. 1/10 volume 3M sodium acetate (pH 5.2), 2 volumes 95% ethanol and 1µL 20mg/mL glycogen was added to each tube and the tubes were incubated at −80°C overnight.

Tubes were centrifuged at full speed for 20 minutes to precipitate the DNA, after which the pellets were washed with 75% ice-cold ethanol and centrifuged again for another 10 minutes. Finally, the ethanol was removed, and the pellets dried and resuspended in 50µL 10mM Tris–HCl (pH 7.4). These FAIRE samples were treated with DNase-free RNase A (30 minutes, 37°C), Proteinase K (1 hour, 55°C) and finally incubated overnight at 65°C (stationary) for decrosslinking.

For preparation of input control DNA, the 150µL clarified lysate removed after sonication was first treated with DNase-free RNase A, Proteinase K and de-crosslinked overnight at 65°C before proceeding with Phenol/Chloroform/Isoamyl alcohol extraction as described for the FAIRE samples. Both the FAIRE DNA and input control DNA were further purified using MinElute PCR purification kit (Qiagen, Hilden, Germany), quantified and run on an agarose gel to check proper sonication efficiency (Figure 0.1).

Input and FAIRE samples were analysed by quantitative PCR (qPCR) using Luna® Universal qPCR Master Mix, following the manufacturer's protocol. Data was analysed using the delta-delta Ct method, normalising FAIRE signal to input signal.

2.3. Secreted protein analysis

Buffers are described in Table 3. Samples for secreted protein analysis were taken at the end of the cultivation run (Day 5, 7 hours post methanol induction) and combined with SDS-PAGE sample buffer prior to analysis. Samples were boiled for 5 minutes at 97°C before proteins were separated by SDS-PAGE on 15% gels. Gels were stained with Coomassie Blue following the Fairbanks method (144).

Gels were covered in Fairbanks buffer A and microwaved until boiling. Gels were then shaken for 10 minutes before buffer A was poured away and the gel was rinsed in distilled water. Buffer B was then added, and the gel was microwaved until boiling before being rinsed in distilled water again. This was repeated for buffer C and D, except the gel was left in buffer D to shake for a few hours or overnight, until the gel completely destained.

Stained gels were imaged using the iBright Imaging System (Thermo Fisher Scientific). Bands were quantified by densitometry using ImageJ using the “Gels” function. Firstly, the rectangle tool was used to select all lanes and plot the profile areas. The band peak was then isolated above the background level using the straight-line tool. Finally, the wand tool was used to calculate the area of the peak in arbitrary units. Target protein band densities were normalised as specified in the figure legends to convey fold change. Total protein staining of the sample derived from quantifying a full lane of a Coomassie brilliant blue stained SDS-PAGE gel.

2.4. Intracellular protein analysis

Buffers are described in Table 3. Cell pellets suspended in TWIRL buffer underwent protein separation using SDS-PAGE gel electrophoresis in duplicate; one gel was used for western blotting and the other for total protein analysis.

For western blotting, resolved proteins were transferred to PVDF membranes, using semi-dry transfer (Bio-Rad) according to the manufacturer's guidelines. Membranes were blocked at room-temperature for 1 hour in 5% (w/v) milk powder in PBS-T. Membranes were incubated with primary antibody, either a polyclonal rabbit unconjugated anti-trypsinogen antibody (NB600-615, dilution: 1 in 250, Biotechne) or rabbit unconjugated anti-PHA-L (AS-2300-1, dilution: 1 in 1000, Vector Labs), in 5% (w/v) milk powder in PBS-T overnight at 4°C. After being washed six times with PBS-T, the membranes were incubated for 1 hour at room temperature with an anti-rabbit HRP-conjugated secondary antibody, diluted in blocking buffer. The membranes were washed a further six times and were then imaged using the iBright Imaging System (Thermo Fisher Scientific) after incubation with the chemiluminescent substrate, SuperSignal West Pico Plus Chemiluminescent Substrate (Thermo Fisher Scientific).

For total protein analysis for normalisation, the SDS-PAGE gel was stained following the Fairbanks method as described above (144). Stained gels were imaged using the iBright Imaging System (Thermo Fisher Scientific).

Western blot images were quantified by densitometry using ImageJ using the “Gels” function as described above. Target protein band densities were normalised as specified in the figure legends to convey fold change. Total protein staining of the sample derived from quantifying a full lane of a Coomassie brilliant blue stained SDS-PAGE gel.

2.5. Cell size measurements

Cells were grown to early log phase (OD_{600} = 1) and harvested by centrifugation. Cells were washed and resuspended in minimal media, YMD-URA, before being spotted onto a microscope slide. Confocal brightfield microscopy was used to image cells. Micrographs were measured using ImageJ (version 1.46, NIH, USA) to determine cell area. 400-600 cells were sampled per biological replicate. Magic cell wand was used to measure cell area.

2.6. Bioinformatics analysis

2.6.1 ATAC-seq and FAIRE-seq analysis

tRNA gene regionsets were obtained from the Genomic tRNA gene database (available at http://gtrnadb.ucsc.edu/). Chromatin accessibility data, ATAC-seq (Accession ID: GSE154330) and FAIRE-seq (Accession ID: GSE131290) datasets were obtained from NCBI. Data was visualised using EAseq (available at https://easeq.net) (145). Heatmap was generated using the 'heatmap' function provided by EAseq. tRNA gene accessibility was calculated using the 'Quantify' function provided in EAseq. Boundary and non-boundary tRNAs were gated into separate region sets and ATAC-seq signal intensity was compared using the 'FillTrack' function on EAseq.

2.6.2 Boundary tDNA identification

tRNA gene regionsets were obtained from the Genomic tRNA gene database (available at http://gtrnadb.ucsc.edu/). Chromatin accessibility data, ATAC-seq (Accession ID: GSE154330) dataset, were viewed using UCSC genome browser (available at https://genome.ucsc.edu) and aligned to the CBS7435 genome. Data was visualised using EAseq (available at https://easeq.net) (145). tRNA gene accessibility was calculated using the 'Quantify' function provided in EAseq. Average accessibility values were calculated for isoacceptors of each amino acid.

2.6.3 Nucleosome occupancy

DNA sequences, obtained from the Genomic tRNA gene database (available at http://gtrnadb.ucsc.edu/) and the *K. phaffii* genome (NCBI), were analysed using NuPoP, an R package for Nucleosome Positioning Predictions. Predictions were made based on the primary sequences around the putative barriers using duration hidden Markov model. The model assumes two oscillating states when analysing a DNA sequence; the nucleosome and the linker region. The nucleosome state has a fixed length of 147bp, and the linker state has a variable length. Further details of this package is described in Xi et al (2010) (146).

2.6.4 Codon demand analysis

Gene sequences were imported from a FASTA file using the readDNAStringSet function from the Biostrings package in R. Sequences were converted into a data frame, with NCBI gene identifiers extracted from sequence headers using the Regex package. Gene lengths were calculated and stored as a separate column. Genes lacking NCBI IDs were separated, and sequences were run through BLAST against the *K. phaffii* nucleotide database to assign NCBI IDs. This dataset containing NCBI gene IDs was imported from a CSV file and merged with the gene sequence data frame to reconcile any missing gene identifiers.

CTC codon counts per gene was obtained by reading the FASTA sequences into a data frame using the Biostrings R package. The trinucleotideFrequency function was used to obtain codon counts for each possible codon, and the dataset was filtered to portray CTC codon counts only. Pie chart was generated using base R to visualise the distribution of genes based on their CTC codon content, dividing genes into two categories: those with zero CTC codons and those with one or more. Codon counts were normalised to gene length.

RNA-seq (GSE159325) and Ribo-seq data (GSE159336) was used to calculate demand for tRNA-Leu-GAG, which was calculated. The calculation was conducted as described below:

$$\frac{\text{mRNA transcript count (tkmp)} \times \text{CUC codon count}}{\text{Gene Length (bp)}} = \text{Demand for CUC codon}$$

Percentage demand was calculated in order to determine:

$$\frac{\text{demand for gene x}}{\text{Total demand}} \times 100 = \text{Percentage demand for gene x}$$

Demand was ordered ascendingly and an additional column for cumulative count was added. Top genes with highest demand for the tRNA were filtered using cumulative count, gating genes until the declared threshold was met. This subset of genes was then carried forward for further analysis.

2.6.5 KEGG enrichment analysis

To identify significantly enriched pathways associated with high demand for tRNA-Leu-GAG, KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway enrichment analysis using the enrichKEGG function. This analysis was applied to genes with highest demand for the tRNA using the *Komagataella phaffii* organism annotation (organism = 'ppa') and NCBI gene identifiers (keyType = "ncbi-geneid"). Results of this analysis were stored as a data frame where significant pathways were stored, accompanied by pathway-specific metrics, including Count (the number of genes enriched in each pathway). This was converted to numeric format for further analysis and visualisation in R.

3. Chapter 3: Exploring the use of tDNA barriers to improve recombinant protein production

3.1. Introduction

3.1.1 The role of epigenetics in recombinant protein production

All eukaryotic species possess the crucial ability to regulate expression of its genes in response to adverse environmental conditions, all with the purpose of increasing its chances of survival. This ability is known as epigenetic regulation. Such regulation is not seen on the primary sequence level and thus unidentifiable via sequencing (147). Epigenetic regulation of gene expression means that populations of cells which are genetically identical can be heterogeneous in terms of their proteomes (147). Even under carefully optimised conditions designed to maximise productivity, single-cell factories still rely on these conserved survival mechanisms to maintain cell viability throughout the production process. While they ensure survival, epigenetic survival mechanisms can conflict with the goal of maximising protein production.

A well-characterised challenge in CHO production systems is unstable transgene expression during production runs, whereby a significant decline in productivity is observed during the approximately 60 generations of growth during the scale-up to production volumes (51, 148, 149). Numerous mechanisms underpin this, but transcriptional repression due to epigenetic silencing is a particularly major issue for recombinant protein production in CHO cells (148, 150). Yields are severely compromised when transgenes are integrated into silenced regions of the genome, and when integrated into a positive environment for expression, transgenes are still vulnerable to the spread of gene silencing activities (150, 151).

DNA elements, such as tRNA genes (tDNAs), are involved in separating epigenetic domains, and examples have been identified in numerous eukaryotes, including in *S. cerevisiae* (107), *S. pombe* (152) and CHO cells (115). These DNA elements function as ‘barriers’ against the spread of gene silencing to protect crucial genes from being silenced (153). It was recently found that the highly conserved barrier function can be exploited in industrial CHO strains, to insulate transgenes and stabilise expression over long term culture (115). This finding was particular exciting as it mitigates a major issue in the manufacturing of biotherapeutics (115).

As a rising host for biotherapeutic production, the aim of this work was to explore the potential for tRNA gene, or tDNA, barriers to enhance transgene expression and recombinant protein production in *K. phaffii*.

3.1.2 Chromatin states and transcriptional activity

Transcription rates are ultimately determined by the activity of RNA polymerase II (Pol II), an enzyme which is recruited by transcription factors, bind to a promoter and transcribes a gene into messenger RNA (mRNA) (154). Binding of these factors is determined by their ability to access their respective binding sites on the DNA strand (155).

DNA is wrapped around protein complexes called histone octamers (156-158). The core histones in the octameric complex, H2A, H2B, H3 and H4, are wrapped by 147 bp of DNA to form the nucleosome (159). A collection of nucleosomes is called chromatin (157, 160). Chromatin is often described as looking like balls on a string, whereby nucleosomes can slide up and down the DNA strand and are incorporated or evicted during gene regulation (161, 162).

Chromatin can be classified into two different types: euchromatin and heterochromatin (Figure 3.1) (157, 160). The former is considered transcriptionally active as it consists of loosely packed and/or widely spaced nucleosomes and so transcription factors can more readily access their binding sites (163). The latter is considered transcriptionally inactive as it consists of tightly packed/densely spaced nucleosomes which impede the binding of transcription factors (163). Transcription of a gene is therefore determined by chromatin state and nucleosome architecture at the promoter, whereby a transition from euchromatin to heterochromatin can inactivate the expression of a gene (157, 160).

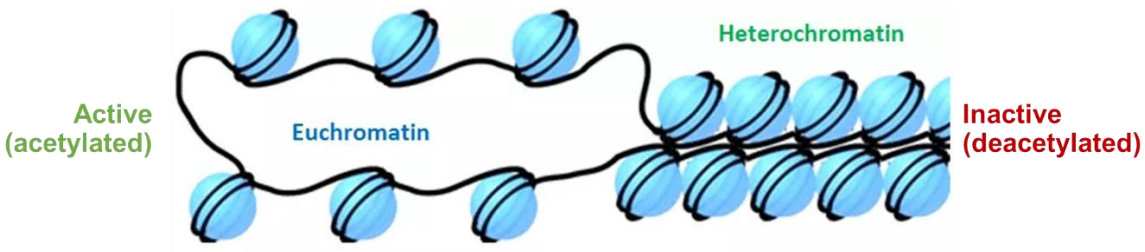


Figure 3.1: Schematic representation of euchromatin and heterochromatin.

3.1.3 Changing chromatin state

Changes in chromatin state are caused by the post-translational addition of small chemical modifications of histones, such as the addition of acetyl and methyl groups, which change the properties of the histones (157, 164, 165). For example, acetylation neutralizes the positive charge of histones, which causes them to drift away from DNA, which has a negative charge (166-169). Therefore, acetylation of histones is typically associated with euchromatin and activation of transcriptional activity. Conversely, deacetylation of histones is associated with transcriptional silencing, such as on lysine 9 of histone H3 (also known as H3K9ac) (166-169). Modifications occur on the N terminal tails of the histones, and they can be modified on several residues, so it has been hypothesised that it is combinations of histone modifications at different genes that cause the complex array of biological outcomes (170-173).

There are numerous proteins involved in chromatin remodelling. Some modify histones, such as histone deacetylases (HDACs), which remove acetyl groups from the lysine residues, which can lead to a compaction of chromatin and the repression of transcription (173-175). Others modify the nucleosome architecture, such as the ATP-dependent RSC complex (Remodelling the Structure of Chromatin), which moves, evicts or rearranges nucleosomes on the DNA strand (165, 176). Proteins such as these respond to environmental triggers to cause dynamic changes in the epigenetic states of many genes and coordinate transcriptional responses (151, 177, 178).

Not all chromatin states are dynamic, however, as heterochromatin exists in both constitutive and facultative forms (179, 180). Facultative chromatin is more dynamic, as it can switch between inactive and active states depending on environmental cues (147, 181). Constitutive heterochromatin is considered permanent, as these regions remain transcriptionally inactive throughout the lifespan of the cell (70, 160, 173, 182). Centromeres and telomeres are constitutively heterochromatic and maintenance of this is vital for cell survival (183, 184).

Constitutive heterochromatin typically contains non-coding DNA such as repetitive satellite DNA (185, 186). It is marked by specific histone modifications such as trimethylation of histone H3 at lysine 9 (H3K9me3), which not only compacts the DNA but recruits epigenetic silencers which help to maintain the constitutive state (182, 184, 187). Such modifications can propagate down the DNA strand and spread into euchromatic regions, causing transcriptional silencing (157, 182). Therefore, epigenetic barriers are required to demark the boundaries between heterochromatin and euchromatin, maintaining silencing in some regions and active transcription in others (188, 189).

The notion that the expression of a gene is influenced by its chromosomal location, particularly when it is relocated near heterochromatic regions of the genome, is known as position effect variegation (PEV) (160, 190). First identified in Drosophila, it was observed that when a euchromatic gene was placed adjacent to centromeric and telomeric heterochromatin, expression of the gene was variegated; being active in some cells and silent in others (190). It was subsequently observed that the same phenomena occurs in *S. cerevisiae* and *S. pombe (190).* For example, the *S. cerevisiae URA3* gene is constitutively expressed at a basal level.

3.1.4 The first tDNA barrier in *S. cerevisiae*

The first tRNA gene (tDNA) barrier was identified in 2001 by Donze and Kamakaka (107). The tDNA, a gene encoding a tRNA-Threonine, was found in the silent HMR (homothallic mating right) locus in *S.* cerevisiae, and was revealed to have an important role in mating-type switching (107). Investigations into the role this tDNA plays revealed that it had an critical function as a barrier, protecting neighbouring genes against the spread of heterochromatin (107).

The HMR locus is found upstream of the mating type-determining MAT locus on chromosome 3 (Figure 3.2). A silenced copy of the MATa allele is stored there in preparation for mating-type switching in haploid yeast cells (191). Genetic information within the MAT locus determines mating type, so mating-type switching is enabled when the MATa allele from the HMR locus, or conversely the MATα allele from the HML locus, is copied into the MAT locus (111, 192). Silencing, or transcriptional repression, of the MATa allele is mediated by the HMR-E silencer (109). This is achieved by the recruitment of silent information regulator (Sir) proteins SIR2, SIR3 and SIR4 (193-195). These bind to the tails of nucleosomes and form a multimeric compound which causes the condensation of the chromatin into transcriptionally silent heterochromatin, which has been found to spread towards neighbouring genes (110,

111). It was found that the downstream end of the HMR locus comprised of inaccessible chromatin, as determined by restriction enzyme accessibility studies (110). Removal of the HMR tRNA-threonine gene led to SIR-dependent repression of a URA3 gene inserted downstream of the deleted sequence (107, 196). Detailed characterisation of the tDNA revealed that the tDNA and its immediate flanking regions were both necessary and sufficient to prevent the spread of silencing from the *HMR* locus (107).

The discovery that tDNAs play a role in genome organisation, specifically in marking the boundaries between heterochromatin and euchromatin, revolutionised our understanding of epigenetics. Beyond simply encoding tRNAs which are vital for decoding mRNA, tRNA genes have an additional role in protecting the expression of nearby genes. This insight challenges our underestimation of gene placements and offers us a potential mechanism to exploit for improving transgene expression for the purposes of recombinant protein production.

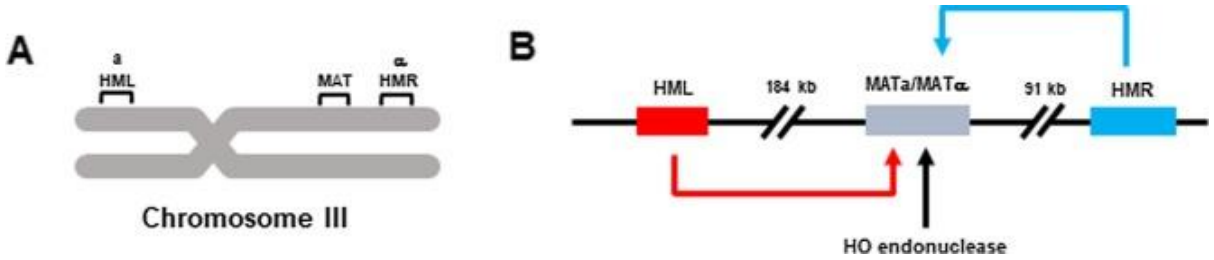


Figure 3.2: Mating type switching in *S. cerevisiae.*

(A) Chromosomal loci of the HML, MAT and HMR loci. (B) Schematic of the Mating-type switching mechanism. Mating-type switching is initiated by a site-specific HO endonuclease, which creates a double-strand break (DSB) in the MAT locus (192). The DSB is repaired by ectopic homologous recombination (gene conversion). One end of the DSB is coated in Rad51, which promotes strand invasion into a short region of homology shared by the MAT locus and the donor sequence. This is followed by copying of the mating-type-specific region, either a or α, and terminated by another short region of homology.

3.1.5 Mechanisms of barrier activity

Whilst the exact mechanism for tDNA barrier activity is not yet known, evidence points towards involvement of its transcriptional machinery. tDNAs are transcribed by RNA polymerase III (Pol III), which are recruited by transcription factors TFIIIB and TFIIIC (197). tDNAs contain internal promoter sequences, known as the A- and B-boxes, which are recognised by TFIIIC (197). TFIIIC then recruits TFIIIB, which binds upstream of the transcriptional start site. TFIIIB then assembles the Pol III complex at the transcriptional start site, and remains bound to the DNA to enable a high rate of transcriptional reinitiation (198, 199).

The SUP53 tRNA-Leu gene in *S. cerevisiae* is not naturally an effective barrier, but deletion of an intron which increases the distance between the A and B boxes by 42 bp causes a dramatic increase in barrier activity (107). It was hypothesised that decreasing the distance between the two TFIIIC binding sites facilitated TFIIIC binding, strongly implicating role of TFIIIC in the barrier mechanism. In concordance with this, introducing point mutations in the B-box, and to a lesser extent, the A- box, causes a loss of barrier activity by affecting TFIIIC binding (107). It is thought that the B box had a particularly significant role in barrier activity as deletion of the A boxes from two tDNAs was found to reduce TFIIIC affinity by only 2- to 5-fold, whereas single base B box substitutions resulted in decreases of 43- to 370-fold (200).

Due to the significant effects of manipulating transcription factor binding on barrier activity, it was thought that being transcriptionally active was a requirement of a strong barrier. However, not all tDNAs have barrier activity (107, 201), and not all tDNAs with barrier activity are transcribed (107). Indeed, a Pol III mutation that prevents transcription initiation has minimal effect on barrier function of the HMR tRNA-Threonine (107). This rules out the role of transcriptional activity in barrier activity, but it remains possible that transcription factor and/or Pol III recruitment is what drives barrier activity.

Further manipulations to tDNAs have revealed that the HMR tDNA barrier's flanking sequences contribute to its activity, although insufficient to reconstitute a barrier alone (107, 201). In fact, that tDNAs with little or no barrier activity can gain this function when their flanking sequences are replaced with sequences found at the native HMR tRNA-Threonine gene (107). It was thought that AT-rich flanking sequences resist nucleosome incorporation and aid with disrupting the propagation of chromatin remodelling activities (202). Supporting this, it was found that histone depletion is necessary for tDNA barrier activity (166). Amongst the most rapidly turned over nucleosomes in *S. cerevisiae* are those adjacent to tDNAs, where nucleosome occupancy is substantially depleted (201, 203-209).

Active nucleosome eviction is also thought to contribute to barrier activity. The RSC complex, a chromatin remodeller involved in maintaining nucleosome depletion, is detected at all tDNAs in *S. cerevisiae* (210). Isw1 and Isw2 remodellers are also detected at tDNAs (181, 211, 212). Deletion of rsc2, a subunit of the RSC complex, and Isw2 abolish and weakens barrier activity, respectively (201, 213). It could be that these create or uphold a gap in nucleosome occupancy which disfavours the spread of heterochromatin. Therefore, competition by either the transcriptional machinery and/or the action of chromatin remodellers with incorporated nucleosomes may contribute towards the formation of the barrier or assist in stabilising the barrier once established (172, 214).

Multimerisation, or the consecutive repeating of the DNA element, strengthens weak barriers (107). tDNA barriers are also often seen in clusters at the boundaries of heterochromatin (209, 215, 216). This suggests that strong barriers require continuous occupancy, as dissociation of barrier components from one tDNA may be compensated if an adjacent tDNA remains bound. Consecutive repeats of tDNAs may also construct a larger gap in the chromatin to further dissuade the propagation of silencing activities.

Of the components in the Pol III transcriptional machinery, TFIIIC seemingly has a particularly significant role in barrier activity. This is evidenced by the fact that upon isolating a B box, it still binds TFIIIC but fails to recruit TFIIIB and pol III, and is sufficient for barrier activity (214). Indeed, orphan B boxes were found in the *S. cerevisiae* genome and were functioning heterochromatin barriers (217). These were named 'extra TFIIIC sites' or ETC sites.

3.1.6 ETC sites

Upon discovery that TFIIIC is a major determinant of barrier activity, it was hypothesised that TFIIIC binding-sites could be sufficient as barriers. Genome-wide mapping of the Pol III machinery in *S. cerevisiae* identified eight loci which are bound by TFIIIC, but not by TFIIIB or Pol III (205). These were named Extra-TFIIIC (ETC) sites and were found to contain the B box promoter motif found in tDNAs but lack the A box (Figure 3.3). Of the eight ETC sites identified in *S. cerevisiae*, two demonstrated an ability to act as heterochromatin barriers as they were able to replace the canonical tRNA-Threonine gene and protect an ADE2 reporter gene from repression in the HMR locus (214). Further analysis of ETC loci revealed that in addition to the B-box sequence, ETC sites contain a further 10bp sequence conserved across different yeast species, but no functions have been discovered for this sequence (214).

S. pombe contains over 60 ETC sites (218). For example, an 500 bp ETC site containing 5 copies of the B-box motif can be found at the 2 kb identical inverted repeats which flank the silenced mating-type locus in *S. pombe* (217). ETC sites in *S. pombe* are also termed chromosome-organising clamp (COC) sites as they have been shown to tether to the nuclear periphery, in close proximity to the nucleolus (217). The presence of TFIIIC at these sites was confirmed by immunofluorescence (217). This finding prompted exploration of the three-dimensional localisation of TFIIIC, where it was found that tDNAs and ETC sites in *S. cerevisiae* also localise to the nuclear periphery in a TFIIIC-dependent manner (215, 216, 219). Whilst it would be intriguing to consider nucleolar association as a contributor to barrier activity, it was shown than an *S. cerevisiae* ETC site retained its barrier activity after release from nuclear periphery localisation (219). However, these observations inspired models of genome organization involving TFIIIC (217).

Three-dimensional clustering of TFIIIC is consistent with evidence that tDNAs in *S. cerevisiae* localise to the boundary of the nucleolus (215, 216). Similarly, six of the eight ETC sites in *S. cerevisiae* are localised at the nuclear periphery, positioning that is lost if the B box is mutated or if TFIIIC is targeted for specific degradation (219). Interestingly, tethering of an ectopic chromosomal locus to the nuclear periphery can be induced by the insertion of an ETC or anchoring of TFIIIC to the target site (219). Therefore, it is likely that introduction of a tDNA will cause tethering of the genomic region to the nuclear periphery.

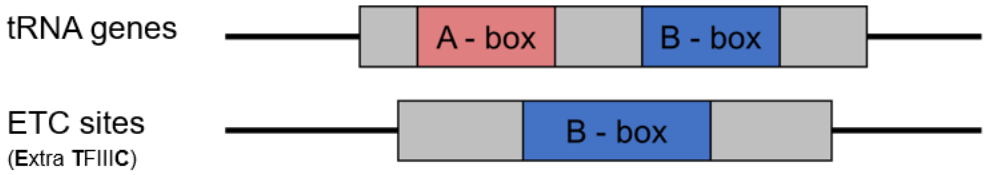


Figure 3.3: Internal promoter elements of tRNA genes and ETC sites.

Research characterising tDNA barriers establish that TFIIIC binding sites are the minimum requirements for barrier activity, but many other components and variables have a role to play in this phenomenon. It remains to be determined how these components uphold boundaries between heterochromatin and euchromatin, but the findings so far are sufficient for successful barriers to be selected and applied in an industrial context. Determining the minimal requirements of a barrier and dissecting its mechanism allows researchers to optimise barriers for increased efficacy, design barriers from scratch and simplify them for easy characterisation. This is especially important in an industrial context as patenting the innovation and incorporating it into standard expression vectors will mean that the genomic element that is developed for use will undergo high levels of scrutiny to meet regulatory or client approval.

3.1.7 Transcription at the *AOX1* locus

In order to investigate the effect of a tDNA barrier on transgene transcription in *K. phaffii* expression platforms, it is imperative to consider the known variables affecting transcription at the *AOX1* locus. Transgenes expressed under the *AOX1* promoter (pAOX1) are induced by the addition of methanol to the *K. phaffii* culture (14, 78). Methanol enters the cell passively and is detected by the cytosolic receptor and transcription factor, Mxr1 (78, 220). Cells can detect and respond to different carbon sources such as methanol at a transcriptional level due to cytosolic receptors detecting the molecule and causing a signalling cascade known as the methanol utilisation or MUT pathway (85, 221).

Depending on the conditions which the cell is under, the *AOX1* promoter undergoes three phases: repression, derepression and activation (222). During growth on glycerol or glucose, the promoter is repressed via the binding of Nrg1, Mig1 and Mig2, which prevents transgene expression while biomass accumulates (223, 224) (Figure 3.4). Proteins such as hexose sensor Gss1 and hexose transporter Hxt1 are also reported to play a role in repression of pAOX1 as their deletion led to the derepression of pAOX1 in response to glucose (225, 226).

During a cultivation run, derepression of pAOX1 occurs as glycerol is depleted, resulting in leaky/uninduced activation of the promoter (227). This is evidenced by *AOX1* mRNA levels being approximately 1-2% of the induced level during derepression (227). In response to methanol inductions, transcription factors Mxr1, Mig1 and Mit1 are the main transcription factors involved in methanol-dependent activation of transcription under the *AOX1* promoter (220, 221). Following the recruitment of transcription factors, RNA polymerase II (Pol II) can bind to pAOX1 and initiate transcription (220, 221). Therefore, epigenetic factors affecting the accessibility of pAOX1 to these transcription factors could affect transgene expression and recombinant protein production.

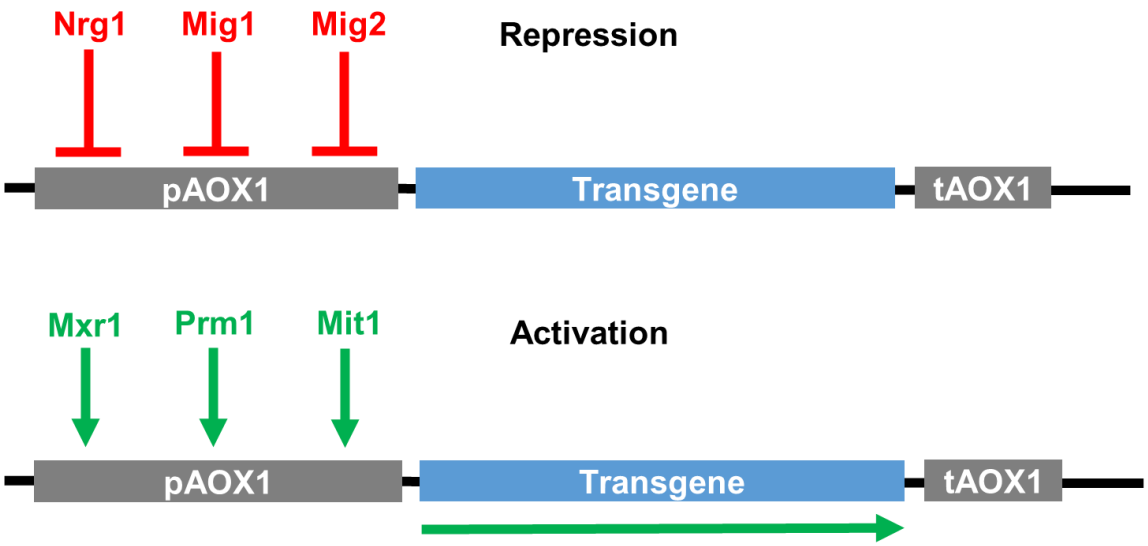


Figure 3.4: Transcription factors affecting expression of genes under the *AOX1* promoter.

3.1.8 Epigenetic regulation in recombinant protein production in *K. phaffii*

The current literature surrounding the epigenetic regulation of the *AOX1* locus is sparse, highlighting an important knowledge gap which once filled could inform innovative work to optimise the epigenetic environment around transgenes and improve their expression. However, some studies have tested certain interventions to improve recombinant protein production in *K. phaffii* which provide insights into the epigenetic variables affecting the *AOX1* locus.

Aiming to implement a novel approach to optimising the pAOX1 sequence, Yang et al demonstrated that the strength of pAOX1 can be increased by manipulating poly (dA:dT) tracts, stretches of adenosine and thymine nucleotide bases, in the primary sequence (228). AT-rich regions are known to be nucleosome resistant as the DNA becomes inflexible which significantly decreases the efficiency of nucleosome incorporation (229, 230). Thus, they hypothesised that introducing these tracts will decrease nucleosome occupancy at certain sites within the promoter and force nucleosome incorporation at other sites within the promoter. Successfully, they generated variants of pAOX1 where intracellular porcine growth hormone (pGH) and Lac Z production were enhanced between 0.25 and 3.5 fold (228). This was found in strains where more poly (dA:dT) tracts were introduced into the *AOX1* promoter (228). This tells us that in the wildtype pAOX1, nucleosome architecture is limiting to transgene expression and recombinant protein production. It has been shown that a single nucleosome positioned at the promoter region can block Pol II loading and formation of the pre-initiation complex (PIC), thereby impeding transcription initiation (231). The RSC complex is responsible for rearranging and evicting nucleosomes within gene promoters to aid transcriptional initiation so introducing poly (dA:dT) tracts may assist RSC in this task (206).

Beyond this study, other works have provided insights into how global epigenetic regulation could be optimised for improvements in recombinant protein production. A study which conducted directed evolution of *K. phaffii* and generated a point mutation in a subunit of the RSC complex (Rsc1-G22V) found that this mutation improves cellulase production, as evidenced by a 20% increase in cellulase yield (21). This suggests that nucleosome architecture and chromatin structure of genes involved in recombinant protein production can have a limiting effect on yields but can be alleviated by modulating RSC activity. This work was limited in that it did not investigate the effect of this mutation on RSC activity and determine its impact on pAOX1 and transgene transcript levels. Studying this would reveal which bottleneck was alleviated by the mutation in Rsc1.

3.1.9 Techniques for measuring epigenetic changes

Numerous well-established techniques are available to measure epigenetic changes in model organisms such as *S. cerevisiae*, where methods for assessing DNA and histone modifications, and chromatin accessibility are highly developed and widely used. Some of these protocols have been adapted in *K. phaffii*, enabling the procurement of data which has open avenues to explore the unique epigenetic landscape within this understudied system. The techniques which have been adapted in *K. phaffii*, and the insights they provide are outlined below.

Chromatin immunoprecipitation (ChIP) assay

Chromatin immunoprecipitation (ChIP) is a powerful technique which allows individual histone PTMs to be mapped to unique DNA sequences (232). Paired with high-throughput sequencing, ChIP-seq provides a detailed profile of histone modifications and nucleosome positioning that is essential for a full understanding of epigenetic regulatory mechanisms (232). This technique is achieved by cross-linking DNA-binding proteins to the DNA *in vivo* through treating the cells with formaldehyde. The chromatin is therefore fixed, creating an association between the DNA sequence and the modified histones. The chromatin is then sheared to facilitate its analysis using sonication. Antibodies raised to the protein of interest, typically a modified histone such as histone H3 lysine 4 trimethylation (H3K4me3), are then used to immunoprecipitate specific DNA-protein complexes. The crosslinks are then reversed and the released DNA is analysed by sequencing or qPCR to determine the level of enrichment of a specific protein on a DNA sequence (232).

The specificity of the antibody is vital for the success of this assay, so resource needs to be devoted to developing an array of antibodies for use in *K. phaffii.* A requirement for this technique is the development of antibodies which specifically recognise these modified histones. As determined from the current literature, it appears that no antibodies have been developed for modified histones in *K. phaffii*.

This issue was circumvented in Coughlan et al, where *K. phaffii* centromeres were mapped using ChIP-seq against centromeric nucleosomes containing CenH3 (Cse4) (70). They generated strains containing *CSE4* gene with a 3xHA (haemagglutinin) tag inserted between amino acids 41 and 42, as their initial attempt to tag *CSE4* at the C-terminus yielded no viable clones (70). Developing strains with protein tags on histones can yield numerous issues such as altering histone properties and functioning, limiting sensitivity and specificity of the assay and increased complexity associated with molecular engineering. Therefore, it remains a significant limitation in *K phaffii* research that there is a lack of specific antibodies available for chromatin immunoprecipitation (ChIP) assays.

Micrococcal nuclease (MNAse) assay

Micrococcal nuclease (MNase) assays have been a core technique for investigating nucleosome architecture since the 1970s (233). Combined with sequencing, it enables the genome-wide mapping of nucleosome occupancy and providing invaluable insights into the accessibility of chromatin (234). This technique involves assaying for the sensitivity of genomic regions to MNase, an enzyme which digests DNA. Nucleosome-free DNA is preferentially digested by MNase compared to nucleosome-bound DNA. After treating isolated chromatin with MNase, only nucleosome-protected DNA remains intact. Removal of the associated-proteins and analysis of the remaining DNA reveals which areas of the genome are occupied by MNase-resistant nucleosomes and indicates areas of low chromatin accessibility.

In *K. phaffii,* MNase assays revealed that the DNA ends of nucleosome are somewhat more accessible, or MNase-sensitive, compared to those of the human nucleosome (235), giving an indication of the degree of flexibility of nucleosome positioning in *K. phaffii*. A limitation of this technique, however, is that it relies on MNase enzymatic activity which may not only be variable but also preferential, as it was shown that A/T-rich nucleosomes are digested faster than G/C-rich nucleosomes (234). Therefore, it is preferrable to use a technique which is not only independent of antibodies, but of enzymes which may provide a skewed indication of nucleosome occupancy.

Assay for Transposase-Accessible Chromatin (ATAC)

Assay for Transposase-Accessible Chromatin (ATAC) is a technique used to determine chromatin accessibility, using the hyperactive transposase Tn5. The procedure involves loading next-generation sequencing (NGS) adapters onto the transposase, allowing the simultaneous fragmentation of chromatin and the integration of the adaptors into open chromatin regions. This generates a library that can be sequenced via NGS, preferentially sequencing regions of the genome with open or accessible chromatin (236).

ATAC-seq lends itself well for mapping chromatin accessibility throughout the entire genome. It was utilised in *K. phaffii* production systems by Brady et al, where new transgene integration sites were identified, using chromatin accessibility data to flag highly accessible regions, which is associated with increased transcriptional activity (236). Optimal sites were predicted using ATAC-seq paired with RNA-seq, and these were characterised, revealing that there are a plethora of promising integration sites beyond the typical *AOX1* and GAP loci (236).

Being independent of biased enzymes and specific antibodies is an advantage of this technique. However, whilst it is particularly powerful, it is also relatively resource intensive, requiring the coupling of high-throughput sequencing. It provides a breadth of data, mapping genome-wide chromatin accessibility, but is not easily implemented in time-course experiments, limiting its ability to study epigenetic changes over time. Therefore, ATAC-seq data was probed using bioinformatics for chromatin accessibility at tDNAs on a genome wide scale, but was not carried forward for experimental implementation.

Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) assay

Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) is another technique for mapping chromatin accessibility, instead using the biochemical properties of protein-bound DNA to separate nucleosome-depleted regions in the genome (237) (Figure 3.5). Like in ChIP assays, cells are subjected to cross-linking, ensuring that the interaction between the nucleosomes and DNA are fixed. Chromatin is sonicated, fragmenting the DNA which can then be separated using a phenol-chloroform extraction. During phenol-chloroform extraction, the chromatin mixture is separated into two phases; an organic and an aqueous phase. DNA fragments cross-linked to nucleosomes will preferentially sit in the organic phase, whereas nucleosome-depleted or 'open' regions will be found in the aqueous phase. By specifically extracting the aqueous phase, only nucleosome-depleted regions will be purified and enriched.

FAIRE-seq data was generated by De et al in order to investigate pseudohyphal growth in *Komagataella phaffii* and the involvement of the flocculin (*FLO*) gene family in its regulation (177). It was found that the expression and repression of FLO400 and FLO5-1 correlated closely with open or closed chromatin regions upstream of these genes, respectively. This indicated that these regions underwent heritable changes in chromatin structure, suggesting an underlying epigenetic mechanism involved in regulating the flocculation phenotype in response to environmental stressors (177).

This simple but powerful technique allows for *K. phaffii* chromatin accessibility to be assayed without the need for specific antibodies or enzymes. In De et al, this technique was combined with sequencing to provide genome-wide data, which was utilised in this work to map chromatin accessibility at and around tDNAs (177). This technique was also adapted to become more high-throughput and combined with quantitative PCR to enable time-course assays to be conducted in order to monitor chromatin accessibility over time.

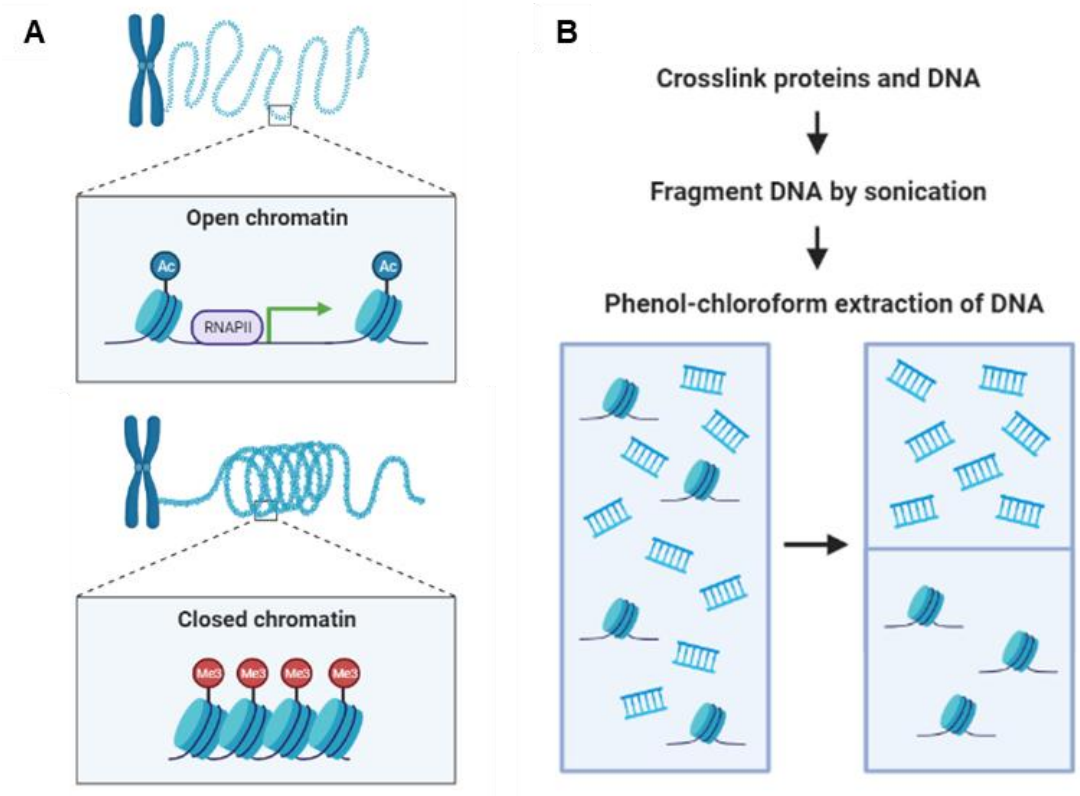


Figure 3.5: Schematic representation of the principles of Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE).

a) Chromatin exists in two states; open chromatin/euchromatin and closed chromatin/heterochromatin. Open chromatin is associated with increased transcriptional activity. b) Overview of the FAIRE procedure. Assaying for chromatin accessibility using FAIRE uses properties of phenol and chloroform to separate nucleosome-bound and nucleosome-free DNA after sonication-induced fragmentation of chromatin.

Summary

The range of techniques available to assay for epigenetic marks and chromatin structure are plentiful but have yet to be sufficiently utilised in the *K. phaffii* expression platform. Key studies highlighted above demonstrate the potential that they have to build our understanding of epigenetic in *K. phaffii* and inspired some of the approaches taken in this thesis. Genome-wide mapping of chromatin accessibility using FAIRE-seq and ATAC-seq data was utilised in this thesis to investigate the effect of tRNA genes on the chromatin environment. Considering the advantages and disadvantages of each technique, FAIRE was then selected for implementation in experiments investigating chromatin accessibility within the strains developed in this study. The simplicity of the assay, requiring no antibodies which are yet to be made commercially available for *K. phaffii* histone modifications, facilitates its easy implementation. Being enzyme-independent removes extraneous effects associated with sequence-biases. All in all, this technique lends itself well towards high reproducibility and therefore enables robust conclusions to be drawn from the data.

3.1.10 Aims and objectives

Feeding into the larger goal of investigating applications of tRNA genes (tDNAs) in improving recombinant protein production, this chapter aims to investigate the barrier phenomenon in *K. phaffii* expression platforms. It first aims to investigate whether such phenomenon exists in *K. phaffii*, and whether there is a potential for this mechanism to be exploited in an industrial context. It then aims to investigate whether epigenetic regulation is a limiting factor in *K. phaffii* expression platforms, first by monitoring transcriptional and epigenetic changes over time, and then by introducing a tRNA with the potential evoke barrier activity and improve transgene expression at the *AOX1* locus.

The objectives of this study are as follows:

- To determine whether barrier activity is conserved in *K. phaffii*
- To test *K. phaffii* candidates for barrier activity
- To investigate factors which contribute towards barrier activity
- To determine whether the epigenetic environment of the transgene is limiting for recombinant protein production in *K. phaffii*
- To develop an integrative vector to insert a tDNA sequence upstream of the *AOX1* locus
- To evaluate the impact of this tDNA on transgene transcription, promoter accessibility, and recombinant protein production

3.2. Results

3.2.1 Chromatin accessibility data indicates the presence of endogenous barriers in *K. phaffii*

In order to investigate whether tDNA barriers exist in *K. phaffii*, chromatin accessibility data was mapped across the genome to identify tDNAs which fall on the boundary of heterochromatin and euchromatin. This was done under the assumption that since tDNA barriers function to prevent the spread of heterochromatin to euchromatin, they are highly likely to be found on the boundaries between the two domains.

ATAC-seq data by Brady et al which maps genome-wide chromatin accessibility was used to generate a heatmap centred around each tDNA in the *K. phaffii* genome. A significant subset of tDNAs were found on the boundaries of accessible regions (Figure 3.6) (Table 13). tDNAs placed on the boundaries between chromatin domains implicate a role of tDNAs in genome organisation and is especially reminiscent of a heterochromatin barrier. Therefore, this data implicates the existence of endogenous heterochromatin barriers and the conservation of the barrier mechanism.

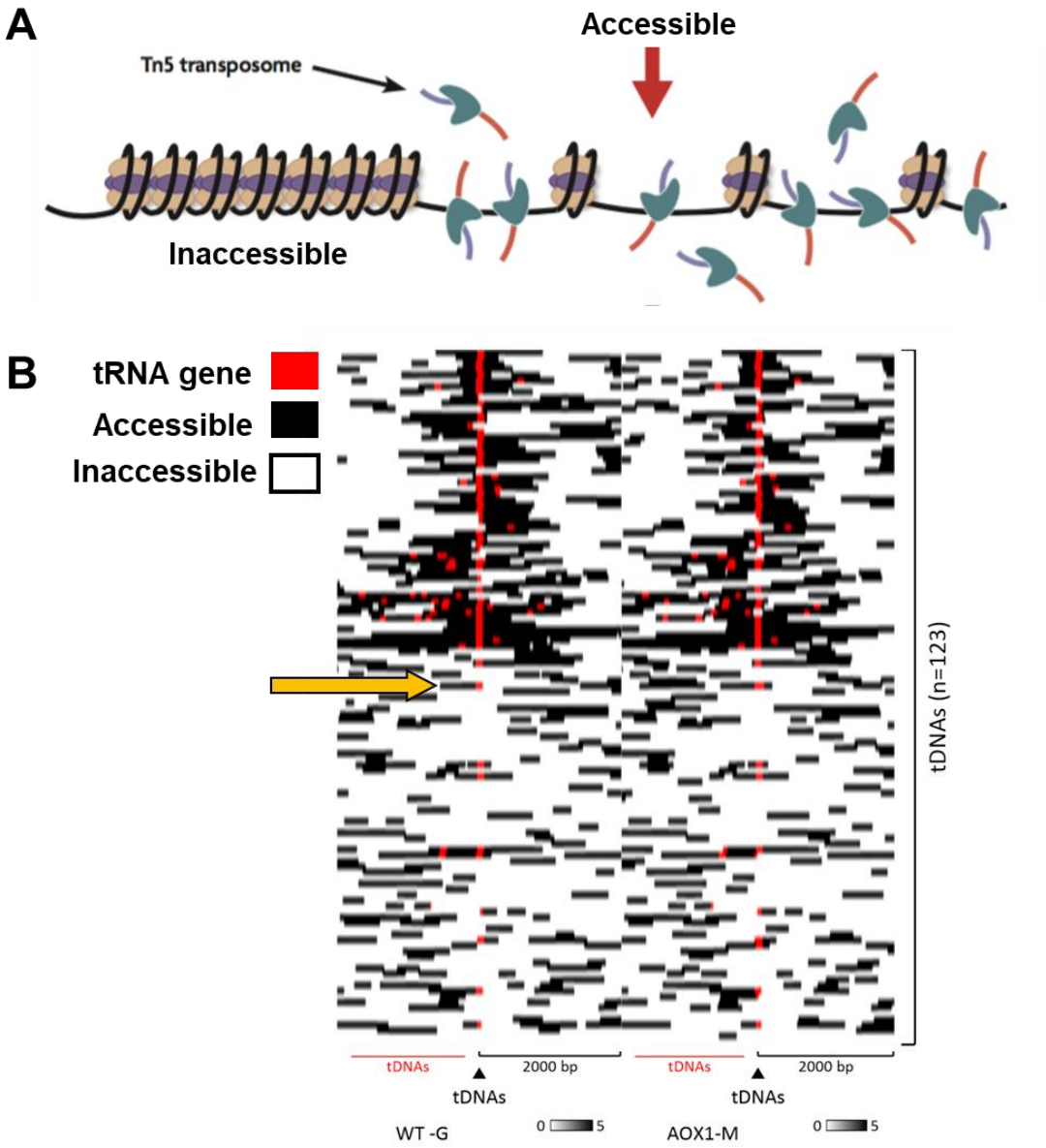


Figure 3.6: Heatmap of chromatin accessibility within a 2kb region surrounding *K. phaffii* tDNAs.

Red regions indicate tRNA genes, or tDNAs. Black regions indicate chromatin regions accessible to the Tn5 transposase and thus accessible to transcription factors. White regions indicate chromatin regions that are inaccessible to the Tn5 transposase and are therefore not accessible to transcription factors. ATAC-seq data was obtained from the NCBI Gene Expression Omnibus (accession number: GSE154330).

3.2.2 Boundary tDNAs are more accessible than non-boundary tDNAs.

To differentiate boundary tDNAs from non-boundary tDNAs and to assess whether boundary tDNAs are likely functioning as barriers, their chromatin accessibility was compared. Figure 3.7 shows that boundary tDNAs appeared to be disproportionately more accessible, according to both ATAC-seq and FAIRE-seq datasets.

It would be expected that barrier tDNAs would have increased accessibility due to the fact that barrier function would require the constitutive recruitment of TFIIIC and other proteins required to uphold the separation between heterochromatin and euchromatin. Thus, nucleosomes over the tDNA would be evicted, creating a nucleosome free region. In contrast, non-barrier tDNAs may not require such constant accessibility, as they are not involved in maintaining chromatin boundaries.

The significant difference in accessibility observed in Figure 3.7 therefore supports the distinct properties of boundary tDNAs compared to non-boundary tDNAs, supporting the notion that boundary tDNAs could bind TFIIIC and have functioning epigenetic barrier activity.

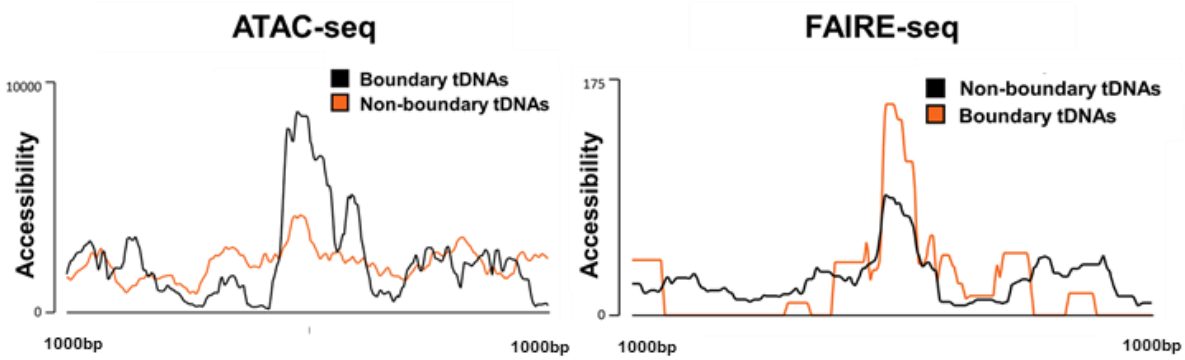


Figure 3.7: Comparison of chromatin accessibility over boundary and non-boundary tDNAs.

ATAC-seq (Accession ID: GSE154330) and FAIRE-seq (Accession ID: GSE131290) datasets were obtained from the NCBI Gene Expression Omnibus, and was used to plot chromatin accessibility over tDNAs and 1000 bp flanking regions.

3.2.3 Adoption of barrier assay in *Saccharomyces cerevisiae*

Having established that there is evidence supporting the presence of tDNA barriers in *S. cerevisiae*, it became imperative to directly test *K. phaffii* tDNAs for barrier activity. David Donze's barrier assay uses the expression of a *S. cerevisiae* mating-type gene, the a1 gene, to assess the strength of putative barriers (107). Strains which express the α1 mating type are transformed with plasmids containing a potential barrier in between the HMR-E silencer and a1 gene (Figure 3.8). When the potential barrier has weak barrier activity, the a1 is silenced by the HMR-E silencer and only the α1 mating type is expressed. Therefore, the strain is able to mate with the mating lawn strain, which expresses the a1 mating-type, and grow on selective media. However, strong barriers will prevent the spread of heterochromatin and therefore prevent the silencing of the a1 gene so that the strain will express both mating types. This means that it will be unable to mate with the mating lawn strain and will be unable to grow on the same selective media. Barrier activity can therefore be measured through the strain's ability to mate, which can be observed by the number of daughter colonies that grow on selective media.

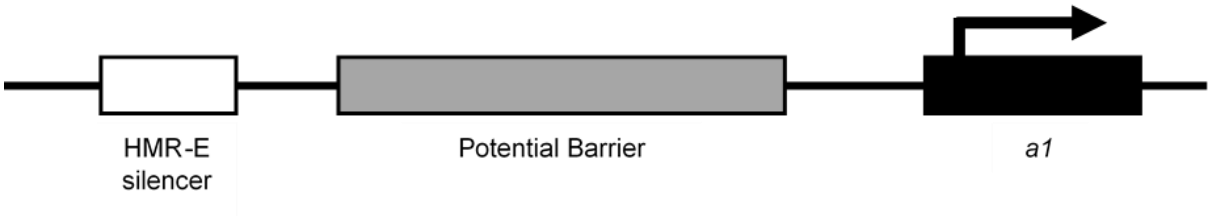


Figure 3.8: Map of the mating-type region in Donze's barrier assay.

The potential barrier will be tested for its ability to inhibit silencing of the a1 gene. It will be placed between the HMR-E silencer region (where heterochromatin formation is initiated) and the a1 gene and expression of the a1 gene will be observed.

The mating assay was successfully adopted, as evidenced by demonstrating robust barrier activity in the canonical *S. cerevisiae* tRNA threonine gene (Figure 3.9). Barrier activity was detected by its ability to allow expression of the a1 mating type gene, which in turn reduced the cell's capacity to mate on the lawn of alpha-mating cells and produce progeny which can grow on selective media. This established a reliable system for detecting barrier activity, as mating efficiency was affected by the presence of the *S. cerevisiae* tRNA gene. With this setup, the assay provided a foundation for testing *K. phaffii* genes by replacing the *S. cerevisiae* tRNA gene with various *K. phaffii* gene candidates.

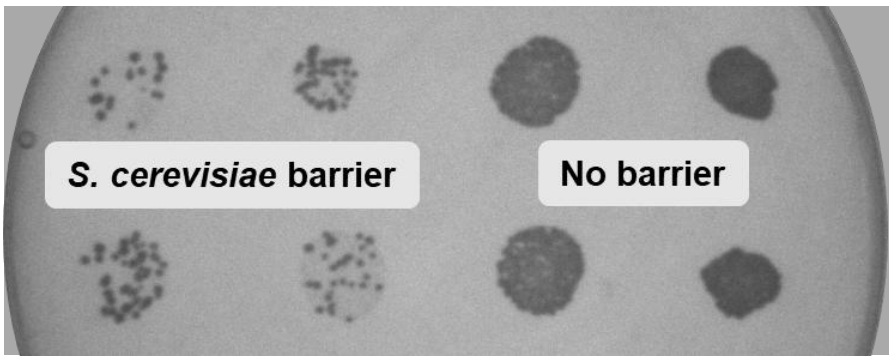


Figure 3.9: Successful adoption of the *S. cerevisiae* mating assay for barrier activity.

The *S. cerevisiae* tRNA-threonine gene was used in the *S. cerevisiae* mating assay for barrier activity. The tRNA-threonine gene was a strong barrier, so prevented silencing of the a1 mating type gene from the HMR-E silencer. This caused a non-mating phenotype due to simultaneous expression of the alpha mating type gene, so it prevented mating with cells on the mating lawn and significantly reduced the amount of progeny that could grow on YMD-URA-HIS plates.

3.2.4 *K. phaffii* tRNA^{Thr} genes were identified as strong heterochromatin barriers in *S. cerevisiae*

Donze and Kamakaka discovered that out of the sixteen tRNA threonine (tRNA^{Thr}) genes in *Saccharomyces cerevisiae,* only one demonstrated strong barrier activity (6). This raised the question of whether any of the seven *K. phaffii* tRNA^{Thr} genes were capable of also acting as heterochromatin barriers.

Six of the seven tRNA^{Thr} genes were cloned into the barrier plasmid between the HMR-E silencer and a1 gene in with 250bp flanking sequences on either side of the gene. It was found that the seventh tRNA^{Thr} gene was found in a tRNA-threonine and tRNA-glutamate gene pair *in situ*, so these were cloned into the barrier plasmid as a pair. Strains transformed with these plasmids were spotted alongside the 'no barrier' negative control and the *S. cerevisiae* tRNA threonine (tRNA^{Thr}), which acted as a positive control.

Lack of growth on YMD-URA-HIS plates indicates that mating did not occur in any of the strains containing the *K. phaffii* tDNAs, suggesting that these tDNAs were effective at preventing a1 silencing, and thus acted as effective epigenetic barriers (Figure 3.10). This finding is particularly remarkable as it is the first heterochromatin barriers to be identified in *K. phaffii*. Further to this, it demonstrates that *K. phaffii* tDNAs can recruit the components required for barrier activity in *S. cerevisiae,* revealing a cross-species adaptation of a conserved epigenetic mechanism.

Not only did all the tDNA candidates demonstrate barrier activity, but they also exhibited varying activity strengths. Three of the barriers; ch1tRNA37, ch3tRNA14 and ch3tRNA27, had a level of barrier activity comparable to that of the positive control, but interestingly, four of the barriers; ch2tRNA8, ch3tRNA5, ch3tRNA9 and the tDNA pair greatly outperformed the positive control. This finding suggests that the chromatin environment of *K. phaffii*, which is largely uncharacterised to date, might have driven tDNAs to evolve more robust barrier activities than those of tDNAs in *S. cerevisiae.*

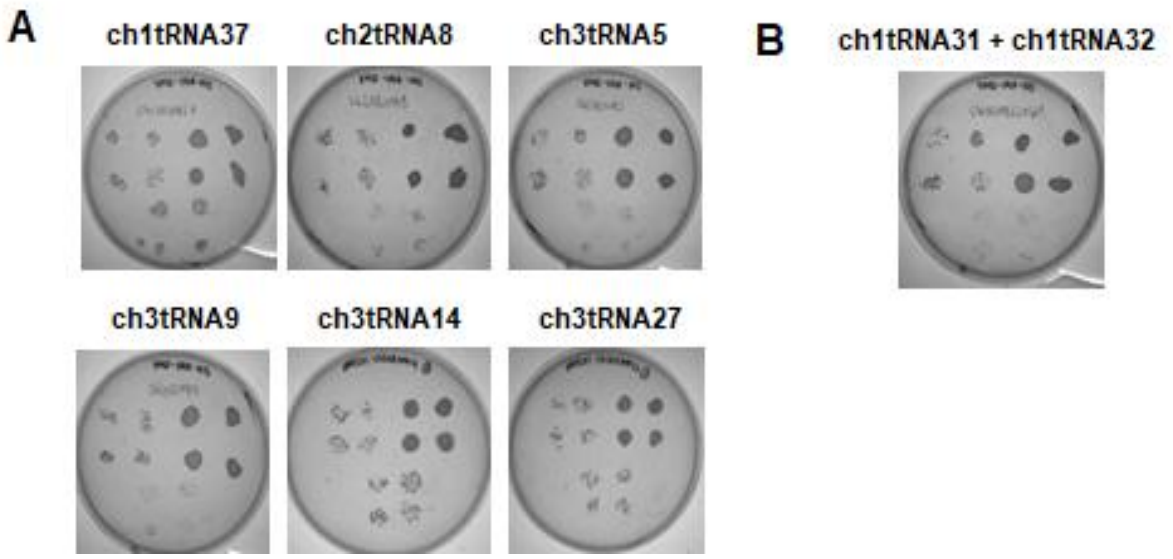


Figure 3.10: *K. phaffii* tDNAs act as heterochromatin barriers in *S. cerevisiae*.

(A) Mating assay plates demonstrating that *K. phaffii* tDNAs display strong barrier activity in *S. cerevisiae*. Lack of colony formation on YMD-URA-HIS plates indicates that the tRNA-Threonine genes effectively prevented silencing of the a1 gene from the *S. cerevisiae* HMR-E silencer. Each plate shows tests in quadruplicate for cells with the *S. cerevisiae* HMR tRNA-Threonine gene (top left), no barrier (top right), or a *K. phaffii* tRNA-Threonine gene (bottom) inserted between a1 and the HMR-E silencer. Identity of the *K. phaffii* tRNA-Threonine genes is indicated above each plate. (B) As in A, except using a fragment of *K. phaffii* DNA carrying closely spaced tRNA-Threonine and tRNA-Glutamate genes.

3.2.5 Identification of putative centromeric ETC sites in *K. phaffii*

Previously identified heterochromatin barriers in yeast species not only consisted of tRNA genes, but also extra-TFIIIC (ETC) sites, which share a common motif with tDNAs known as the B box (217). It is thought that barrier activity is shared by these elements due to the B-box being a binding of TFIIIC (217). To investigate whether ETC sites with barrier activity exist in *K. phaffii*, genome analysis was conducted to identify putative ETC sites. As ETC sites are not yet annotated in *K. phaffii*, these could not be analysed in a genome-wide screen using ATAC-seq and FAIRE-seq datasets as was conducted for *K. phaffii* tDNAs. Two consensus sequences were assessed for their ability to identify ETC sites.

Moqtaderi and Struhl (2004) analysed the evolutionary conservation of B blocks and ETC loci across species of yeast by aligning sequences to the B block consensus derived from the analysis of 274 *S. cerevisiae* tRNAs (10). They identified a 6bp core sequence ‘TTCGAA’ which was used to identify ETC sites within the genome of the wildtype *K. phaffii* strain, CBS7435. Using the Regular Expression (Regex) package in Python, 7116 instances of this 6bp core were identified as shown in Figure 3.11. To narrow down the candidates, searching instead for a 15bp sequence 'TTCGAANCCNNNNNG' suggested 54 putative TFIIIC binding sites within *K. phaffii.* The core consensus sequences aligned well with the successful tDNA barriers identified previously in this study (Figure 3.12).

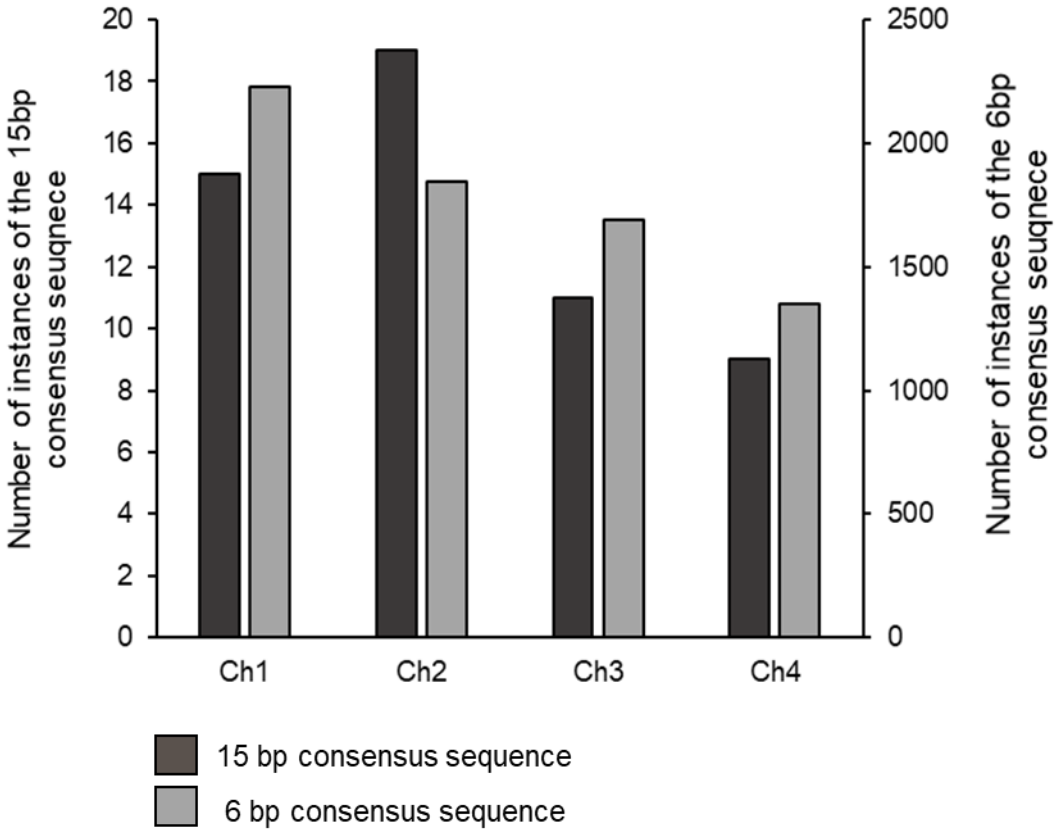


Figure 3.11: Number of potential ETC sites identified in each chromosome of the *K. phaffii* strain CBS7435 using two variations of the B box consensus sequence.

Each chromosome of the *K. phaffii* genome was searched for the number of instances of each consensus sequence using SnapGene Viewer. Using the longer consensus sequence vastly narrowed down the number of potential ETC sites, helping to increase the likelihood of finding an ETC site that functions as a heterochromatin barrier.

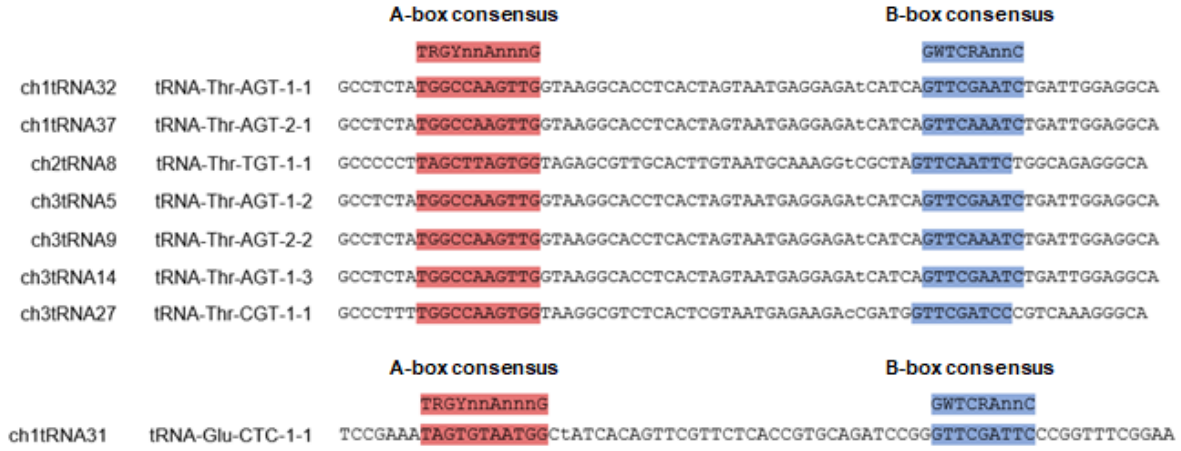


Figure 3.12: Alignment of A and B box consensus sequences to tDNAs tested in *S. cerevisiae* barrier assay.

Seven tRNA-Threonine alignments with yeast A- and B-box consensus motifs and one tRNA-Glutamate alignments with yeast A- and B-box consensus motifs. Consensus sequences TRGYnnAnnnG (A-box) and GWTCRAnnC (B-box) were identified by Marck et al (238) through global distance analysis of all tDNA sequences in 10 different yeast species; Saccharomyces *cerevisiae, Schizosaccharomyces pombe, Saccharomyces castellii, Candida glabrata, Kluyveromyces waltii, Kluyveromyces lactis, Eremothecium gossypii, Debaryomyces hansenii, Candida albicans* and *Yarrowia lipolytica*.

3.2.6 *K. phaffii* orphan A- and B-boxes did not exhibit barrier activity in *S. cerevisiae*.

To further narrow down candidates to be tested in the *S. cerevisiae* mating assay, putative ETC sites were not only selected by considering the sequence similarity to the 15bp core consensus sequence. Chosen candidates were also selected based on their proximity to the centromere, which is a known silenced region and therefore would require heterochromatin barriers to prevent silencing of nearby genes. The centromere has also previously been associated with strong barrier elements, as tRNA genes found here in *Schizosaccharomyces pombe* have been shown to be strong barriers (7, 9, 10).

Four candidates, Ch1.cenETC1, Ch1.cenETC2, Ch3.cenETC1 and Ch3.cenETC2&3. Ch1.cenETC1, were chosen for their varying properties (Figure 3.13). Ch1.cenETC1 contains a single B box as identified by the 15bp consensus sequence and was also cross-validated by another 9bp consensus sequence. Ch3.cenETC2&3 also contained the 15bp consensus sequence, and when cross-validated, contained one 9bp consensus and two 6bp consensus sequences, increasing our confidence of a functioning B-box being present. Ch1.cenETC2 and Ch3.cenETC1 both contained cross-validated B-box consensus sequences, and also contain an A-box, as identified by an 11bp consensus sequence. Containing both an A-box and a B-box means that these two candidates are most similar to tDNAs, which exhibited strong barrier activity in the *S. cerevisiae* assay.

Interestingly, despite containing cross-validated B-box sequences and A-box sequences, none of the candidates exhibited barrier activity, as demonstrated by growth of progeny on selective media (Figure 3.14). A number of conclusions could be inferred from these findings. It could be that despite aligning with B-box consensus sequences, none of these sequences do not contain TFIIIC binding sites, as TFIIIC binding is considered a requirement of barrier activity (107, 214). A more likely interpretation is that these sequences bind TFIIIC but are missing other qualities which tDNAs possess which allow other essential components for barrier activity to be recruited. To determine why tDNAs exhibit barrier activity but orphan A- and B- boxes do not, a detailed comparison needs to be conducted, which can aid in investigations into the still unknown mechanism of barrier activity.

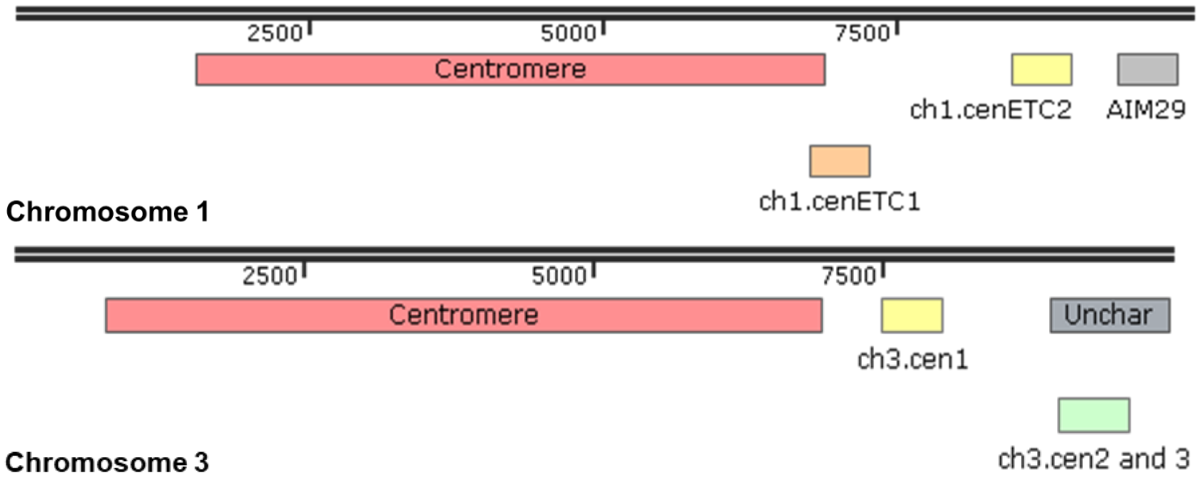


Figure 3.13: Chromosome locations of putative ETC sites.

SnapGene viewer was used to view putative ETC site loci. Centromeres were identified using findings from Coughlan et al (70). Neighbouring genes to the *K. phaffii* centromeres were mapped using the Pichia genome browser (http://pichiagenome-ext.boku.ac.at/). Putative sites were selected within 2500bp from the centromere on their respective chromosomes.

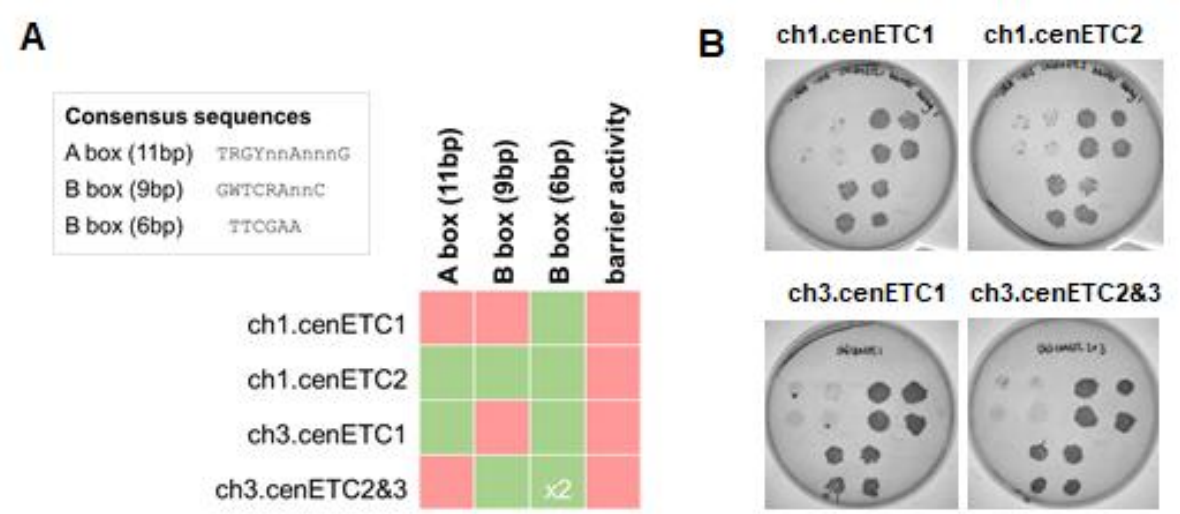


Figure 3.14: Putative ETC sites do not have barrier activity

(A) Consensus sequences present in candidates tested for barrier activity. Candidates consist of 506bp centromeric regions; a 6bp core consensus sequence with 250bp flanking regions. Pink and green indicate an absence and presence of the consensus sequence or barrier activity respectively. (B) Mating assay plates demonstrating that *K. phaffii* orphan A- and B-boxes do not exhibit barrier activity in *S. cerevisiae*. Each plate shows tests in quadruplicate for cells with the *S. cerevisiae* HMR tRNA-Threonine gene (top left), no barrier (top right), or a *K. phaffii* candidate barrier (bottom) inserted between a1 and the HMR-E silencer. Identity of the *K. phaffii* candidate barrier is indicated above each plate.

3.2.7 Nucleosome-resistant flanking sequences have a minimal role in barrier activity.

It has been hypothesised that AT-rich flanking sequences are a requirement for the barrier mechanism, as these prevent nucleosome incorporation and create a gap in the chromatin which prevents the propagation of chromatin remodelling activities (107, 201, 202, 239). This hypothesis was examined when comparing why tDNAs exhibit barrier activity while orphan A- and B-boxes do no by searching for these nucleosome-resistant sequences within each of the candidates.

Nucleosome occupancy across the *K. phaffii* genome has yet to be mapped so in order to investigate the nucleosome structure around the candidates, nucleosome occupancy predictions were carried out computationally. NuPoP, an R package for Nucleosome Positioning Predictions, was used to carry out this analysis. Predictions were made based on the primary sequences around the putative barriers using duration hidden Markov model proposed in Xi et al (2010) (146). This model has been previously used to predict genome nucleosome occupancies in *K. phaffii* by Yang et al (2018) which investigated the manipulation of *AOX1* promoter strength by inserting poly (dA:dT) tracts (228).

Nucleosome-resistant flanking were successfully detected in strong tDNA barriers ch2tRNA8 and the ch1tRNA31-ch1tRNA32 gene pair (Figure 3.15). A peak indicating a nucleosome directly on top of the tDNA was observed, but this nucleosome is likely to be evicted by the binding of TFIIIC and Pol III, so a sufficiently large gap in the chromatin could be present which disrupts the propagation of chromatin remodelling activities. Evidence of nucleosome eviction has been shown in *S. cerevisiae* (240), but experimental nucleosome occupancy data in *K. phaffii* would be required to validate this.

The remaining successful tDNA barriers did not have nucleosome-resistant flanking sequences. ch3tRNA27 was predicted to be flanked by nucleosome-free regions but these regions are further apart; therefore, the upstream region was not captured in the barrier assay and did not contribute to the strong barrier activity observed in Figure 3.10 (Figure 3.15). Further to this, failed barriers ch1cenETC2 and ch3cenETC2&3 were also predicted to have nucleosome-resistant flanking regions (Figure 3.16). This suggests that nucleosome-resistant flanking regions caused by AT-rich sequences is unlikely to have a significant role in barrier activity. However, this data does not indicate nucleosome remodelling activities caused by recruited factors, so it is possible that nucleosome eviction contributes to barrier activity but further investigation would be required.

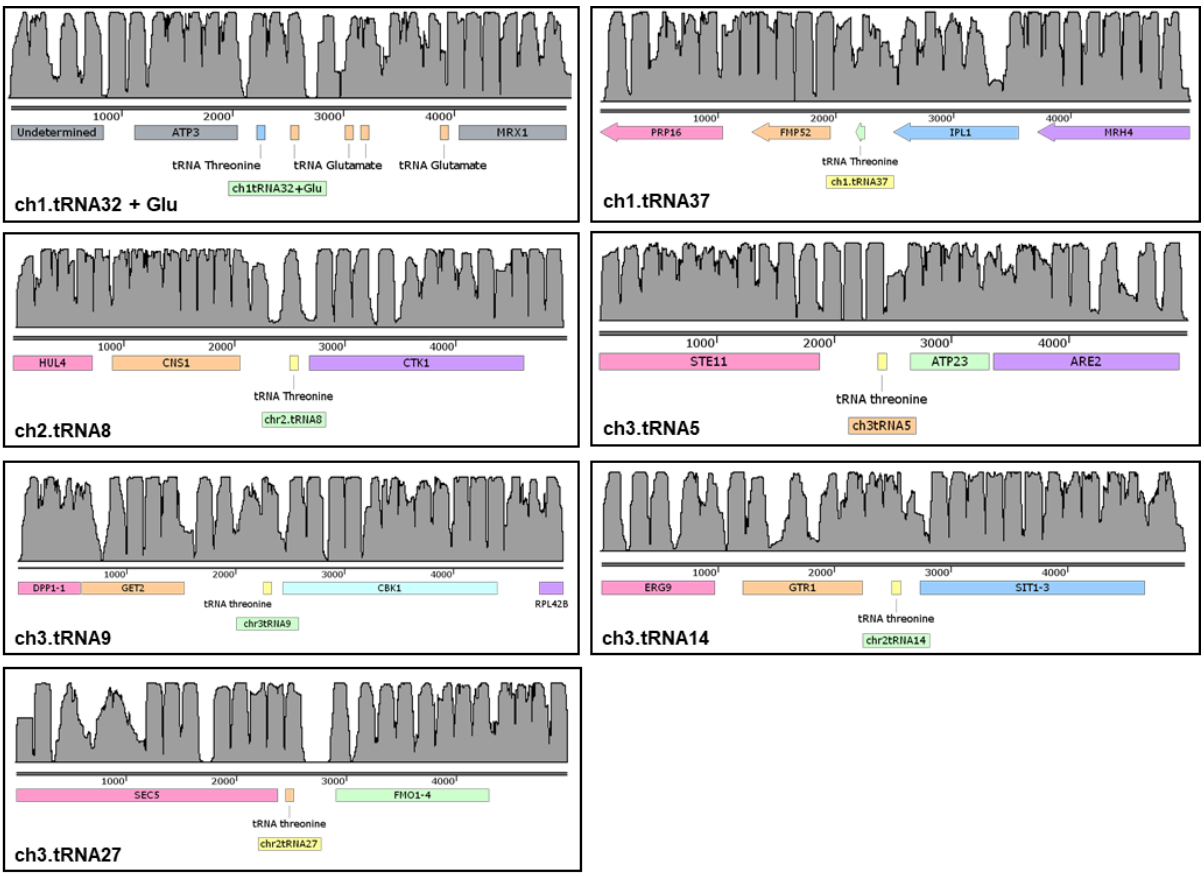


Figure 3.15: Chromosomal positioning and nucleosome occupancy of tRNA threonine candidate barriers.

The potential barrier will be tested for its ability to inhibit silencing of the a1 gene. It will be placed between the HMR-E silencer region (where heterochromatin formation is initiated) and the a1 gene and expression of the a1 gene will be observed. Features below tRNA gene labels indicate the sequence that was amplified and inserted into the mating assay. Predictions were carried out using NuPoP, an R package for Nucleosome Positioning Predictions (Xi et al., 2010). Neighbouring genes were mapped using the Pichia genome browser (http://pichiagenome-ext.boku.ac.at/).

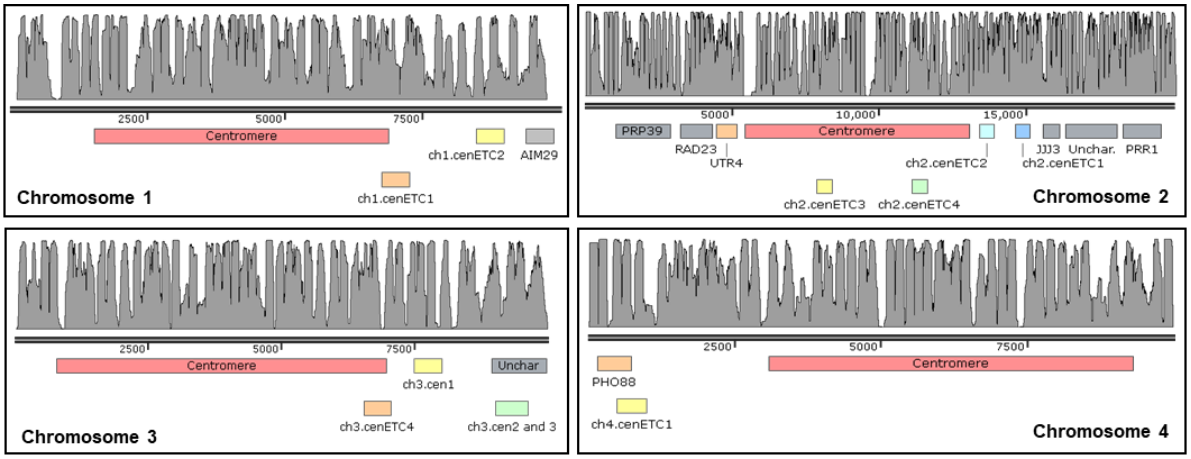


Figure 3.16: Chromosomal positioning and nucleosome occupancy of candidate ETC sites.

The potential barrier will be tested for its ability to inhibit silencing of the a1 gene. It will be placed between the HMR-E silencer region (where heterochromatin formation is initiated) and the a1 gene and expression of the a1 gene will be observed. *K. phaffii* centromeric sequences on chromosome I and (E) *K. phaffii* centromeric sequences on chromosome III. Predictions were carried out using NuPoP, an R package for Nucleosome Positioning Predictions (Xi et al., 2010). Neighbouring genes to the *K. phaffii* centromeres were mapped using the Pichia genome browser (http://pichiagenome-ext.boku.ac.at/).

3.2.8 tDNAs near the *AOX1* locus do not exhibit barrier activity.

Having established that *K. phaffii* has tDNAs capable of acting as heterochromatin barriers, it raised the question as to whether any tDNAs were located near the *AOX1* locus, and whether these were capable of barrier activity. When conducting the genomic analysis for tDNAs near the *AOX1* locus, it was observed that it is located on chromosome 4, within 138kb from the centromere and telomere, known sources of heterochromatin and the *K. phaffii* mating locus (Figure 3.17). This observation has yet to be documented in the literature but could imply that the *AOX1* locus is vulnerable to epigenetic regulation spreading from these sources, resulting in position effect variegation. Homologous recombination occurring at the mating type locus, in particular, could contribute towards this, causing transient changes in expression levels at the *AOX1* locus.

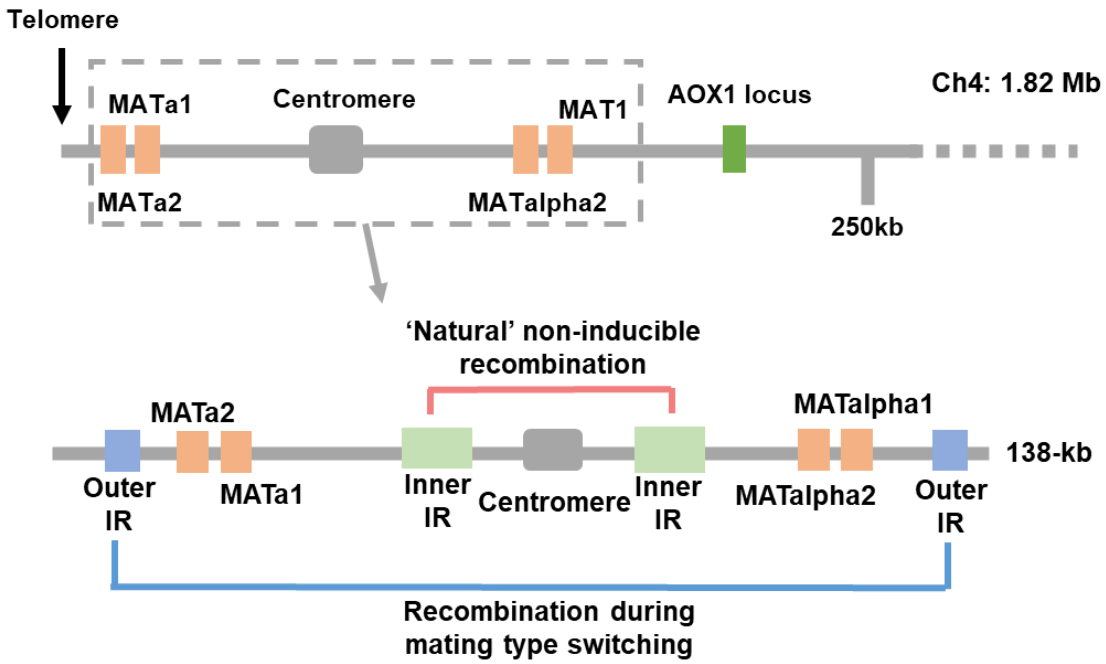


Figure 3.17. Schematic representation of *K. phaffii* mating type locus and areas of homologous recombination.

Centromeric loci and structure as well as inner and outer inverted repeats loci were described in Coughlan et al (70). Genomic loci of the mating type genes were found using the Pichia genome browser (http://pichiagenome-ext.boku.ac.at/).

First, to test these hypotheses, tDNAs found between the centromere and the *AOX1* locus were tested for barrier activity in the *S. cerevisiae* mating assay in order to test whether the *AOX1* is protected by tDNA barriers. Out of the three tDNAs tested, none demonstrated barrier activity, suggesting that they may not be capable of protecting the *AOX1* locus from heterochromatin spread in their endogenous location (Figure 3.18). Therefore, it is possible that the *AOX1* locus could be subject to transient changes in epigenetic regulation which may limit transgene expression.

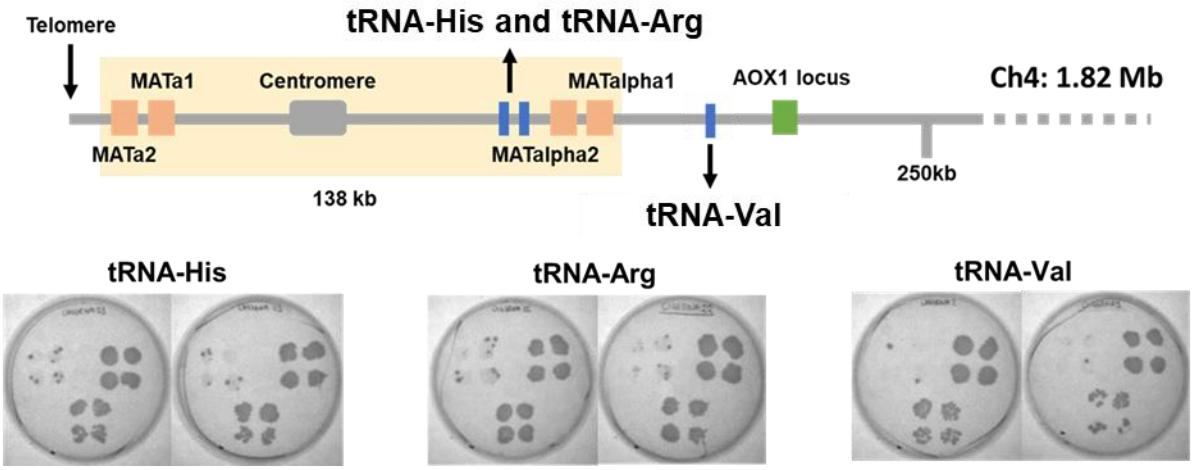


Figure 3.18: tDNAs near the *AOX1* locus do not have barrier activity.

Genomic loci of tDNAs between regions of homologous recombination and *AOX1* locus with mating assay results demonstrating tDNAs with weak barrier activity in *S. cerevisiae*. Eight biological repeats, top left; *S. cerevisiae* tRNA threonine barrier as positive control, top right; no barrier negative control, bottom centre; *K. phaffii* tDNA.

3.2.9 pAOX1 exhibits sharp response rates to methanol.

In order to investigate whether transient changes in transgene expression occurs at the *AOX1* locus, expression of the *Sus scrofa* trypsinogen gene was monitored over time. This enabled the observation of the dynamics and kinetics of expression, which set the stage for further investigations. Transgene transcript levels were measured throughout the growth stage on glycerol, where transgene expression is expected to be repressed, and during the production stage where cultures were subject to regular methanol infusions which induce transgene expression.

A slow incline in transcript levels were seen during growth phase, which indicates ‘leaky’ or uninduced expression of the transgene (Figure 3.19). Sharp response curves were seen in trypsinogen mRNA levels which indicate high activation and deactivation rates, likely correlating with the rapid infusion and depletion of methanol. Sampling occurred every 24 hours during growth phase, and 1 hour, 4 hours and 7 hours post methanol induction during production phase. This captured transcriptional activity at a level of granularity over long-term culture that has yet to be documented, revealing the fluctuating expression in response to methanol induction. This insight highlighted the importance of considering the proximity to methanol inductions when sampling for transgene expression and informed the decision to sample at 7 hours post methanol in subsequent experiments.

A slight reduction in transcriptional activity occurred after 192 hours, or day 7, which may be due to a decline in cell viability (Figure 3.19). No anomalous patterns in transcriptional activity could be observed as the tight regulation of transcriptional activity with methanol masked any significant fluctuations.

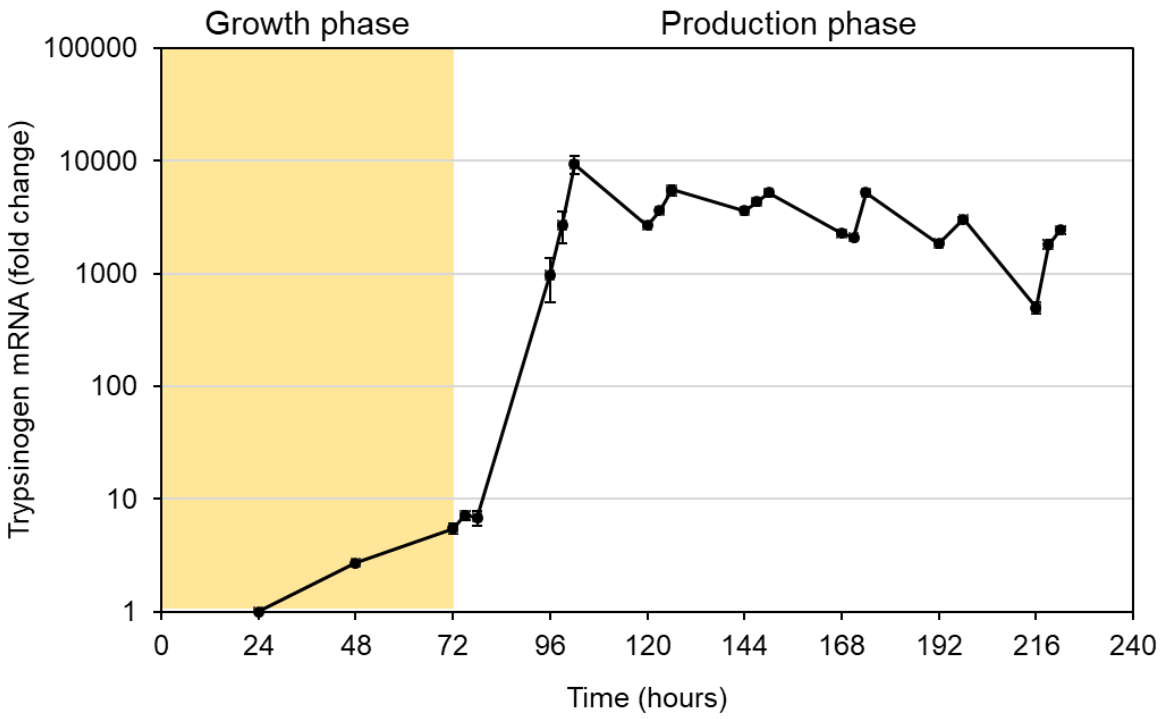


Figure 3.19: Trypsinogen mRNA levels over time, taken with three time points post-methanol induction.

Transcript levels were measured every 24 hours; 1 hour, 4 hours and 7 hours post methanol induction and were normalised to GAPDH mRNA levels. RT-qPCR data was analysed by the delta-delta Ct method. The mean of eight replicates are presented. Error bars indicate SEM.

3.2.10 The rate of transcription activation increases between day 1 and day 2

Due to the inducible nature of the *AOX1* promoter, it was challenging to observe transcriptional fluctuations independent of the inducer. Therefore, it became essential to examine fluctuations while controlling for the time elapsed since the last methanol induction. Transcriptional activity on each day of the production phase, 7 hours post-methanol induction, was compared. It was revealed that despite being subjected to identical methanol inductions, upon the first methanol induction, the transcriptional response is minimal, but upon the second methanol induction, a very large transcriptional response is evoked (Figure 3.20). These distinct types of transcriptional responses resemble a phenomenon observed in similar promoter systems in related yeast species called transcriptional memory, whereby certain inducible genes respond much faster upon re-exposure to the same stimulus. This phenomenon has been observed and well characterised in the *S. cerevisiae* GAL1 and INO1 promoters (241-243) which, like the *AOX1* promoter, is inducible by carbon sources.

Metabolic changes during the production phase, such as glucose starvation between methanol inductions, are known to affect cellular regulatory networks and global gene expression. Proteins involved in metabolism, such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH), can also influence transcriptional regulation under these nutrient-limited conditions, affecting promoter activity. These metabolic shifts may contribute to the enhanced transcriptional response observed on day 2.

Reduced transcriptional activity on days 3 to 5 suggests that prolonged cultivation and repeated inductions may trigger stress responses or metabolic feedback mechanisms that suppress promoter activity. This could involve limitations in cellular resources, accumulation of toxic by-products, or epigenetic changes leading to transcriptional repression.

Ultimately, the data shown in Figure 3.20 demonstrates that day 2 exhibits high transcriptional activity, and during the remaining days of cultivation, transcription is limited. Therefore, it is possible that alleviating this could improve transgene transcription and recombinant protein production.

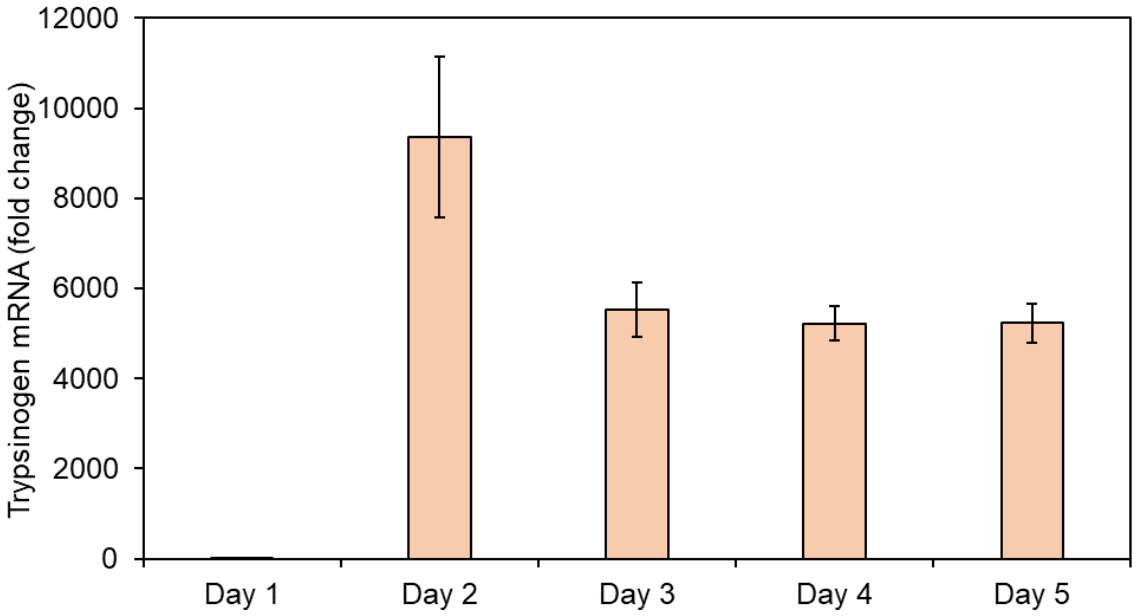


Figure 3.20: Trypsinogen mRNA levels over time, sampled daily.

Transcript levels were measured every 24 hours; 7 hours post methanol induction and were normalised to GAPDH mRNA levels. RT-qPCR data was analysed by the delta-delta Ct method. Means of eight biological replicates are presented. Error bars indicate SEM.

3.2.11 FAIRE-qPCR maps chromatin accessibility across pAOX1

To further investigate whether epigenetic regulation can explain these transcriptional patterns, it became imperative to find a technique which can measure such variables. Formaldehyde-assisted isolation of regulatory elements (FAIRE-seq) has been previously applied in *K. phaffii* to study genes related to its flocculation phenotype (177). In this work, this method was adapted to increase throughput. FAIRE-qPCR was used to enable temporal analysis of chromatin accessibility. Primers were designed tiling the *AOX1* promoter to assess whether any specific transcription factors are particularly regulated by nucleosome remodelling activities. Chromatin accessibility was measured on day 1, day 2 and day 5. If chromatin accessibility is limiting for transcriptional activity, it would be expected that the FAIRE-qPCR data would correlate highly with patterns in transcriptional activity. This is because chromatin state would affect the ability of methanol-dependent transcription factors to bind to pAOX1.

Figure 3.21 shows that chromatin accessibility of pAOX1 does not change across the time points sampled. This suggests that chromatin at pAOX1 is constitutively open and is not a limiting factor in transcription for days 1 and 5. It remains possible that it could be limiting the day 2 response, but it is not possible to infer this from this data. Therefore, it was concluded that while transcriptional activity is limited on four out of five days of cultivation, there could be a limiting epigenetic factor, but chromatin accessibility is unlikely to be it.

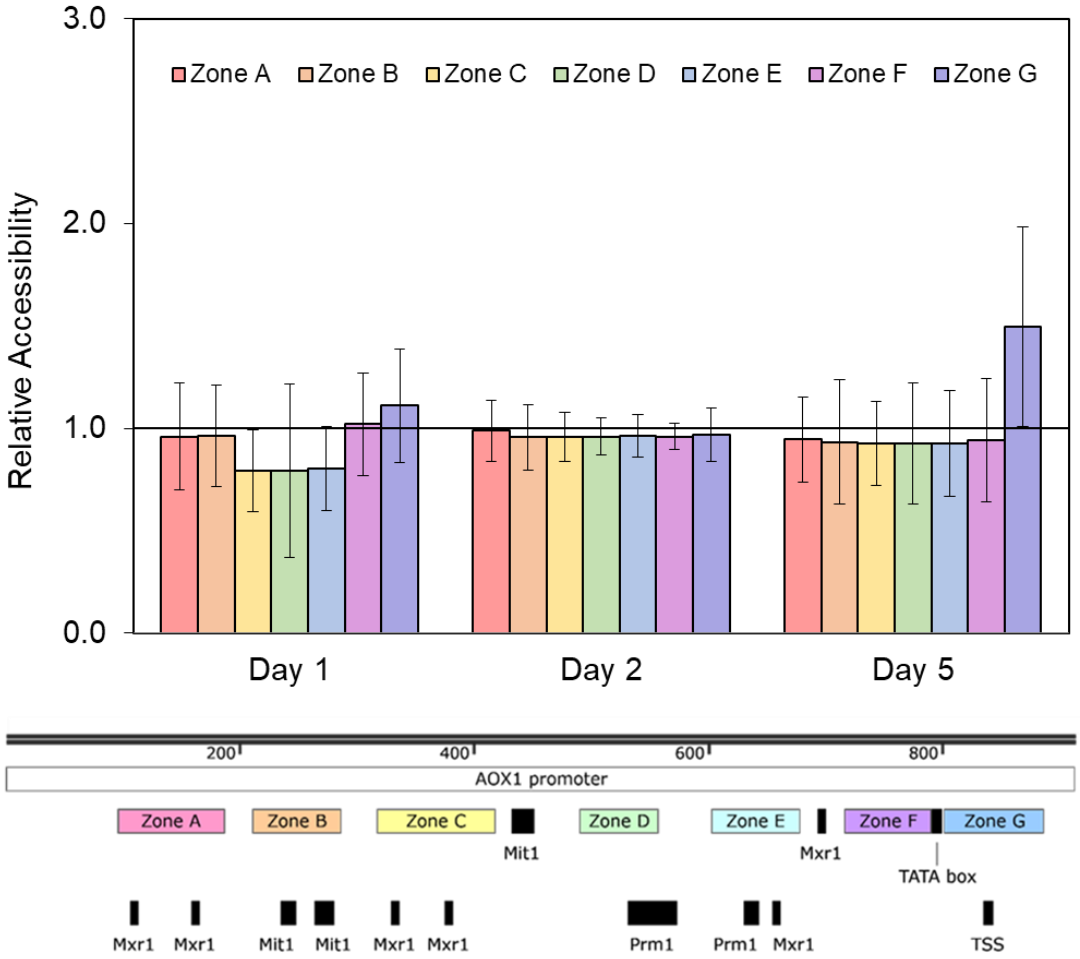


Figure 3.21: Chromatin accessibility across pAOX1 over time.

FAIRE signal was analysed using the delta-delta-Ct method, normalised to input DNA and expressed relative to centromeric DNA on chromosome 4 (CEN4), which is constitutively heterochromatic (70). The binding sites of transcription factors Mxr1, Mit1 and Prm1, which are involved in the methanol-dependent induction of transcription in pAOX1 (227), are displayed below. Results were obtained from three biological replicates. Error bars represent SEM.

3.2.12 Development a barrier construct for *K. phaffii*

Investigations so far have established that *K. phaffii* contains tDNAs which can act as effective heterochromatin barriers in *S. cerevisiae*, the *K. phaffii AOX1* locus is not yet protected by tDNA barriers and that there is scope to improve transcriptional activity of transgenes at the *AOX1* locus. The key aim of this chapter is to explore the possibility of utilising tDNAs to improve transgene expression in *K. phaffii*. Therefore, it remains to test the effect of introducing a tDNA barrier on transgene expression.

The tDNA pair, chr1.tRNA31 and chr1.tRNA32, was chosen as the candidate barrier for testing in *K. phaffii*, as it proved to be an effective barrier in the *S. cerevisiae* screen for barrier activity. An integrative plasmid was designed with 1000bp homology to a site upstream of the *AOX1* locus (Figure 3.22). Restriction sites XbaI and XhoI were included to allow the tDNA barrier to be swapped for other candidates in further investigations. The transgenes can also be substituted using the standard EcoRI and NotI sites common to other *K. phaffii* integrative vectors. The plasmid was linearised at *NdeI*, which lies in the centre of the site for homologous recombination, instead of the standard linearisation sites, such as PmeI, which lie in the centre of the *AOX1* promoter.

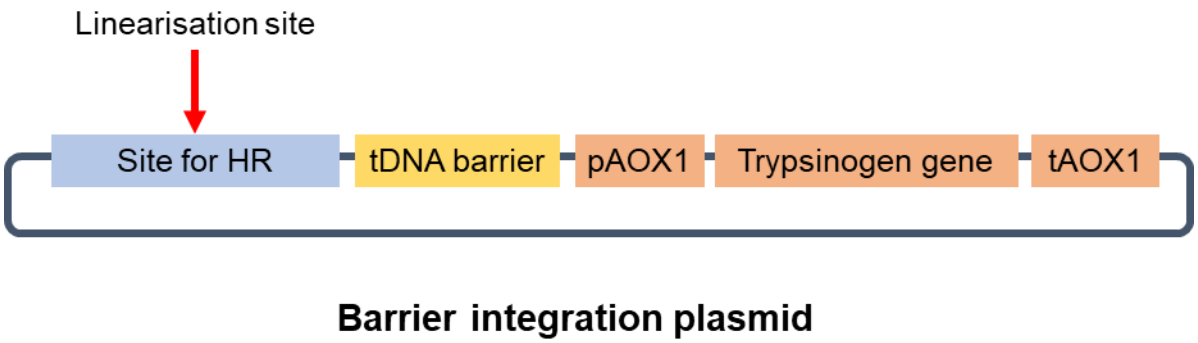


Figure 3.22: Map of barrier integration plasmid.

Schematic representation of the plasmid designed to integrate the tDNA pair upstream of the *AOX1* promoter. 1000 bp homology to the region upstream of the *AOX1* promoter was used for homologous recombination. The linearisation site indicated in the figure represents the *NdeI* site located in the middle of the site for homologous recombination. The tDNA barrier is flanked by restriction sites *XbaI* and *XhoI* to facilitate exchanging of the barrier in future work.

Upon linearisation, the plasmid was transformed into *K. phaffii* cells using electroporation and plated on selective media. Of the colonies subsequently grown under selection, 12 clones were selected for colony PCR screening, with primers targeting a small region of the trypsinogen gene. Of 12 colonies selected, only 3 were positive for the presence of the trypsinogen gene (Figure 3.23). This is quite a low number, suggesting that the integration efficiency of the new site for homologous recombination is low.

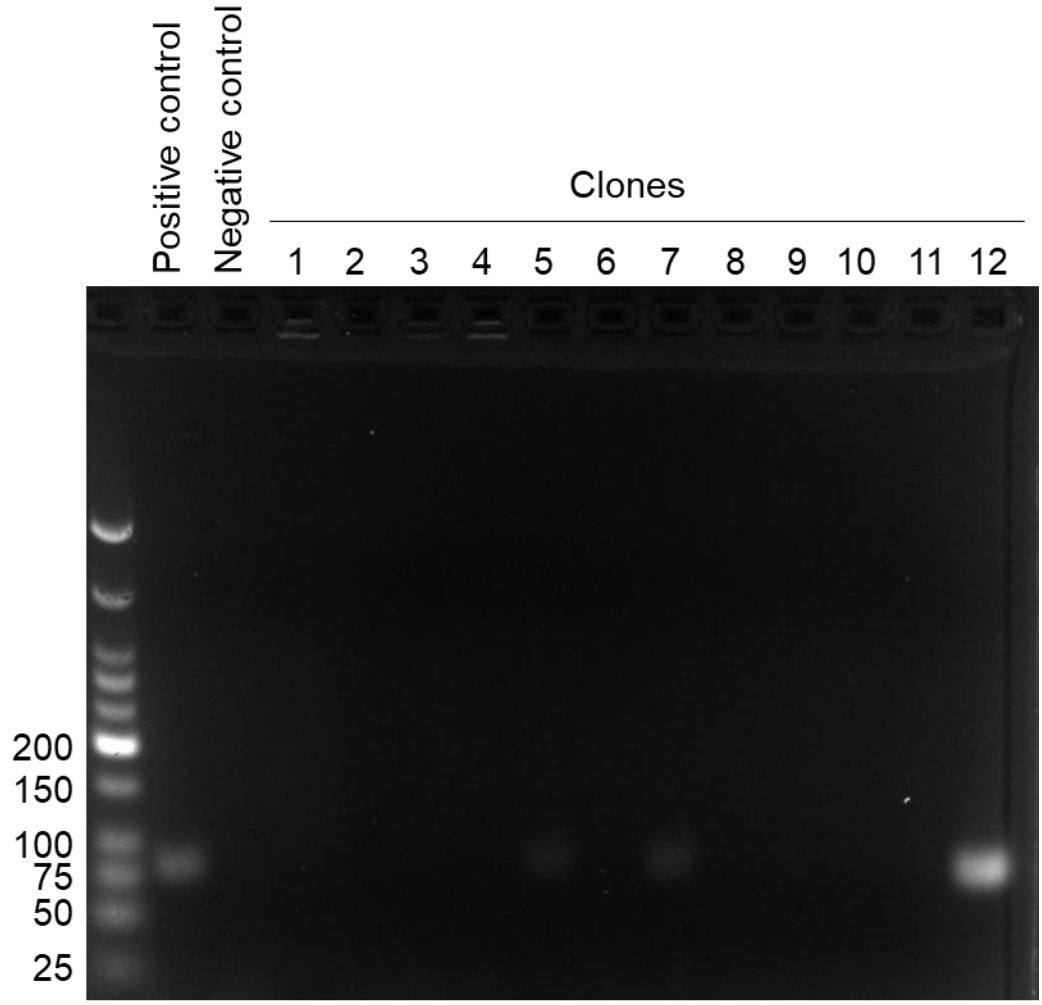


Figure 3.23: Selection of successfully transformed clones.

Colony PCR was used in *K. phaffii* transformants to select positive clones after transformation with the linear barrier plasmid using Transformation protocol 2. Attempts to use Transformation protocol 1 were unsuccessful. Primers designed for qPCR were utilised, amplifying a small region of the trypsinogen gene. Nucleotide bases for the molecular barrier are indicated to the left. Twelve clones were selected for screening but only three of which were identified to be positive clones.

3.2.13 Effect of tRNA on secreted protein

The three successful clones were carried forward for cultivation experiments, where trypsinogen production was assessed. Surprisingly, it was observed that little to no trypsinogen could be detected from these clones (Figure 3.24). An extremely faint band could be observed for Barrier clone 1 but was difficult to distinguish from the background. The expected increase in trypsinogen production upon addition of the tDNA barrier was not observed. This raised the question of whether gene expression was negatively affected by the presence of the tDNA.

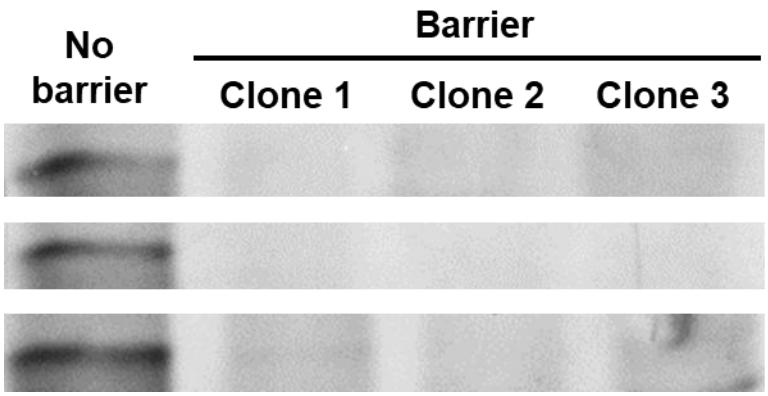


Figure 3.24: Cumulative secreted trypsinogen levels for strains with and without the *K. phaffii* barrier.

Supernatant of cultures cultivated for five days, with daily methanol inductions were analysed for strains expressing trypsinogen with and without the tDNA barrier. Supernatant was analysed using SDS-PAGE and stained with Coomassie Brilliant Blue. All three positive clones were tested but no improvement could be seen in trypsinogen production upon addition of the tDNA barrier. Three biological repeats were conducted.

3.2.14 tDNA barrier disrupts transgene transcription

Trypsinogen transcript levels were measured on days 1, 2 and 5 of the cultivation run in strains with and without the tDNA barrier. In accordance with amount of protein produced, trypsinogen transcription levels were negligible in the barrier containing clones (Figure 3.25). This suggests that the tDNA could have detrimental effects to the chromatin environment around the *AOX1* locus and impede transgene transcription.

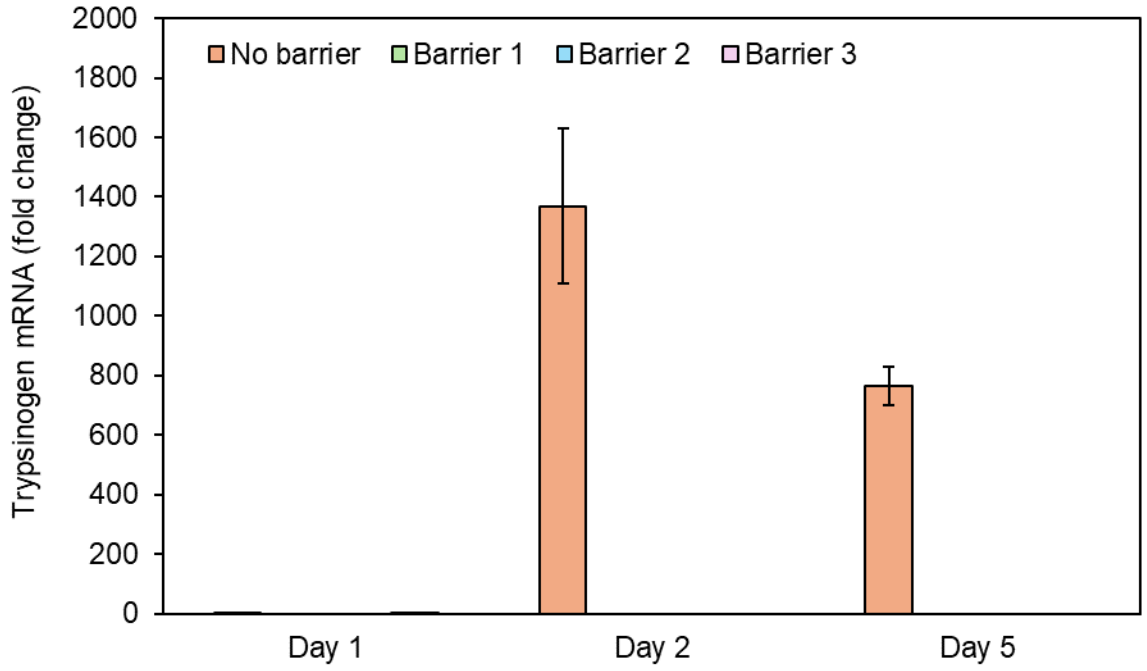


Figure 3.25: Trypsinogen expression levels in the absence and presence of the barrier.

Transcript levels were determined by RT-qPCR using TAF10 as an internal control and analysed using the delta-delta-Ct method. Transcript levels of each gene variant are expressed relative to the no barrier control. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per strain.

3.2.15 tDNA barrier increases pAOX1 chromatin accessibility

In order to investigate whether the tDNA may have an effect on chromatin accessibility, FAIRE-qPCR was conducted for the no barrier control and the barrier 1 strain, sampling on days 1, 2 and 5. As demonstrated previously, for the no barrier strain no detectable change in chromatin accessibility could be observed. However, a significant increase in chromatin accessibility over time could be observed for the barrier strain (Figure 3.26). The fact that transcriptional activity did not increase with the increase in chromatin accessibility supports the conclusion that chromatin accessibility is not limiting for transgene expression and recombinant protein production. However, further investigations are required to explain why an increase in chromatin accessibility causes a decrease in transcriptional activity.

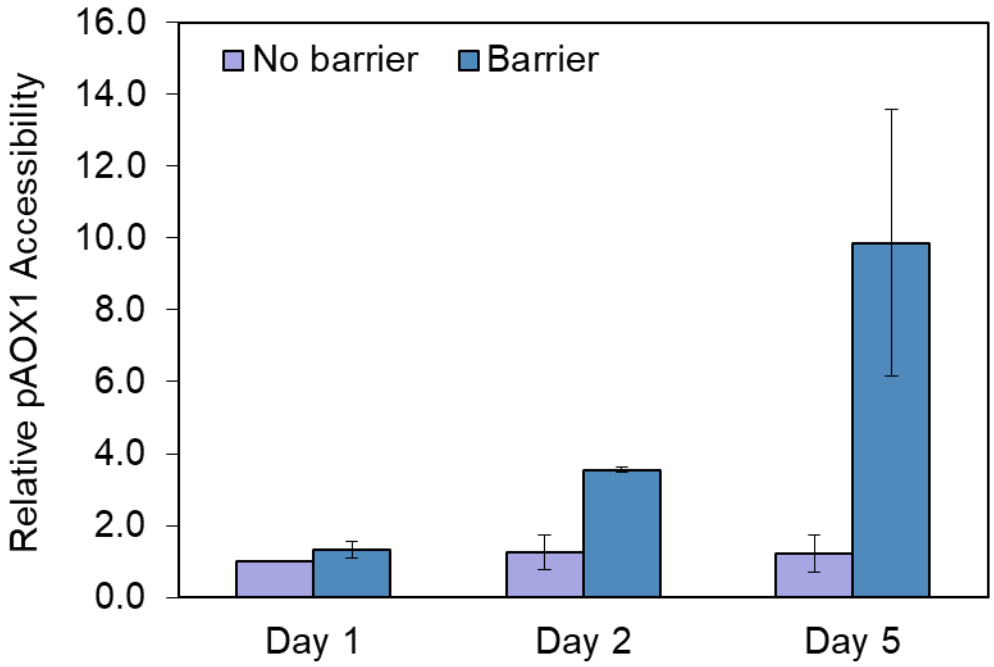


Figure 3.26: Chromatin accessibility with and without a tDNA barrier over time.

FAIRE signal was analysed over time in strains expressing trypsinogen with and without the tDNA. qPCR data was analysed using the delta-delta-Ct method, normalised to input DNA and expressed relative to the no barrier control. Results were obtained from three biological replicates. Error bars represent SEM.

3.3. Discussion and conclusions

This chapter explores the possibility that transgene expression in *K. phaffii*, specifically at the *AOX1* locus, may be subject to epigenetic regulation that could limit recombinant protein production. It investigates the potential for the phenomenon of epigenetic barriers, genetic elements capable of disrupting the propagation of histone modifying and chromatin remodelling activities which cause gene silencing, to improve transgene expression and recombinant protein production. This chapter focuses on the most highly characterised of which; tRNA genes or tDNAs.

3.3.1 Conservation of barrier activity in *K. phaffii*

The data in this chapter presents evidence to support the notion that barrier activity is a conserved mechanism in *K. phaffii*. Not only are many tDNAs located on the boundaries of euchromatin and heterochromatin (Figure 3.6), but they are distinctly accessible to TFIIIC binding compared to non-boundary tDNAs (Figure 3.7). When transferred into a species with confirmed barrier activity, *K. phaffii* tDNAs succeeded at demonstrating barrier activity in a *S. cerevisiae* mating assay (Figure 3.10). To further support this hypothesis, further characterisation of *K. phaffii* tDNAs is required. For example, boundary tDNAs can be assayed for chromatin accessibility and histone marks associated with barriers using the different techniques outlined in the introduction of this chapter. Deletion or mutation experiments manipulating these tDNAs and determining the effect on neighbouring genes would also help to determine unequivocally whether endogenous *K. phaffii* tDNAs can act as heterochromatin barriers.

A particularly interesting finding in this study is that *K. phaffii* tDNAs not only exhibited barrier activity in *S. cerevisiae,* but some even outcompeted the canonical native tRNA-Threonine. Despite the evolutionary distance between these species, the chromatin boundary function of *K. phaffii* tDNAs have been maintained within the primary sequence of the tRNA gene, and is compatible with *S. cerevisiae* barrier machinery. It is unlikely, but not impossible, that *K. phaffii* tDNAs preserved the ability to function as barriers without retaining this property in its own species.

The lack of understanding of the barrier mechanism is a clear hindrance in this investigation, as the particular requirements of barrier activity could not be tested in *K. phaffii*. Upon confirming the ability of *K. phaffii* tDNAs to exhibit barrier activity in their native species, it would also be interesting to test whether *S. cerevisiae* tDNAs could also be capable of barrier activity in *K. phaffii*. This line of investigation could contribute towards our understanding of the barrier mechanism through comparisons of the molecular machinery in *S. cerevisiae* and *K. phaffii*, assuming that the barrier mechanism would be highly similar between the two species. However, barrier activity has been retained in a wide diversity of species, including metazoans, who have diverged significantly from its eukaryotic ancestors. Therefore, it is possible that despite sharing similar characteristics, the mechanism is substantially different in *S. cerevisiae, K. phaffii* and other species which have conserved this phenomenon.

3.3.2 Mechanisms of barrier activity

This chapter attempts to investigate mechanisms of barrier activity by considering particular hypotheses in the literature and comparing properties of successful barriers, *K. phaffii* tDNAs, and unsuccessful barriers, putative *K. phaffii* ETC sites (Figure 3.11). First, the presence of internal promoter element A- and B- box consensus sequences were considered, as these were TFIIIC binding sites and TFIIIC binding is considered to be the minimal requirement for barrier activity. Interestingly, while a *K. phaffii* tDNA, which contains an A and a B box, exhibited barrier activity, a centromeric sequence containing both an A-box and a B-box failed to exhibit barrier activity. This discrepancy suggests a feature of tDNAs which the centromeric sequence lacks which confers their ability to act as a barrier. It would be interesting to determine using a ChIP assay whether TFIIIC is enriched at this centromeric site. It would be expected that TFIIIC occupancy correlates strongly with barrier activity, and thus TFIIIC does not bind the *K. phaffii* centromeric putative ETC site. This could explain why barrier activity was not observed for any of the centromeric sequences tested and could aid in determining the variables affecting TFIIIC recruitment and binding.

Furthermore, this work considers the hypothesis that flanking sequences contribute towards creating an environment that promotes barrier activity (107, 201). In particular, it is thought that AT-rich sequences cause DNA strands to be less flexible and resistant to nucleosome incorporation (202), and create a gap which disrupts the propagation of chromatin condensing activities (239). However, nucleosome occupancy predictions suggest that primary sequences' innate affinity for nucleosome occupancy does not have a great contribution towards barrier activity. Two of seven strong tDNA barriers had nucleosome-free flanking regions (Figure 3.15), but two of four failed centromeric candidates were also predicted to be nucleosome depleted (Figure 3.16). It is likely that active chromatin remodellers such as the RSC complex (210) have a much more significant role in constituting barrier activity than the primary sequence of the DNA and therefore assaying for nucleosome occupancy using MNAse-seq, ATAC-seq or ChIP-seq would provide a better insight into what is happening at these tDNA barriers.

ChIP-seq would be a particularly powerful technique in investigating the mechanism of barrier activity in *K. phaffii*. A wide array of candidates have been implicated in TFIIIC-mediated barrier activity in *S. cerevisiae,* and it would be interesting to validate the involvement of these components in the *K. phaffii* barrier mechanism. It would also be beneficial to use this technique when examining the surprising effects of the tDNA barrier observed in this work. For example, cohesin is a protein complex with the primary role of facilitating the cohesion of sister chromatids to ensure proper chromosome segregation during cell division (244). Mutations in its smc1 and smc3 subunits compromise the barrier functioning of the HMR tRNA-Threonine gene (196). Cohesin is enriched at tDNAs, and functions at the centromeres (244, 245), so may have a role in affecting transgene expression upon addition of the tDNA upstream of the *AOX1* locus. Namely, cohesin has been implicated in establishing chromatin loops (246), and thus recruitment of cohesin at the *AOX1* locus, which is in close proximity to the centromere, could form a chromatin loop that is particularly disruptive to the transcriptional activity of the transgene.

3.3.3 The role of chromatin accessibility in recombinant protein production

By investigating the interplay between chromatin accessibility and transcriptional activity, the hypothesis that epigenetic regulation of the transgene is limiting to recombinant protein production could be interrogated. Two parallel experimental strategies to explore the three-component relationship between chromatin accessibility, transcription rates and protein production. FAIRE-qPCR was used to monitor chromatin accessibility and RT-qPCR was used to measure mRNA levels. No changes in chromatin accessibility of the *AOX1* promoter were observed throughout the cultivation run (Figure 3.21). This presented a discrepancy with the varying levels of transcriptional activity observed of the transgene expressed under this promoter (Figure 3.20). In CHO systems, FAIRE enrichment correlated highly with transcriptional activity of the transgene (247), suggesting that chromatin accessibility is a limiting factor in transgene expression in CHO and is simply not limiting in *K. phaffii* production platforms.

A key difference between CHO and *K. phaffii* production platforms which could explain these findings is the use of a constitutive promoter in CHO and an inducible promoter in *K. phaffii*. The most abundantly used promoter for transgene expression in CHO is the CMV promoter. Constitutive promoters are constantly expressed, and therefore regulation of expression of these genes are likely to be undertaken by epigenetic mechanisms, in order to modulate gene expression according to environmental stressors. However, inducible promoters such as pAOX1 are highly regulated by their environments through transcriptional cascades, such as the addition of methanol to the media causing the binding of transcription factor Mxr1 (221). Therefore, if inducible promoters are also highly regulated by chromatin state, it would limit its responsiveness via other mechanisms, would likely be unfavourable for survival and would be dissuaded from evolutionary conservation.

3.3.4 Epigenetic mechanisms affecting transgene expression at the *AOX1* locus

Two key forms of epigenetic mechanism were investigated in this chapter. The first of which is position effect variegation (PEV). This is the phenomenon whereby the expression of a gene is influenced by its chromosomal location, particularly when it is relocated near heterochromatic regions of the genome (167, 190, 248). In this chapter, it was observed that the *AOX1* locus resides in close proximity to two key heterochromatic regions; the telomere and centromere of chromosome 4. It was therefore hypothesised that transgene expression could be affected by its genomic loci. Testing this would require comparing expression of the transgene in this locus with expression of the transgene in a different locus and would fall out of the remit of this thesis. However, local tDNAs were tested for barrier activity to evaluate whether the *AOX1* locus is protected by endogenous tDNA barriers from these potential threats. Indeed, it was found that the *AOX1* locus is not protected by tDNA barriers (Figure 3.18). It was hypothesised that introducing an effective epigenetic barrier would also indicate whether this phenomenon is occurring, but the tDNA had substantial unexpected impacts on transgene expression. Therefore, the notion that pAOX1 could be affected by PEV remains possible and would be an interesting line of research to pursue.

Despite no changes in chromatin accessibility being observed, interesting patterns in transcriptional activity emerged during the *K. phaffii* cultivation run. Minimal transcriptional response was observed for the first methanol induction, which was followed by extremely large transcriptional response to the second methanol induction (Figure 3.20). It was noticed that this pattern is strongly reminiscent of transcriptional memory. Transcriptional memory refers to when cells that have been primed with a particular stimulus show increased rates of gene expression after re-stimulation at a later time (241, 242, 249, 250). This event was shown to take place in yeast during growth in galactose (249, 251) and inositol (241), in similarly carbon-source inducible promoters. It was found that the galactose signalling pathway of *S. cerevisiae* has the potential for reliably storing information on previous galactose exposures for hundreds of generations (251). This implies that if this mechanism were to exist in *K. phaffii*, it could be exploited to prime the *AOX1* promoter for repeated exposure by adjusting feeding strategies and optimising concentrations of methanol used during cultivation.

Further characterisation of transcriptional responses to methanol as well as assaying for markers of transcriptional memory, including interactions with the nuclear pore complex and a characteristic histone modification: histone H3 lysine 4 dimethylation (H3K4me2) (252). This histone modification permits recruitment of a memory-specific pre-initiation complex, poising RNA polymerase II at the promoter (252).

In addition to revealing interesting patterns of transcriptional activity during *K. phaffii* cultivation runs, these data also highlight the importance of considering sampling for transcriptomics experiments using fed-batch cultures. Dramatic changes in transcript levels occur within the first 7 hours of methanol induction, and it is expected that further changes occur at later time points, and thus it should be imperative that the time post methanol induction is reported in publications presenting data on transgene transcript levels.

3.3.5 Effect of the tDNA on recombinant protein production in *K. phaffii*

The overarching goal of this chapter is to test the utility of tDNAs in improving transgene transcription and recombinant protein production. From the data presented in this chapter, it can be concluded that a tDNA pair which exhibited strong barrier activity in *S. cerevisiae* did not only not improve recombinant protein production, but in fact abolished it (Figure 3.24). This surprising result was supported by the elimination of transcriptional activity of the transgene (Figure 3.25), but interestingly was coupled with increased chromatin accessibility (Figure 3.26).

These findings bring into question whether mechanisms other than barrier activity could be occurring in the barrier strains generated in this work. For example, it has been found in *S. cerevisiae* that tDNAs can strongly inhibit transcription from adjacent pol II promoters *in vivo*, in a phenomenon known as tRNA-gene mediated (tgm) silencing (253). Tgm silencing was found to share components associated with barrier activity, but has been distinguished as a different mechanism entirely (254). The investigation of this behaviour led to the observation that tRNA genes are clustered at the nucleolus in *Saccharomyces cerevisiae*, nucleolar localisation is a determinant for tgm silencing (215). Therefore, to test the hypothesis that the tDNA barrier selected in this study is capable of tgm silencing, markers of this mechanism such as nucleolar localisation could be examined, in addition to testing other tDNAs for their effect on transgene transcription and recombinant protein production.

4. Chapter 4: Exploring the role of wobble-mediated translation in recombinant protein production

4.1. Introduction

Through exploring the utility of tDNA barriers in affecting transgene transcription in *Chapter 3* of this thesis, it became imperative to explore the uses of their products, tRNAs, on processes downstream of transcription in the recombinant protein production pipeline. The most studied process in which tRNAs are involved in is protein translation. Common optimisations of this step include codon optimisation, which aims to modify the codon composition of transgenes based on a number of assumptions about tRNA binding and supply. The following chapter aims to interrogate some of these assumptions and suggest alternative approaches towards improving recombinant protein production through introducing exogenous tRNAs.

4.1.1 Protein translation

Messenger RNA, or mRNA, molecules are the blueprint for protein synthesis as these contain the information which is decoded into a polypeptide chain, the precursor for a mature protein (255). mRNA contains codons, or triplets of nucleotide bases, which denote specific amino acids that are recruited during protein translation (Figure 4.1). This process takes place either in the cytosol or in the rough endoplasmic reticulum, where ribosomes mount and scan the mRNA molecule (106, 256). Protein translation is divided into three phases; initiation, elongation and termination, the first two are highly dependent on tRNAs.

Initiation of translation starts with the formation of a preinitiation complex (PIC) containing an initiator methionine tRNA (iMet-tRNA) which scans for AUG start codons on mRNA molecules. Recognition of a start codon is required for the final initiation complex to be completed and for translation to begin (257). The next step, elongation, involves the binding of incoming aminoacyl-tRNAs to mRNA codons, a process facilitated by elongation factors. The tRNA-bound codon of the mRNA then moves to the P site of the ribosome, where the growing polypeptide chain is transferred to the aminoacyl-tRNA, forming a peptidyl-tRNA. As the ribosome shifts, the tRNA moves to the E site (exit site), and the growing polypeptide chain is passed to the next tRNA at the P site. The uncharged tRNA detaches from the mRNA and is released (258). Finally, the termination of translation is tRNA-independent. It requires recognition of the stop codon by a release factor protein which catalyses the cleavage of the bond between the completed polypeptide chain and the final tRNA (258).

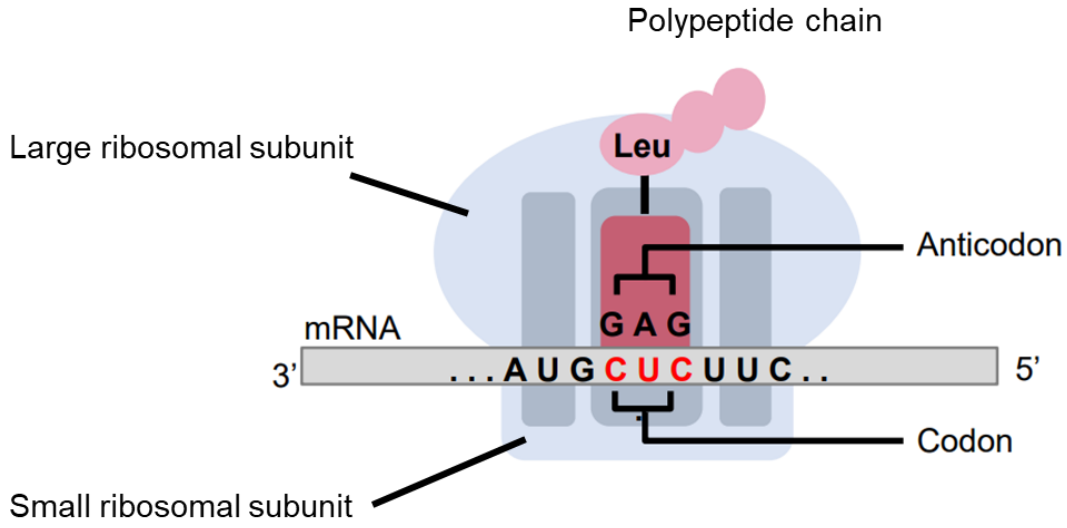


Figure 4.1. Schematic representation of protein translation.

mRNA molecules are decoded by tRNAs carrying anticodons complimentary to codons on the mRNA transcript. Shown here is a peptidyl-tRNA bound to mRNA at the P site of the ribosome. The growing polypeptide remains bound to a peptidyl tRNA until released during translation termination.

4.1.2 Watson-Crick base pairing

The binding of tRNA anticodons to mRNA codons during translation is governed by universal rules which allow for correct decoding. Firstly, each mRNA codon corresponds to a specific amino acid, or a signal such as a start or stop codon (106). The codons are read in a sequential and continuous manner, such that the reading frames do not overlap or skip nucleotides (259). The genetic code is also degenerate, meaning multiple codons can encode the same amino acid (259). For example, six codons; UUA, UUG, CUA, CUG, CUU, and CUC, encode the leucine amino acid. The group of leucine-tRNAs carrying the corresponding anticodons are called isoacceptors as despite carrying different anticodons, they all accept the same amino acid (259).

tRNAs bind to mRNAs via complementary base-pairing, whereby guanine (G) always pairs with cytosine (C) and vice versa, and thymine (T) always pairs with adenine (A) whilst adenine bases (A) always pairs with uracil (U). This set of rules is known as Watson-Crick base-pairing (259). The process of Watson-Crick base pairing is rapid and efficient, with a relatively low error rate and thus is the most common type of base-pairing occurring during translation (259-261).

For the 20 different amino acids used in protein production, there are 61 possible codons. However, for most organisms, fewer than 61 anticodons are available to decode these codons (262). For example, in *S. cerevisiae*, there are only 57 anticodons in its portfolio. This raises the question as to how the remaining four codons are decoded. Through investigating this, it was found that there is some ambiguity in the binding of the third position within the anticodon, allowing the remaining 4 codons in *S. cerevisiae* to be translated by a non-canonical anticodon. This type of decoding is known as wobble base-pairing (260, 261).

4.1.3 Wobble base pairing

Wobble base-pairing refers to the ability of a tRNA anticodon to recognise more than one mRNA codon due to flexibility in the third position of the codon. It is this phenomenon that contributes to the built-in degeneracy of the genetic code, whereby multiple codons can encode for the same amino acid (260). Contributing to the wobble phenomenon is space within the ribosome around the third nucleotide position in the A site, allowing for small conformational adjustments which gives the name ‘wobble’ (260).

Typically, isoacceptors are utilised undergo wobble-mediated translation, resulting in the addition of the same amino acid as the codon demands. However, there is an increased risk of error compared to Watson-Crick base-pairing, causing a point mutation within the protein sequence (263). This raises the question as to why has the wobble phenomenon been retained through evolution. Translational accuracy is often critical for cell viability as inaccurately translated proteins can disrupt essential cellular processes, so it would be expected that a phenomenon which increases error rate would have been selected against.

However, despite being slower and more error prone, the wobble phenomenon remains conserved among all eukaryotes (118). This suggests that it must provide some important advantages which may be lost upon removing it. One such advantage is thought to be that slowed rates of elongation creates the time and space for proper protein folding and translocation into the ER (264). Therefore, decreasing the cell’s reliance on wobble may cause an increase in misfolded proteins, which can cause unfavourable consequences.

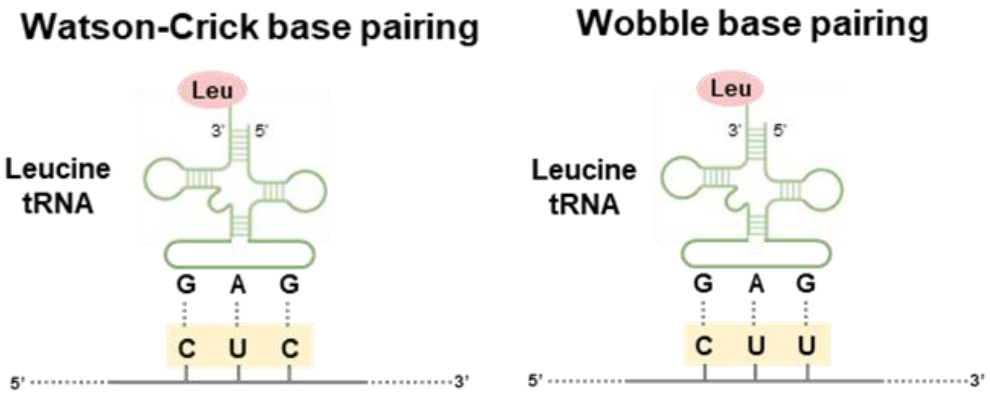


Figure 4.2: Schematic representation of Watson-Crick vis Wobble base pairing.

Codons on mRNA molecules can be decoded by tRNAs through two different mechanisms: Watson-Crick base pairing and Wobble base pairing. Shown here, a tRNA-Leucine carrying the GAG anticodon can decode the CUC codon using complementary Watson-Crick base pairing. However, a tRNA-Leucine carrying the GAG anticodon can also decode the CUU codon.

4.1.4 Mechanisms of non-standard codon:anticodon pairings

Codons and anticodons pairings occur through hydrogen bonding between the two bases. Specificity is conferred by the chemical structure of the bases, which fall into two classes; purines and pyrimidines, which can only pair with each other to form a stable bond (265). If two purines tried to pair, they would be too large causing steric hinderance which would disrupt the structure of the bond, whereas two pyrimidines would be too small and will not be able to achieve proximity to form a stable interaction (265). Watson-Crick base-pairing is therefore the more stable and energetically favourable interaction.

To participate in wobble base pairing, some tRNAs have to undergo specific chemical modifications at the anticodon loop's wobble position, typically at position 34, to be able to form stable interactions with an incompatible base (118). For example, deamination of adenosine results in inosine, which can pair with U, C, or A within mRNA codons, providing a broader decoding capacity (266). Modifications such as these expand the pairing capacity of the tRNA without altering standard Watson-Crick base-pairing rules at other positions. However, there are some naturally permissible non-standard, wobble interactions, such as that between guanine (G) and uracil (U). This G-U pairing forms because the two bases can establish two hydrogen bonds, making the pairing stable enough in the flexible wobble position of the tRNA’s anticodon.

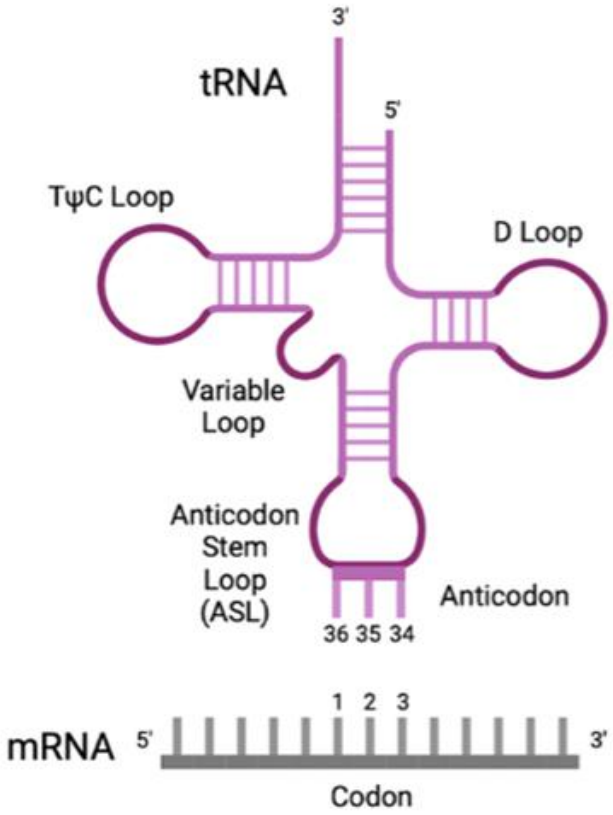


Figure 4.3: Schematic representation of a mature tRNA.

Mature tRNAs are folded into secondary structures, often depicted as a cloverleaf shape with four arms, or stems. The Anticodon stem contains the anticodon stem loop (ASL) which carries the anticodons which bind to codons via Watson-Crick or wobble base pairing. Anticodons are often reported by base 34, 35, 36 and codons are reported according to codon base 1, 2 and 3. Codon base 1 binds anticodon base 36, codon base 2 binds anticodon base 34 and codon base 3 binds anticodon base 34.

4.1.5 Codon usage bias

A bias exists whereby in a certain genome, some codons are used more abundantly than others. This non-random, unequal usage of synonymous codons is called the codon usage bias. It was found that the most highly expressed genes are most impacted by codon usage bias (267), suggesting that it has a role in affecting protein production efficiencies. Indeed, it was found that substituting rare codons with their more abundant synonymous codon greatly improves the translation rate of the gene (267).

Abundances of tRNA species are thought to be a determinant in codon usage bias. For example, the cognate anticodon for a highly used codon may also be relatively high in abundance compared to cognate anticodons for rare codons. Evidence supports this notion, as it was found that codons decoded by abundant tRNAs are more frequent in coding sequences than their synonyms (268-270). In contrast, rare codons accumulate in lowly expressed genes, and slow down translation rates (271, 272). Therefore, there is a relationship between the demand for an anticodon, through the abundance of its codon, and its supply.

Interestingly, the balance between tRNA supply and demand has been shown to adapt according to environmental influences (273). Yeast cells dedicate a significant proportion of cellular effort towards tRNA production as evidenced by the fact that approximately 10% of RNA are tRNAs (273). Transcription of tRNAs consume an estimated 10-15% of the nucleotides used by dividing cells for RNA synthesis (274, 275). Under stress conditions, it has been shown that yeast cells induce the expression of tRNAs whose codons are enriched within stress-induced mRNAs (276), prioritising cellular efforts towards increasing the pool of tRNAs according to increased demands.

Wobble decoding is slower than Watson-Crick decoding (277), but offers increased flexibility in decoding, so is utilised in instances where tRNA supply is limiting. The Percudani rule states that tRNAs wobble with a synonymous codon only if there is no better tRNA for that codon (278). This suggests that wobble decoding is likely to be overrepresented in rare codons, where its cognate tRNA supply is low, or generally where tRNA supply is not aligned with its demand.

4.1.6 Codon optimisation

Codon usage bias evolved to align tRNA supply and demand such that protein translation is not a bottleneck in protein production. However, in the context of recombinant protein production, exogenous genes are transferred into a completely different species, so it is expected that the host environment is not perfectly synchronous with the recombinant protein of interest. Therefore, extensive efforts have been made to optimise the adaptation of the gene to the new species in order to reduce the translational bottleneck.

Several different approaches to improving translation rates have been implemented to improve recombinant protein production. For example, codon optimisation is a popular tool used to tailor the codon composition of a gene to be in accordance with the cellular hosts codon biases (279). It considers codon usage patterns for a particular species and assumes that highly used codons are more optimally translated and endeavours to replace rare codons with these (5, 280). Except for methionine and tryptophan, which are decoded by a single codon/anticodon, this is possible for all amino acids as codon optimisation algorithms can choose from around 2–6 synonymous codons (281).

Mauro et al (2014) published a critical analysis of codon optimisation in human therapeutics, outlining a number of assumptions made during the process of codon optimisation which could be misguiding innovations to improve recombinant protein production (281). One assumption made is that rare codons are rate-limiting for protein synthesis. Here, they highlight that other factors such as tRNA supply may be the true limiting factor to consider, and that it is in fact limiting supplies of anticodon that are rate-limiting for protein synthesis. Further assumptions made are that synonymous codons are interchangeable without affecting protein structure and function, and that replacing rare codons with frequently used ones will increase protein production. Indeed, it has been found that the codon composition of an mRNA molecule can affect its stability through its effect on its secondary structure (281). In *K. phaffii* expression systems, it has been shown that codon usage bias has a role in supporting gene expression and correct protein folding (280).

In summary, informed codon optimisation can be beneficial for recombinant protein production, but often algorithms can be limited in the variables they consider. By attempting to optimise translation of the recombinant protein, bottlenecks can be introduced elsewhere in the protein production pipeline. As an alternative, universal translation systems are a lesser used approach which address these limitations by engineering the host cell's translational machinery rather than modifying the transgene.

4.1.7 Universal translation systems

To negate the need for codon optimisation, innovations like the *E. coli* Rosetta strains have been developed to improve the production of eukaryotic proteins in prokaryotes (117). These strains express tRNAs carrying cognate anticodons for rare codons in *E. coli*, in order to compensate for the lack of endogenous tRNAs capable of decoding them. For example, *E. coli* strain Rosetta(DE3) compensates tRNAs for the AGG, AGA, AUA, CUA, CCC and GGA codons which rarely are used in *E. coli* but are not uncommon in the human genome. This facilitates the production of therapeutic proteins in a prokaryotic host. Proteins produced in Rosetta strains have demonstrated increased yields compared to *E. coli* strains without supplemented tRNAs (117).

To date, it appears as though *E. coli* is the only species for which strains with supplemented tRNAs have been developed. This presents a significant gap in progress for other species commonly used for recombinant protein production. Being a prokaryotic species, it is understandable that the anticodon pool will be the most incompatible with therapeutic proteins, which are often derived from human or animal origins. However, eukaryotic microbes such as *S. cerevisiae* and *K. phaffii* may also stand to benefit from this approach. This thesis investigates the supplementation of a tRNA carrying an anticodon otherwise missing from the tRNA pool in *K. phaffii* expression system. This work takes the first initial steps in determining whether tRNA supplementation is a promising approach to improving translation rates and recombinant protein production efficiencies.

4.1.8 Aims and Objectives

This chapter looks at the role of tRNAs in decoding mRNA molecules and evaluates the effects of Watson-Crick versus wobble decoding in recombinant protein production. Through examining the production of gene variants with increasing reliance on wobble, and then introducing a tRNA carrying an anticodon which theoretically abolishes that reliance on wobble, the potential for increasing yields of *K. phaffii* production hosts by supplementing tRNAs can be examined.

The aims of this chapter are as follows:

- To engineer strains producing proteins with synonymous codon changes that increase their reliance on wobble
- To investigate the effect of synonymous codon changes on transgene transcription
- To investigate the effect of introducing wobble on recombinant protein production
- To evaluate the effect of introducing an exogenous tRNA on recombinant protein production

4.2. Results

4.2.1 Identification of a missing anticodon in *K. phaffii*

To investigate whether wobble-mediated translation of the transgene limits the rate of recombinant protein production, it was necessary to design gene variants with differing degrees of reliance on wobble. To achieve this, a codon was selected for which there was no tRNA carrying the cognate anticodon encoded in the *K. phaffii* genome. It was hypothesised that if gene variants were designed to include codons for which there is no available anticodon, this codon will be translated using wobble.

The genomic tRNA database (http://gtrnadb.ucsc.edu/) was used to view the *K. phaffii* anticodon portfolio and identify the missing anticodons. Four anticodons (GAT, CGC, GAG and TCA) were found to be missing in *K. phaffii* (Table 9). tRNAs with these anticodons deliver Isoleucine, Alanine, Leucine and Selenocysteine residues respectively. Selenocysteine, known as the 21st amino acid, is not commonly used in yeast species, so the TCA anticodon was omitted as a candidate for further study. The tRNA libraries of 8 other yeast species were also examined to identify a tRNA gene carrying the anticodons missing from *K. phaffii* (Table 9). A tRNA-Leu-GAG gene, found to be present in *S. cerevisiae*, was selected for further use as *S. cerevisiae* strains were readily available.

Anticodon	*K. phaffii*	*S. cerevisiae*	*S. pombe*	*C. glabrata*	*K. lactis*	*E. gossypii*	*C. albicans*	*C. lusitanaie*	*D. hansenii*
tRNA-Ile-GAT	0	0	0	0	0	0	1	1	0
tRNA-Ala-GGC	0	0	1	0	0	2	0	1	1
tRNA-Leu-GAG	0	1	0	1	1	1	0	0	0
tRNA-SelCys-TCA	0	0	0	0	0	0	0	0	0

Table 9: Comparison of missing anticodons in *K. phaffii* and their occurrences in other yeast species.

The four anticodons absent in *K. phaffii* and their corresponding frequencies in other yeast species; *Saccharomyces cerevisiae, Schizosaccharomyces pombe, Candida glabrata, Kluyveromyces lactis, Eremothecium gossypii, Candida albicans, Candida lusitaniae* and *Debaryomyces hansenii*. Values were obtained from genomic tRNA database (http://gtrnadb.ucsc.edu/).

Interestingly, out of six possible anticodons for leucine, *S. cerevisiae* only has four while *K. phaffii* has five (Table 10). This suggests that *K. phaffii* naturally has a higher capacity for recognising and translating leucine codons than *S. cerevisiae*, and as such, leucine-rich proteins would be more efficiently translated in *K. phaffii.* However, *S. cerevisiae* has more than double the number of copies of tRNA-Leucine genes in its genome, with 21 tRNA-Leucine genes compared to only 9 in *K. phaffii* (Table 10). Since it has been shown that tRNA copy number correlates highly with tRNA supply, this suggests that whilst *K. phaffii* has an increased diversity of anticodons available, the amount of tRNA available for translation may be less than in *S. cerevisiae*.

A study assessing the essentiality of leucine tRNA genes in *S. cerevisiae* found that whilst its three tRNA-Leu-TAG genes were essential for viability, its single tRNA-Leu-GAG gene was shown not to be essential, and thus, CUC codons in *S. cerevisiae* can also be translated using wobble upon deletion of the tRNA gene (282). This supports the notion that the CUC codon, cognate to the GAG anticodon, is a good candidate for investigating wobble-mediated translation.

Anticodon	*S. cerevisiae*	*K. phaffii*
AAG	0	2
GAG	1	0
CAG	0	2
TAG	3	1
CAA	10	3
TAA	7	1
Total	21	9

Table 10: Comparison of leucine tRNA gene copy numbers in *S. cerevisiae* and *K. phaffii*.
Copy numbers of tRNA genes corresponding to each leucine anticodon in *S. cerevisiae* and *K. phaffii*. Values were obtained from genomic tRNA database (http://gtrnadb.ucsc.edu/).

Furthermore, in order to assess the ability of *K. phaffii* to use wobble-mediated translation, codon usage was analysed for the codons are decoded by these missing anticodons (Figure 4.4). Interestingly, the CUC codon has a relatively high codon usage level, meaning that the codon is naturally prevalent within the *K. phaffii* genome. This suggests that this codon is abundantly translated by wobble in endogenous proteins within *K. phaffii* and is therefore appropriate for this study.

***Pichia pastoris* [gbpln]: 137 CDS's (81301 codons)**

fields: [triplet] [frequency: **per thousand**] ([number])

UUU	24.1 (1963)	UCU	24.4 (1983)	UAU	16.0 (1300)	UGU	7.7 (626)
UUC	20.6 (1675)	UCC	16.5 (1344)	UAC	18.1 (1473)	UGC	4.4 (356)
UUA	15.6 (1265)	UCA	15.2 (1234)	UAA	0.8 (69)	UGA	0.3 (27)
UUG	31.5 (2562)	UCG	7.4 (598)	UAG	0.5 (40)	UGG	10.3 (834)
CUU	15.9 (1289)	CCU	15.8 (1282)	CAU	11.8 (960)	CGU	6.9 (564)
CUC	7.6 (620)	CCC	6.8 (553)	CAC	9.1 (737)	CGC	2.2 (175)
CUA	10.7 (873)	CCA	18.9 (1540)	CAA	25.4 (2069)	CGA	4.2 (340)
CUG	14.9 (1215)	CCG	3.9 (320)	CAG	16.3 (1323)	CGG	1.9 (158)
AUU	31.1 (2532)	ACU	22.4 (1820)	AAU	25.1 (2038)	AGU	12.5 (1020)
AUC	19.4 (1580)	ACC	14.5 (1175)	AAC	26.7 (2168)	AGC	7.6 (621)
AUA	11.1 (906)	ACA	13.8 (1118)	AAA	29.9 (2433)	AGA	20.1 (1634)
AUG	18.7 (1517)	ACG	6.0 (491)	AAG	33.8 (2748)	AGG	6.6 (539)
GUU	26.9 (2188)	GCU	28.9 (2351)	GAU	35.7 (2899)	GGU	25.5 (2075)
GUC	14.9 (1210)	GCC	16.6 (1348)	GAC	25.9 (2103)	GGC	8.1 (655)
GUA	9.9 (804)	GCA	15.1 (1228)	GAA	37.4 (3043)	GGA	19.1 (1550)
GUG	12.3 (998)	GCG	3.9 (314)	GAG	29.0 (2360)	GGG	5.8 (468)

Leu

Figure 4.4: K. phaffii codon usage table.

Leucine codons are indicated by a black box. Codon highlighted in yellow was selected for investigating wobble-mediated translation. Values were obtained from the Kazusa codon usage database (https://www.kazusa.or.jp/).

4.2.2 Development of model genes with increasing reliance on wobble

Upon identification of CUC as wobble-translated codon, variants of transgenes encoding trypsinogen and PHA-L were engineered. T1 and P1 are genes encoding trypsinogen and PHA-L respectively, which have been codon-optimised for production in *K. phaffii*. T2 and P2 variants encode the same proteins, but with half of all leucine codons changed to CUC. T3 and P3 variants have all their leucine codons were changed to CUC. Therefore, for both trypsinogen and PHA-L, genes 1, 2 and 3 exhibit an increasing reliance on wobble-mediated translation (Figure 4.5).

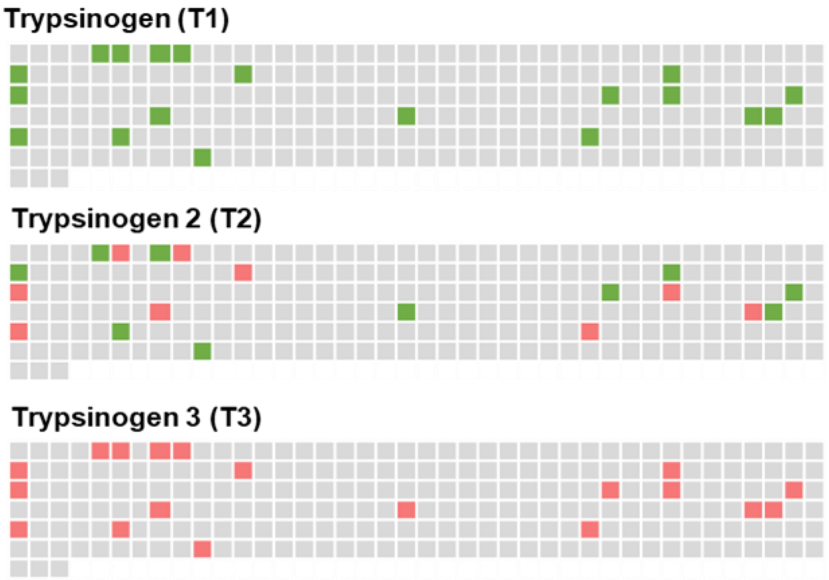


Figure 4.5: **Map of trypsinogen variants**.

Pink squares represent CUC codons and green squares represent other Leu codons.

Comparing the codon-optimised trypsinogen variant (T1) against its native gene sequence revealed that codon optimisation of the gene involved replacing the variety of leucine codons with TTG (or UUG on the mRNA transcript), the leucine codon with the highest codon usage in *K. phaffii* (Figure 4.6). The cognate anticodon for TTG is CAA, for which it has the highest tRNA gene copy number amongst the other leucine anticodons (Table 10). Codon-optimisation is based on the assumption that (a) frequently used codons are translated more efficiently than infrequently used codons, and (b) tRNA gene copy number correlates with tRNA availability, leading to the conclusion that TTG is the optimal leucine codon for recombinant protein production in *K. phaffii*. In the native gene, CTG is the most abundantly used leucine codon, aligning with CUG being the most abundantly used leucine codon in *Sus scrofa* (Figure 4.7). Despite CUC being the second most used leucine codon in *Sus scrofa*, it was completely removed from the codon-optimised gene (Figure 4.6).

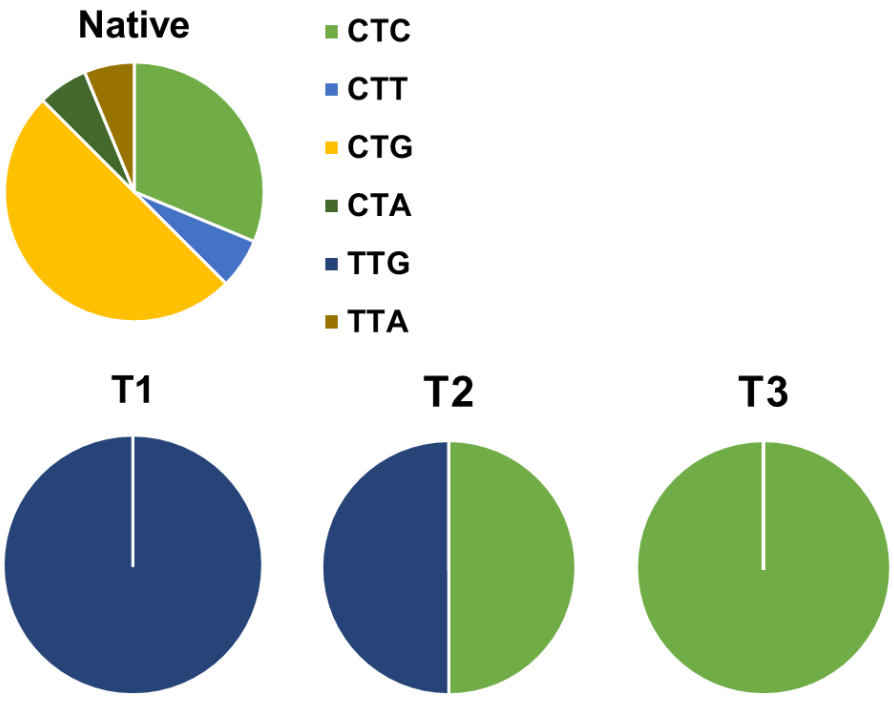
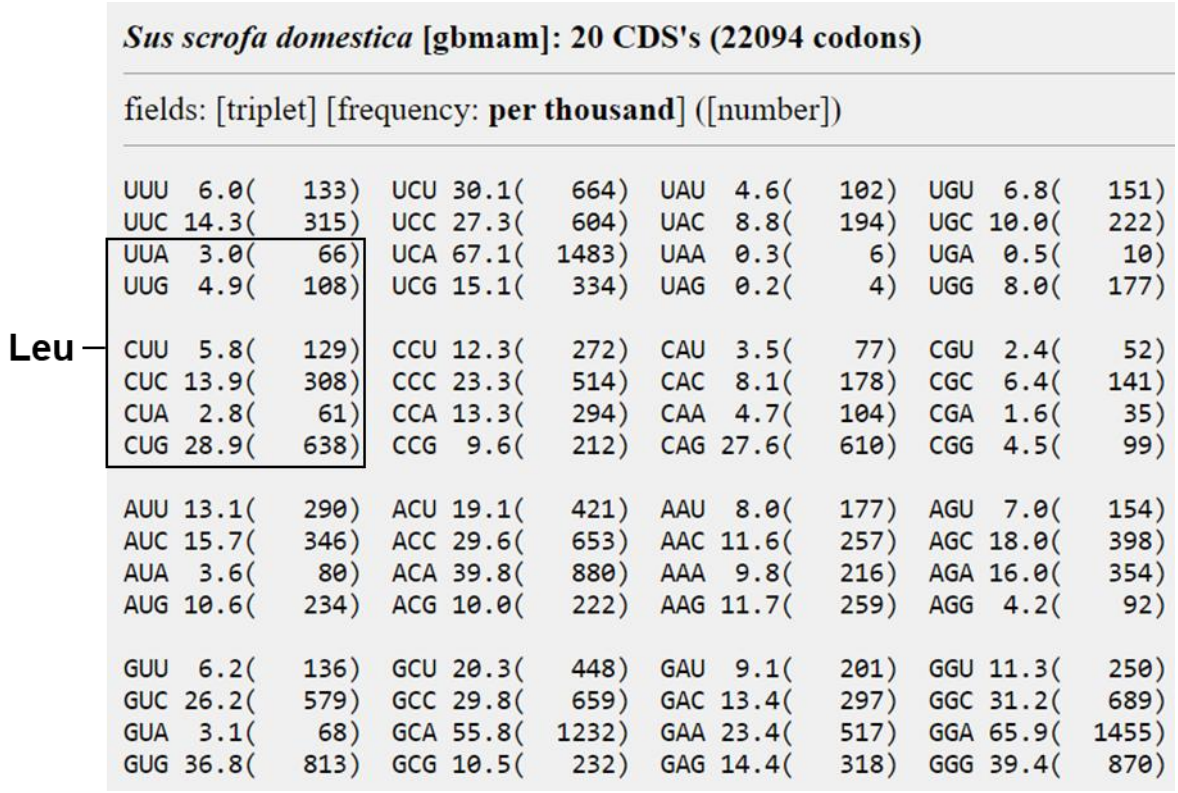


Figure 4.6: Codon usage in trypsinogen gene variants and native gene sequence.
Codon usage was determined using SnapGene Viewer, where the number of instances of each codon was counted and presented as a proportion of total leucine codons.



***Sus scrofa domestica* [gbmam]: 20 CDS's (22094 codons)**

fields: [triplet] [frequency: **per thousand**] ([number])

UUU	6.0 (133)	UCU	30.1 (664)	UAU	4.6 (102)	UGU	6.8 (151)
UUC	14.3 (315)	UCC	27.3 (604)	UAC	8.8 (194)	UGC	10.0 (222)
UUA	3.0 (66)	UCA	67.1 (1483)	UAA	0.3 (6)	UGA	0.5 (10)
UUG	4.9 (108)	UCG	15.1 (334)	UAG	0.2 (4)	UGG	8.0 (177)
CUU	5.8 (129)	CCU	12.3 (272)	CAU	3.5 (77)	CGU	2.4 (52)
CUC	13.9 (308)	CCC	23.3 (514)	CAC	8.1 (178)	CGC	6.4 (141)
CUA	2.8 (61)	CCA	13.3 (294)	CAA	4.7 (104)	CGA	1.6 (35)
CUG	28.9 (638)	CCG	9.6 (212)	CAG	27.6 (610)	CGG	4.5 (99)
AUU	13.1 (290)	ACU	19.1 (421)	AAU	8.0 (177)	AGU	7.0 (154)
AUC	15.7 (346)	ACC	29.6 (653)	AAC	11.6 (257)	AGC	18.0 (398)
AUA	3.6 (80)	ACA	39.8 (880)	AAA	9.8 (216)	AGA	16.0 (354)
AUG	10.6 (234)	ACG	10.0 (222)	AAG	11.7 (259)	AGG	4.2 (92)
GUU	6.2 (136)	GCU	20.3 (448)	GAU	9.1 (201)	GGU	11.3 (250)
GUC	26.2 (579)	GCC	29.8 (659)	GAC	13.4 (297)	GGC	31.2 (689)
GUA	3.1 (68)	GCA	55.8 (1232)	GAA	23.4 (517)	GGA	65.9 (1455)
GUG	36.8 (813)	GCG	10.5 (232)	GAG	14.4 (318)	GGG	39.4 (870)

Figure 4.7. Sus scrofa codon usage table.

Leucine codons are indicated by a black box. Codon highlighted in yellow was selected for investigating wobble-mediated translation. Values were obtained from the Kazusa codon usage database (https://www.kazusa.or.jp/).

Interestingly, the codon-optimised P1 gene retained some variety in its leucine codon usage. Like T1, all the CTC codons were removed and were instead replaced with CTG (Figure 4.8). CTC is the most used codon in the native PHA-L gene despite it not having the highest codon usage in *Phaseolus vulgaris* (Figure 4.9). Evolving a heavy usage of a widely lesser used codon in the native gene could potentially implicate an evolutionary advantage of this codon in limiting translation rates. In P1, the majority of leucine codons were CTG, or CUG (Figure 4.8). CUG is not the most abundantly used codon in *K. phaffii*, so this emulates the potentially limiting CTC codon in *P. vulgaris* (Figure 4.4).

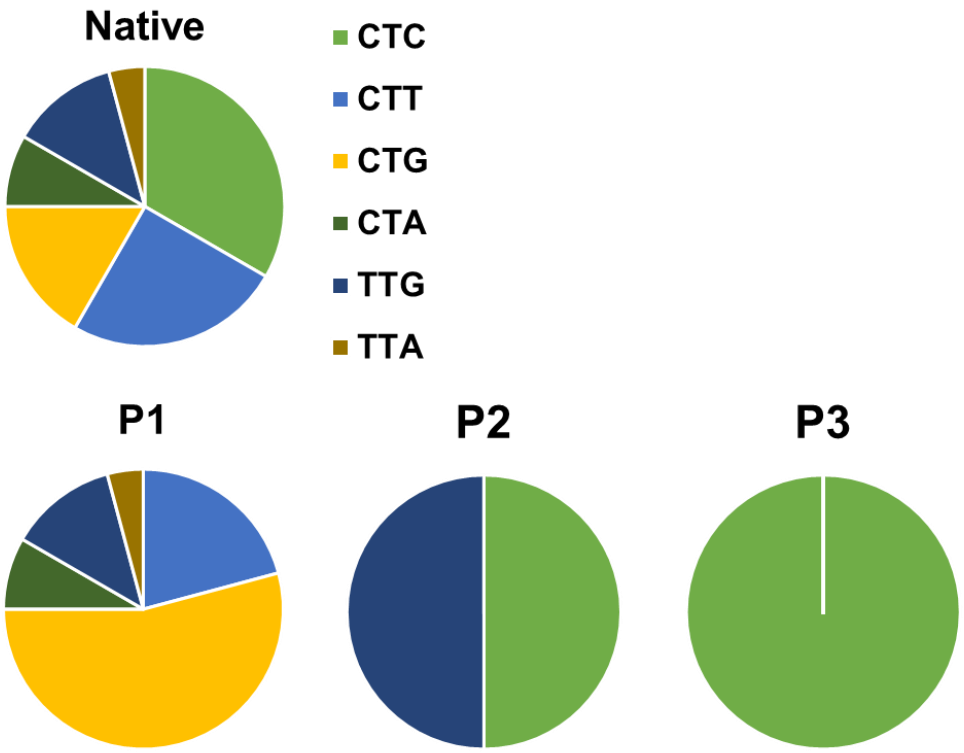


Figure 4.8: **Codon usage in PHA-L gene variants and native gene sequence.**
Codon usage was determined using SnapGene Viewer, where the number of instances of each codon was counted and presented as a proportion of total leucine codons.

Phaseolus vulgaris **[gbpln]: 303 CDS's (132048 codons)**

fields: [triplet] [frequency: **per thousand**] ([number])

UUU 22.6(2979)	UCU 21.5(2843)	UAU 15.6(2064)	UGU 7.9(1047)
UUC 23.3(3074)	UCC 16.7(2205)	UAC 15.4(2031)	UGC 8.3(1092)
UUA 9.7(1279)	UCA 15.8(2088)	UAA 0.7(92)	UGA 0.9(115)
UUG 22.3(2947)	UCG 4.9(649)	UAG 0.7(95)	UGG 13.5(1777)
CUU 23.6(3113)	CCU 17.7(2333)	CAU 12.8(1696)	CGU 7.1(941)
CUC 18.2(2409)	CCC 10.7(1417)	CAC 10.7(1418)	CGC 5.5(720)
CUA 9.4(1244)	CCA 17.2(2273)	CAA 18.7(2472)	CGA 4.2(548)
CUG 13.3(1760)	CCG 4.1(541)	CAG 14.6(1923)	CGG 2.8(372)
AUU 22.2(2928)	ACU 16.1(2130)	AAU 23.4(3088)	AGU 12.1(1603)
AUC 18.7(2470)	ACC 17.0(2244)	AAC 27.1(3579)	AGC 11.8(1556)
AUA 12.0(1583)	ACA 15.0(1981)	AAA 29.9(3946)	AGA 13.2(1744)
AUG 21.3(2807)	ACG 5.8(770)	AAG 32.4(4272)	AGG 12.1(1592)
GUU 23.6(3117)	GCU 24.7(3266)	GAU 32.6(4305)	GGU 22.7(2997)
GUC 12.2(1605)	GCC 17.2(2265)	GAC 21.3(2819)	GGC 13.7(1809)
GUA 7.6(1007)	GCA 19.8(2610)	GAA 31.9(4207)	GGA 23.2(3057)
GUG 21.8(2880)	GCG 5.3(696)	GAG 29.3(3867)	GGG 12.8(1691)

Leu

Figure 4.9: *Phaseolus vulgaris* codon usage table.

Leucine codons are indicated by a black box. Values were obtained from the Kazusa codon usage database (https://www.kazusa.or.jp/).

4.2.3 Codon optimisation is not deleterious for transgene transcription

A key assumption made when codon optimising a gene for production in a different species is that synonymous codon changes will not have a deleterious effect on mRNA levels. However, it is well documented that the codon composition of an mRNA molecule can affect its stability through its effect on its secondary structure, which can in tern impact transcript abundance (281).

To test the effect of the synonymous codon changes made in this study, trypsinogen and PHA-L transcript levels were measured 7 hours post methanol induction on day four of cultivation in *K. phaffii* as a representative marker of transcription levels throughout the cultivation run. Data in Chapter 3 was used to inform this, as it was shown that transcription rates are variable in response to methanol during the first few days of cultivation.

It would be expected that there would be no significant difference in transcript levels between strains expressing T1, T2 and T3, as well as between strains expressing P1, P2 and P3. Indeed, no significant difference was found in transcript levels between T1, T2 and T3 (Figure 4.10). This contrasts with observations in the PHA-L constructs where a significant decrease in transcript level was seen in P2 compared to P1, suggesting that the codon changes did cause a change in transcription rate of the gene (Figure 4.11). This defect in transcription was, however, rescued in P3 due to an insignificant difference in P1 and P3 transcript levels, suggesting compensatory effects of specific codon modifications.

Taken together, these results demonstrate that while synonymous codon changes may not universally alter mRNA abundance, their effects can be context-dependent. This highlights the need for careful evaluation of codon optimisation strategies to avoid unintended negative impacts on transcription and mRNA stability, which may ultimately affect recombinant protein yield.

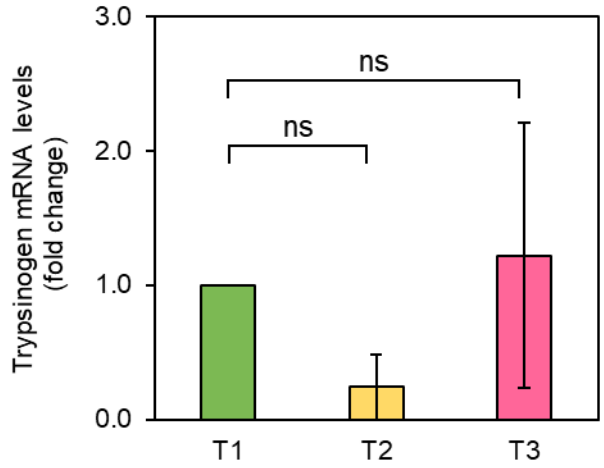


Figure 4.10: **Mean relative transcript levels of trypsinogen gene variants.**
Transcript levels were determined by RT-qPCR using TAF10 as an internal control and analysed using the delta-delta-Ct method. Transcript levels of each gene variant are expressed relative to T1. Error bars depict the standard error of the mean (SEM), with n = 4 biological replicates per variant. Student's T-tests were carried out to test for statistical differences between conditions. 'ns' denotes 'not significant'.

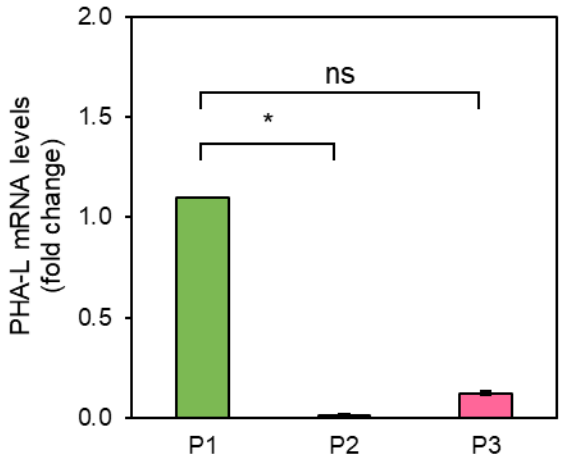


Figure 4.11: **Mean relative transcript levels of PHA-L gene variants.**
Transcript levels were determined by RT-qPCR using TAF10 as an internal control and analysed using the delta-delta-Ct method. Transcript levels of each gene variant are expressed relative to P1. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per variant. Student's T-tests were carried out to test for statistical differences between conditions. [ns] denotes 'not significant'. [*] denotes $p < 0.05$. [***] denotes $p < 0.001$.

4.2.4 Wobble-mediated translation does not limit recombinant protein yields in *K. phaffii*

To investigate whether the synonymous codon changes introducing wobble-mediated translation affects recombinant protein yields, cumulated secreted titres of trypsinogen and PHA-L were examined on the final day of cultivation. Comparison of the three trypsinogen gene variants revealed no significant difference in yield (Figure 4.12). This suggests that increasing the reliance on wobble decoding does not affect the rate of trypsinogen production, and thus wobble decoding is not rate-limiting.

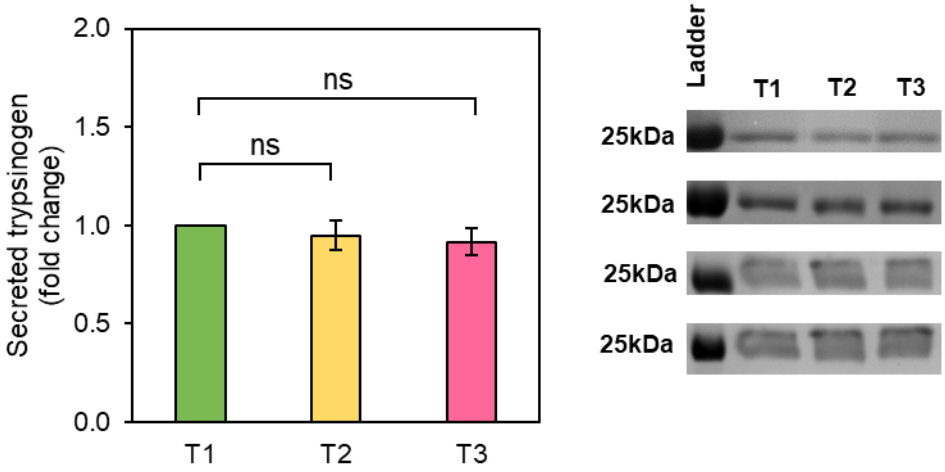


Figure 4.12: **Secreted trypsinogen levels for strains expressing T1, T2 and T3.**
Supernatant of cultivated cultures were collected on day 5 of cultivation and analysed via SDS-PAGE and staining with Coomassie Brilliant Blue. Secreted protein level was determined by quantitative densitometry, normalised to OD_{600} and are expressed relative to T1. Error bars depict the standard error of the mean (SEM), with n = 4 biological replicates per variant. Student's T-tests were carried out to test for statistical differences between conditions. [ns] denotes 'not significant'.

Figure 4.13 displays the amount of secreted protein per strain, normalised to mRNA transcript levels. Compared to T1, there was a significant reduction in the amount of protein per mRNA for T2, suggesting that the synonymous codon changes could have a role in decreasing the rate of translation. However, T3 showed a significant increase in the amount of protein per mRNA, suggesting an increased rate of protein production. As no changes in secreted protein levels were seen despite the changes in protein production rate suggests that there could be a downstream bottleneck preventing changes in secreted protein. The opposing effects of synonymous codon changes for trypsinogen suggests that rates of translation are significantly affected by codon usage. T2 containing two types of leucine codon was produced less efficiently than T3 which contained only one type of leucine codon.

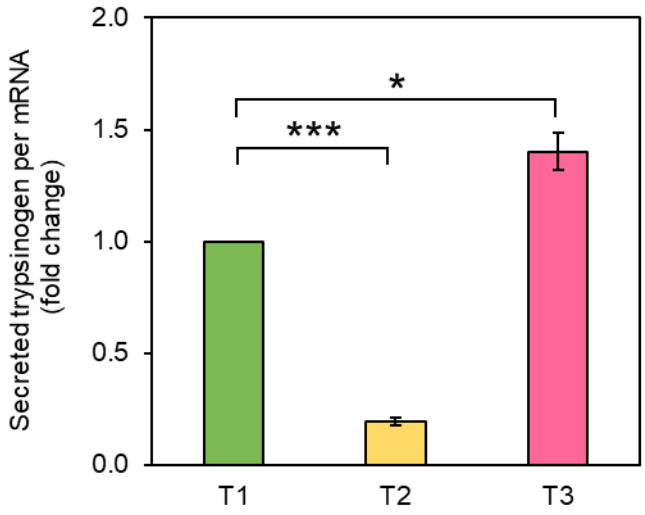


Figure 4.13: **Secreted trypsinogen levels per mRNA molecule.**
Secreted protein levels were normalised using trypsinogen transcript levels and are expressed relative to T1. Error bars depict the standard error of the mean (SEM), with n = 4 biological replicates per variant. Student's T-tests were carried out to test for statistical differences between conditions. [ns] denotes 'not significant'.

Like for the trypsinogen, comparison of the three PHA-L gene variants also revealed no significant difference in yield, despite there being a significant difference in transcription rate of P2 (Figure 4.14). Again, this suggests that there is a significant bottleneck in the secretory process.

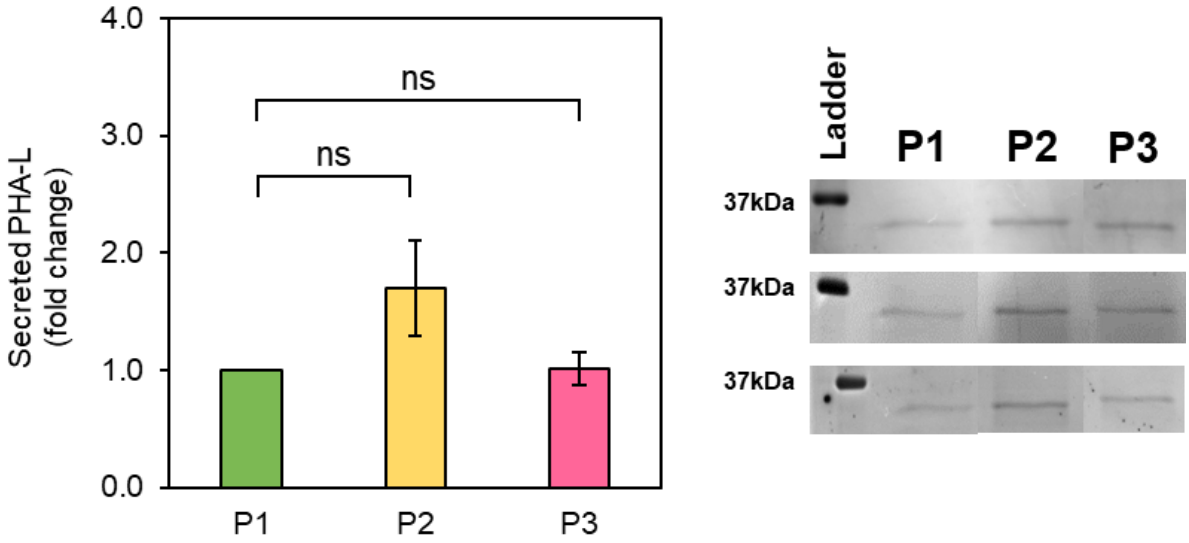


Figure 4.14: Secreted PHA-L levels for strains expressing P1, P2 and P3.

Supernatant of cultivated cultures were collected on day 5 of cultivation and analysed via SDS-PAGE and staining with Coomassie Brilliant Blue. Secreted protein level was determined by quantitative densitometry, normalised to OD600 and are expressed relative to P1. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per variant. Student's T-tests were carried out to test for statistical differences between conditions. [ns] denotes 'not significant'.

Assessing the amount of secreted protein per mRNA shows that compared to P1, P2 showed a significant increase protein production rate (Figure 4.15). P3, however, showed no change in production rate. The differing outcomes for trypsinogen and PHA-L highlight the protein-specific effects of altering translational efficiency on recombinant protein production. Various factors may explain why synonymous codon changes lead to such diverse impacts. Introducing the missing anticodon to eliminate wobble-mediated translation provides a targeted approach to assess how translation rates influence recombinant protein yield, helping clarify the role of wobble-mediated translation in production efficiency.

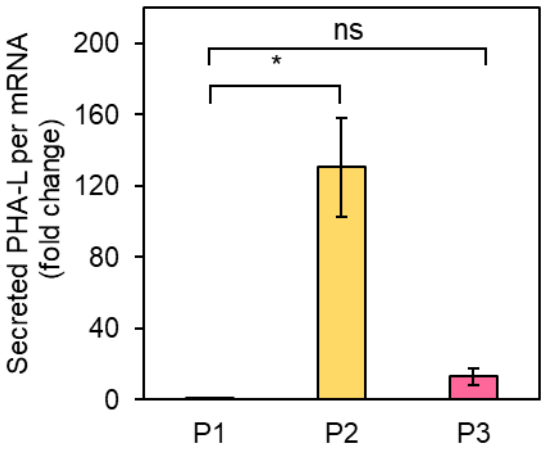


Figure 4.15: **Secreted PHA-L levels per mRNA molecule.**

Secreted protein levels were normalised using PHA-L transcript levels and are expressed relative to T1. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per variant. Student's T-tests were carried out to test for statistical differences between conditions. [ns] denotes 'not significant'. [*] denotes $p < 0.05$.

4.2.5 Development of strains co-expressing transgene variants and exogenous tRNA

The *S. cerevisiae* tRNA-Leu-GAG gene, along with 150bp flanking sequences, was cloned into the transgene expression vectors (Figure 4.16). Upon integration, the tDNA lies 2646 bps away from the *AOX1* promoter, mitigating any potential epigenetic effects of the tDNA discussed in Chapter 3 of this thesis.

Figure 4.16: Schematic representation of the integrative plasmid.
The plasmid contains the *S. cerevisiae* tRNA-Leu-GAG gene, flanked by 150bp upstream and downstream sequences, inserted into the pPICZ expression vector. Within the 2646bp separating the tRNA gene and *AOX1* promoter include the transgene, terminator and the Zeocin resistance expression cassette.

Primers were designed to specifically detect the pre-tRNA transcripts of the *S. cerevisiae* tRNA-Leu-GAG gene, in order to determine whether there was a significant difference in tRNA gene expression between the strains (Figure 4.17).

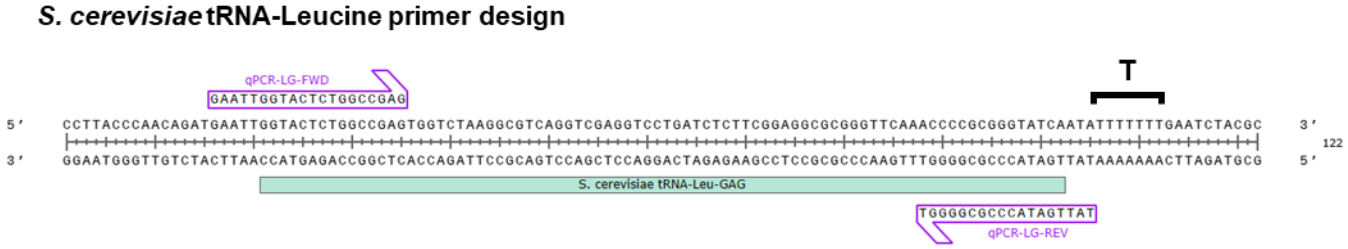


Figure 4.17: **Primers designed for specific detection of the S. cerevisiae tRNA-Leu-GAG gene.**
The forward and reverse primers target the *S. cerevisiae* tRNA-Leu-GAG gene, ensuring specific amplification of pre-tRNA by overlapping slightly with flanking regions. This primer design allows for precise detection and quantification of the tRNA gene in genomic assays, avoiding amplification of endogenous tRNA genes in *K. phaffii*. The Genomic tRNA database (https://gtrnadb.ucsc.edu/) was used to identify the transcriptional start site and analysis of the sequence identified the transcriptional termination site indicated by 'T' in the figure.

Indeed, comparing the strains containing each of the gene variants and the exogenous tRNA gene, there was no significant difference in tRNA expression (Figure 4.18, Figure 4.19), with wide error bars reflecting significant biological noise when comparing statistically highly similar samples. This suggests that each of the cassettes were integrated as a single copy, as tDNA copy number correlates highly with expression levels. This allows us to eliminate differences in tRNA supply as an extraneous variable in further experiments.

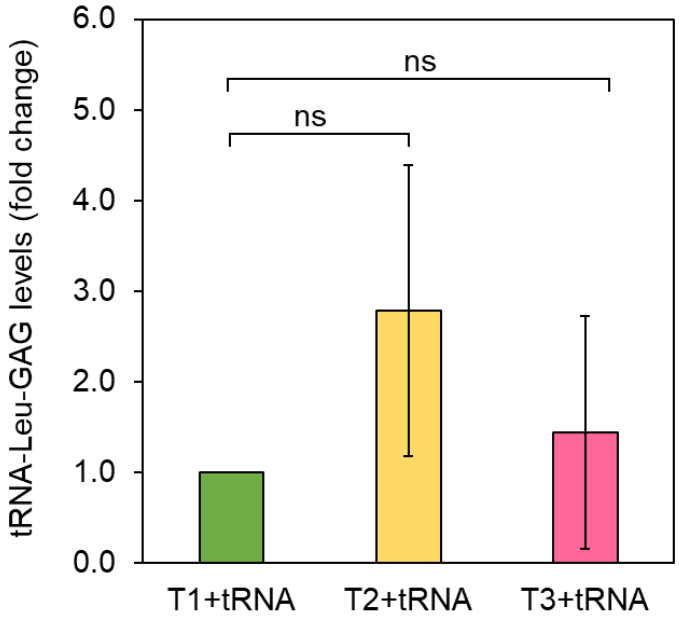


Figure 4.18: Comparison of relative tRNA-Leu-GAG expression levels in trypsinogen-producing strains.

Transcript levels were determined by RT-qPCR using TAF10 as an internal control and analysed using the delta-delta-Ct method. Transcript levels of each gene variant are expressed relative to P1. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per variant. Student's T-tests were carried out to test for statistical differences between conditions. [ns] denotes 'not significant'.

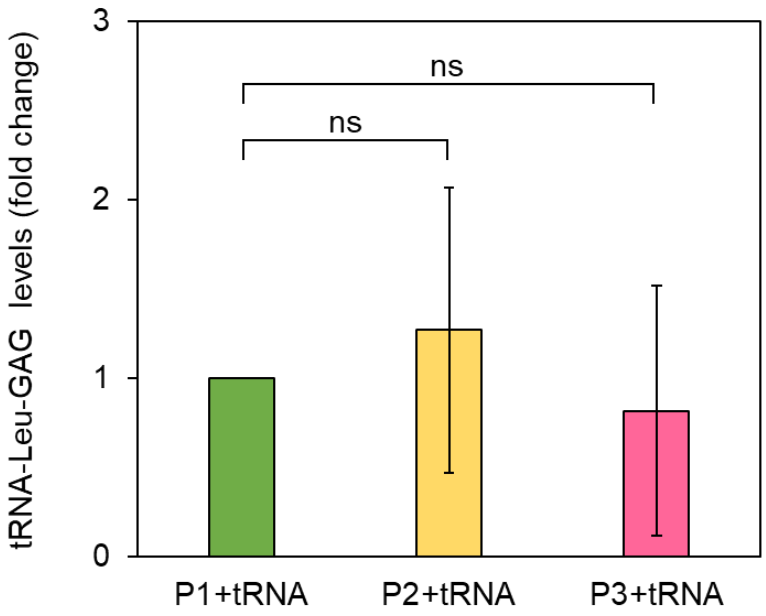


Figure 4.19: Comparison of relative tRNA-Leu-GAG expression levels in PHA-L-producing strains.

Transcript levels were determined by RT-qPCR using TAF10 as an internal control and analysed using the delta-delta-Ct method. Transcript levels of each gene variant are expressed relative to P1. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per variant. Student's T-tests were carried out to test for statistical differences between conditions. [ns] denotes 'not significant'.

4.2.6 Cell size is not affected by exogenous tRNA expression

To determine whether introduction of the exogenous tRNA affects recombinant protein production, both titre needs to be normalised to the optical density of cultures at 600nm (OD_{600}) as a proxy for cell number. Introduction of a tRNA can affect global translation rates, which may affect cell size and change the relationship between optical density and cell count of a culture, skewing the normalisation calculation. To investigate whether this could occur, the average cell size for each strain was measured. Cell size was measured in exponential-phase cells grown in glucose. Growth in glucose means that transgene expression is repressed but tDNA transcription will still occur. Therefore, the effect of only the exogenous tDNA on cell size can be determined.

No effect of the tRNA on cell size was observed, allowing secreted protein to be normalised to OD_{600} (Figure 4.20). However, T3 cell size was shown to be significantly increased, which could be explained by leaky expression of T3. As cells expressing T3 have the greatest reliance on wobble, it could be that a greater demand for CUC decoding could have an impact on the translation of endogenous proteins, resulting in this unexpected phenotype. This will be explored further in Chapter 5.

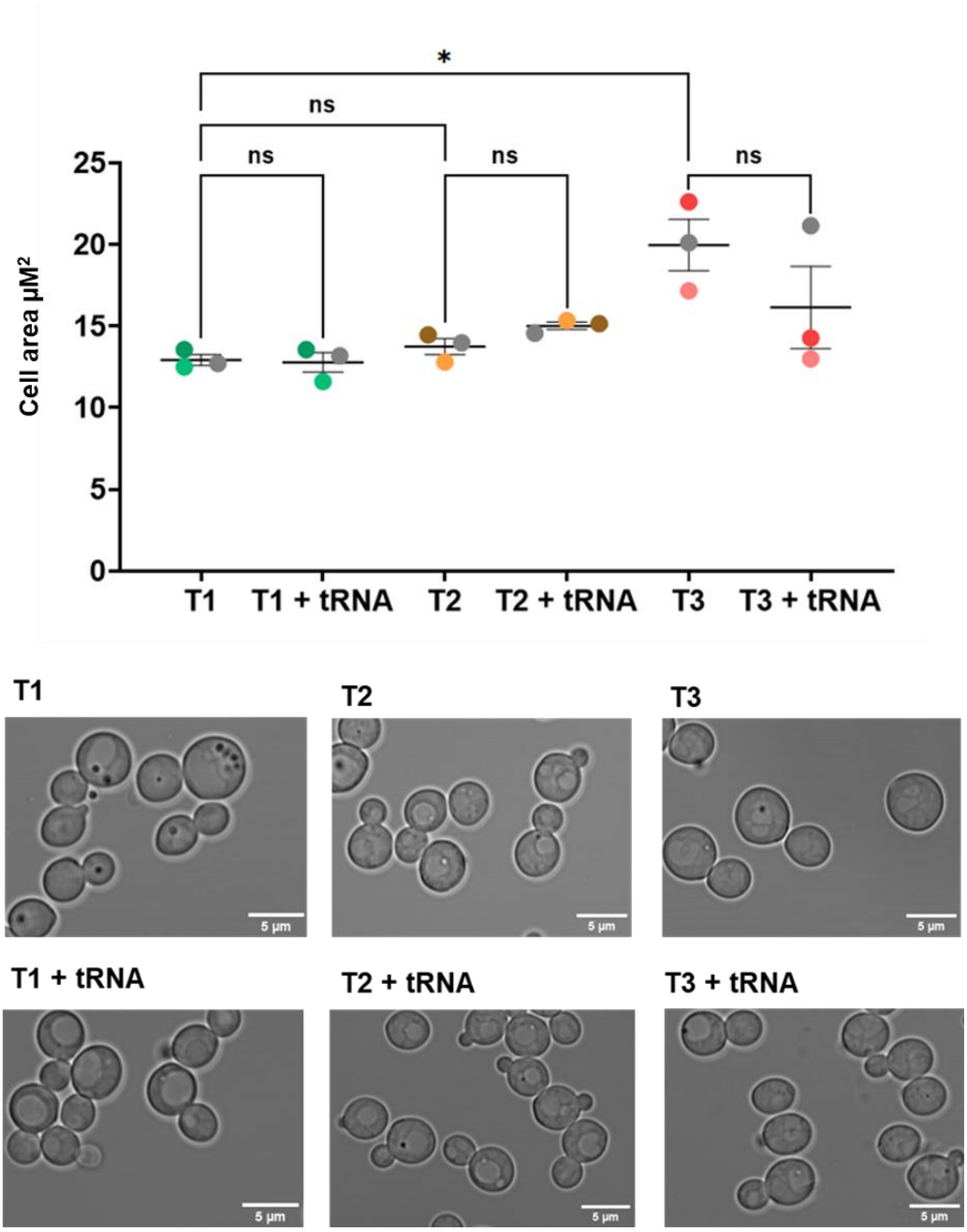


Figure 4.20: **Cell size comparisons for strains expressing each trypsinogen gene variant, with and without the exogenous tRNA gene.**

Cell size was measured using confocal microscopy, followed by analysis with ImageJ to determine cell area in µm². Representative images of the cells are displayed below. Cell size was quantified using the magic cell tool outlining individual cells in each strain. The strains were compared under two conditions: in the absence and presence of the exogenous S. cerevisiae tRNA-Leu-GAG gene. Data are presented as mean cell area with standard error of the mean (SEM) for each condition. A two-way ANOVA was used to determine statistical differences between conditions. [ns] denotes 'not significant' and [*] denotes 'p > 0.05'.

4.2.7 Exogenous tRNA has no effect on transgene transcription

Furthermore, to compare protein production levels between strains, mRNA transcript levels were measured. Normalisation to transcript level allows for any differences in protein production to be attributed to differences at post-transcriptional stages only, including translation and protein secretion. Figure 4.21 and Figure 4.22 show no significant differences between transcript levels in strains with and without the tRNA, demonstrating no effect of the tRNA on transgene transcription. T3+tRNA displayed especially wide error bars, suggesting a degree of clonal heterogeneity despite clones being genetically identical. It is possible that this could be due to epigenetic factors, but further investigation is required to confirm this.

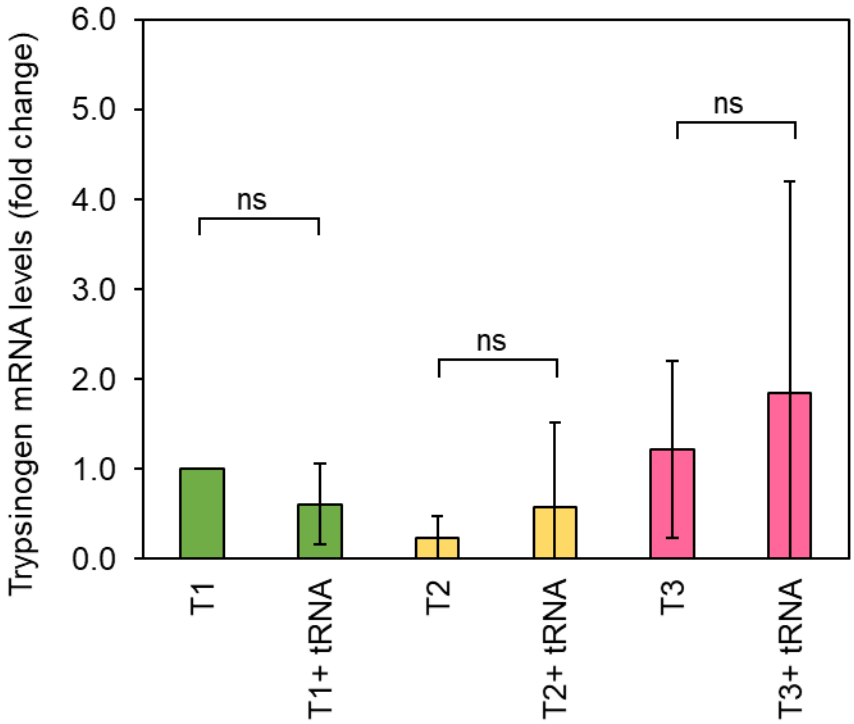


Figure 4.21: Comparison of relative trypsinogen transcript levels in strains with and without tRNA-Leu-GAG.

Transcript levels were determined by RT-qPCR using TAF10 as an internal control and analysed using the delta-delta-Ct method. Transcript levels of each gene variant are expressed relative to P1. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per variant. Student's T-tests were carried out to test for statistical differences between conditions. [ns] denotes 'not significant'.

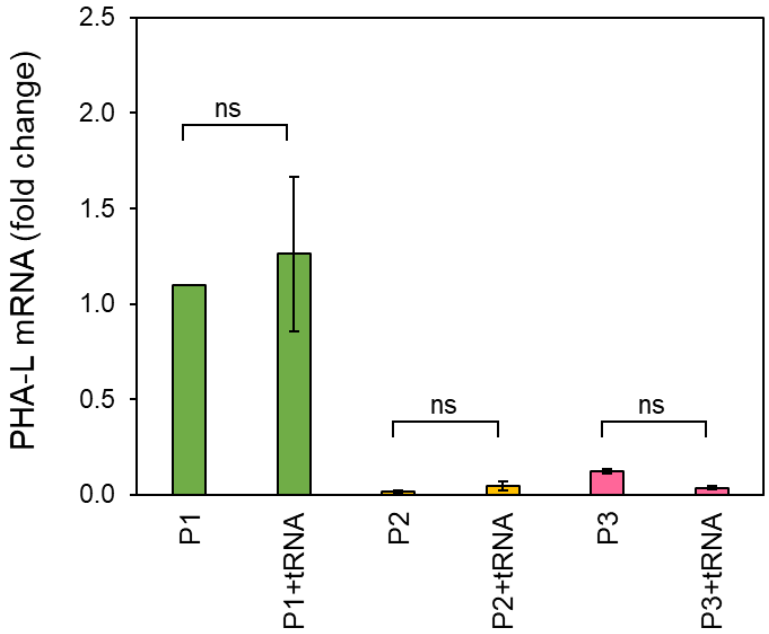


Figure 4.22: Comparison of relative PHA-L transcript levels in strains with and without co-expression of *S. cerevisiae* tRNA-Leu-GAG.

Transcript levels were determined by RT-qPCR using TAF10 as an internal control and analysed using the delta-delta-Ct method. Transcript levels of each gene variant are expressed relative to P1. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per variant. Student's T-tests were carried out to test for statistical differences between conditions. [ns] denotes 'not significant'.

4.2.8 Reliance on wobble-mediated decoding could improve trypsinogen production in *K. phaffii*

Recombinant secreted trypsinogen levels were measured to determine the effect of the exogenous tRNA on recombinant protein production. Co-expressing the tRNA with T1 and T2 significantly improved production approximately 1.5-fold, but this was not observed for T3 (Figure 4.23). This suggests that the previous interpretations that there was a secretory bottleneck impeding changes in yields was either incorrect or was somewhat alleviated by addition of the tRNA.

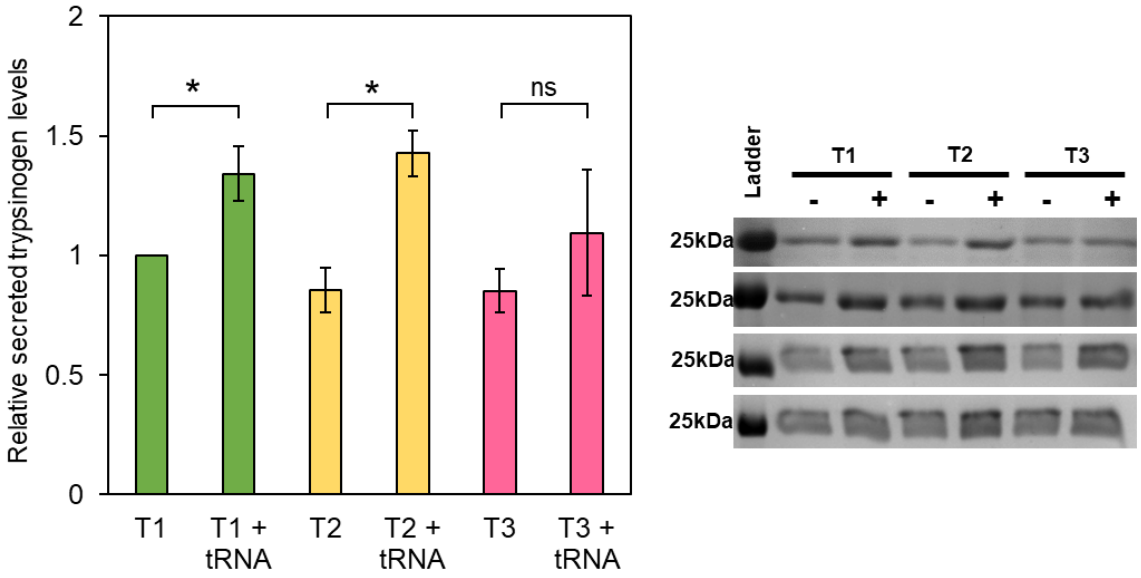


Figure 4.23: Secreted trypsinogen yields in strains with and without co-expression of S. cerevisiae tRNA-Leu-GAG.

Secreted trypsinogen levels were assessed by SDS-PAGE and Coomassie blue staining. Four independent biological repeats were obtained. Means are displayed relative to T1 and error bars indicate SEM. Student's T-tests were carried out to test for statistical differences between conditions. 'ns' denotes 'not significant'.

Considering that the tRNA had no significant effect on the amount of secreted trypsinogen per mRNA for T1 and T2 supports the notion that addition of the tRNA did not improve translation rates but may have affected the production of some endogenous proteins which affect recombinant protein production (Figure 4.24). Interestingly, T3 displayed a decrease in the amount of protein produced per mRNA. which may mask the global effect of the tRNA on recombinant protein production. The data suggests that the higher rate of protein production

per mRNA in T3 is likely to be driven by introduction of wobble, as this is abolished to the T2 baseline upon addition of the tRNA.

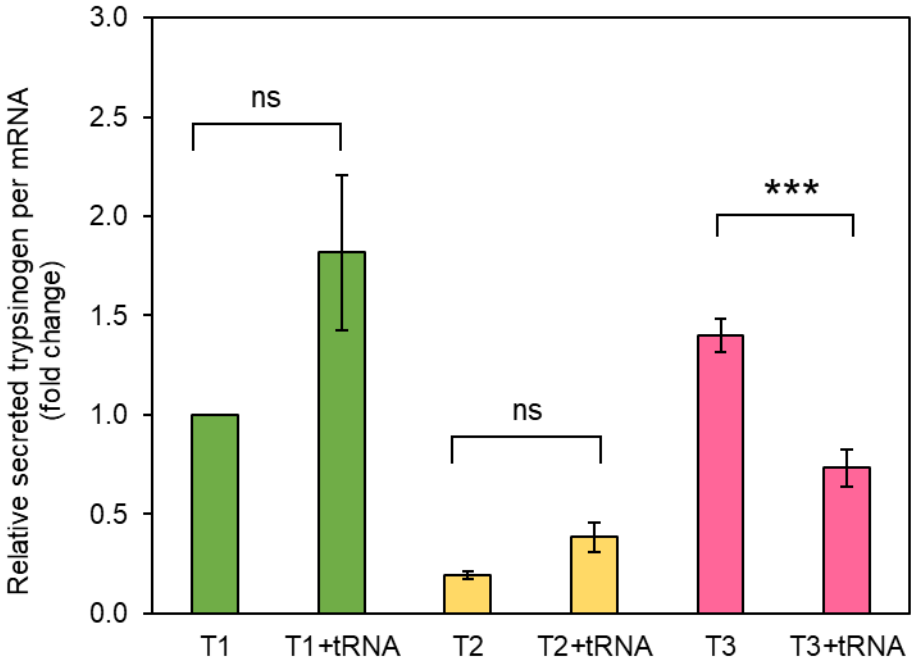


Figure 4.24: Secreted trypsinogen per mRNA in strains with and without co-expression of *S. cerevisiae* tRNA-Leu-GAG.

Secreted trypsinogen levels per mRNA were assessed by SDS-PAGE and Coomassie blue staining. Four independent biological repeats were obtained. Means are displayed relative to T1 and error bars indicate SEM. Student's T-tests were carried out to test for statistical differences between conditions. 'ns' denotes 'not significant'.

4.2.9 Moderate reliance on wobble-mediated translation improves PHA-L production in *K. phaffii*

In contrast with the results observed for trypsinogen, no improvement of secreted protein levels was observed upon addition of the exogenous tRNA (Figure 4.25). This suggests that the tRNA could improve endogenous processes which are specific to trypsinogen production and not PHA-L production.

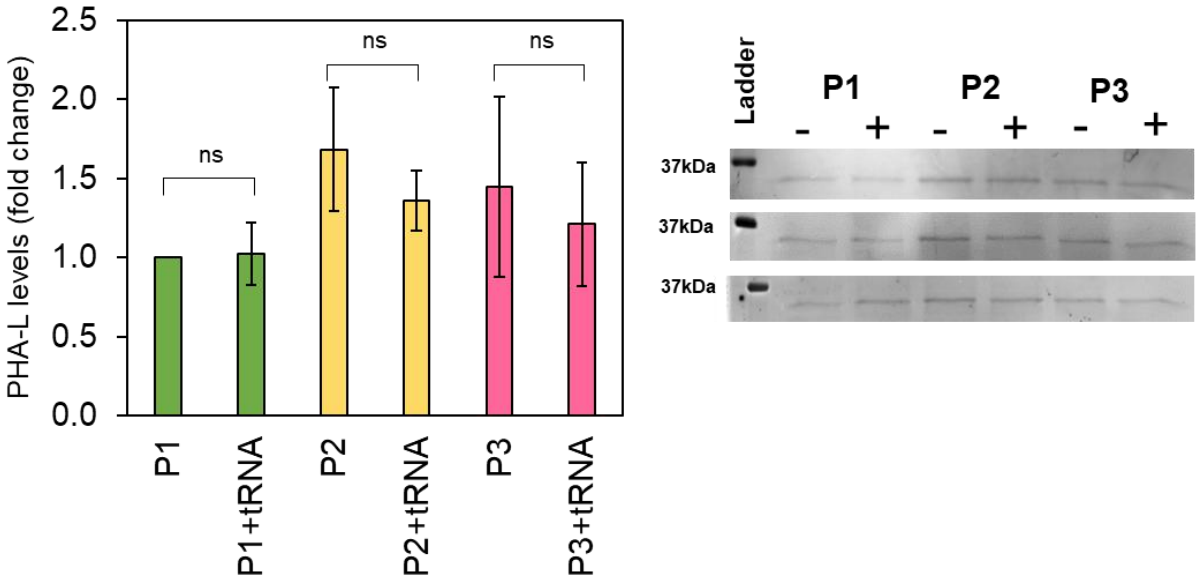


Figure 4.25: Secreted PHA-L levels strains with and without co-expression of *S. cerevisiae* tRNA-Leu-GAG.

Secreted PHA-L levels were assessed by SDS-PAGE and Coomassie blue staining. Four independent biological repeats were obtained. Means are displayed relative to P1 and error bars indicate SEM. Student's T-tests were carried out to test for statistical differences between conditions. 'ns' denotes 'not significant'.

Introducing the tRNA does not affect the protein per mRNA for P1 and P2, but it causes a significant decline in protein production for P3 (Figure 4.26). This indicates that wobble-mediated translation was advantageous for producing PHA-L in P3, and the addition of the tRNA disrupts this benefit.

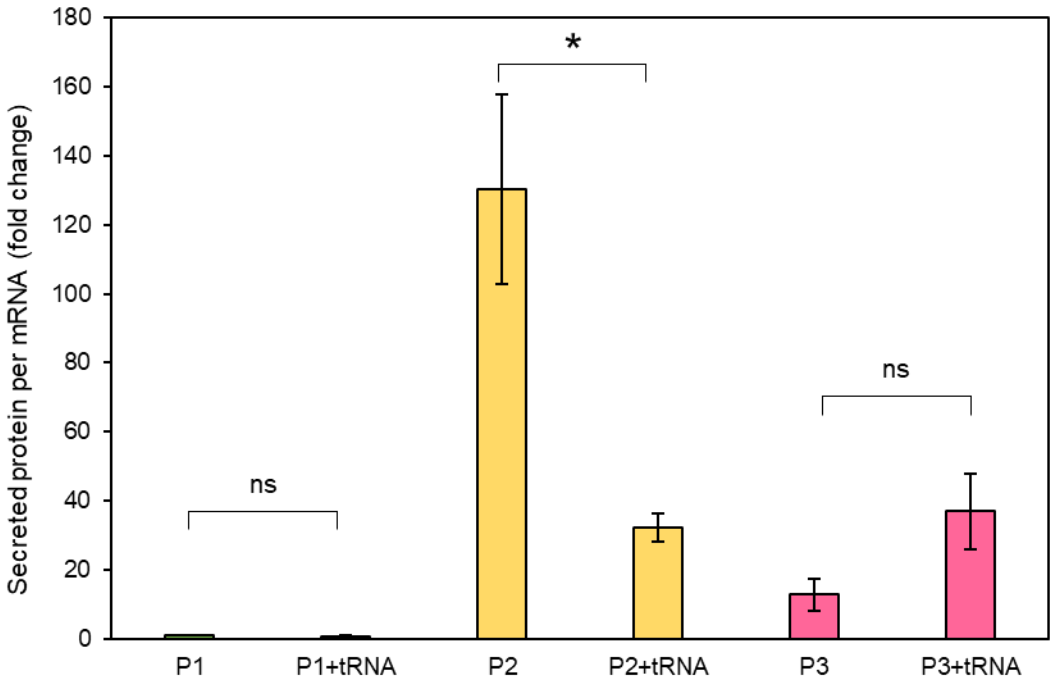


Figure 4.26: Secreted PHA-L levels per mRNA strains with and without co-expression of *S. cerevisiae* tRNA-Leu-GAG.

Secreted PHA-L per mRNA levels were assessed by SDS-PAGE and Coomassie blue staining. Four independent biological repeats were obtained. Means are displayed relative to P1 and error bars indicate SEM. Student's T-tests were carried out to test for statistical differences between conditions. 'ns' denotes 'not significant'.

4.2.10 Exogenous tRNAs may have an effect on protein secretion

While extracellular T1 levels increased with addition of the exogenous tRNA, no change in intracellular levels are observed (Figure 4.27), coinciding with no change in the amount of intracellular protein per mRNA (Figure 4.28). This suggests that addition of the tRNA has no direct effect on T1 production as T1 does not contain any CUC codons. The increase in secreted T1 suggests that addition of the tRNA may have improved a part of the recombinant production process, likely a processing or secretion step.

In the absence of the tRNA, intracellular T2 levels were increased when compared with T1 levels (Figure 4.27), suggesting that an introducing wobble may have created a bottleneck in T2 secretion. Slightly increased intracellular protein per mRNA in T2 compared to T1 suggests that the synonymous codon changes may slightly improve production, but addition of the tRNA does not abolish this effect, suggesting that wobble-mediated decoding is not responsible for this (Figure 4.28). Addition of the tRNA does however cause a significant reduction in the amount of intracellular T2 (Figure 4.27), which, together with increased extracellular T2 (Figure 4.23), further implicates a role of the tRNA in alleviating a secretion bottleneck.

Similar to extracellular levels, intracellular T3 levels were not affected by the presence of the tRNA (Figure 4.27, Figure 4.28). The reduction in secreted protein per mRNA upon addition of the tRNA (Figure 4.24), suggested that it alleviated bottlenecks in translation. Introducing the tRNA likely disrupted this optimised translation mechanism, resulting in a decrease in protein production efficiency. This indicates that for T3, the reliance on wobble translation may have provided an advantage in balancing translation rates with the cellular processes involved in protein folding, translocation, and secretion.

This led us to strive to validate the expression levels of the tRNA in order to test the assumption that introducing a single copy of this gene to each strain would mean the gene is expressed similarly in each strain.

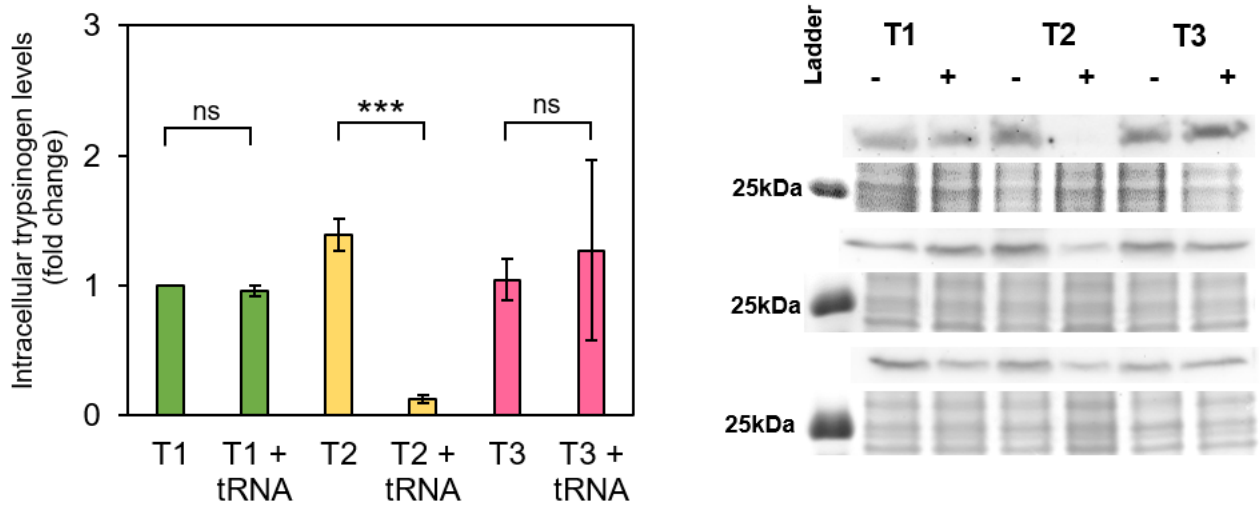


Figure 4.27: Intracellular trypsinogen levels strains with and without co-expression of *S. cerevisiae* tRNA-Leu-GAG.

Crude lysates were obtained by resuspending cells in TWIRL buffer. Lysates were analysed by immunoblot. Signal was normalised to total protein, which was determined by SDS-PAGE and Coomassie stain. Intracellular protein was determined by western blotting and quantified using quantitative densitometry. Mean values are displayed relative to T1. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per strain. Student's T-tests were carried out to test for statistical differences between conditions. 'ns' denotes 'not significant'. '***' denotes a p value < 0.001.

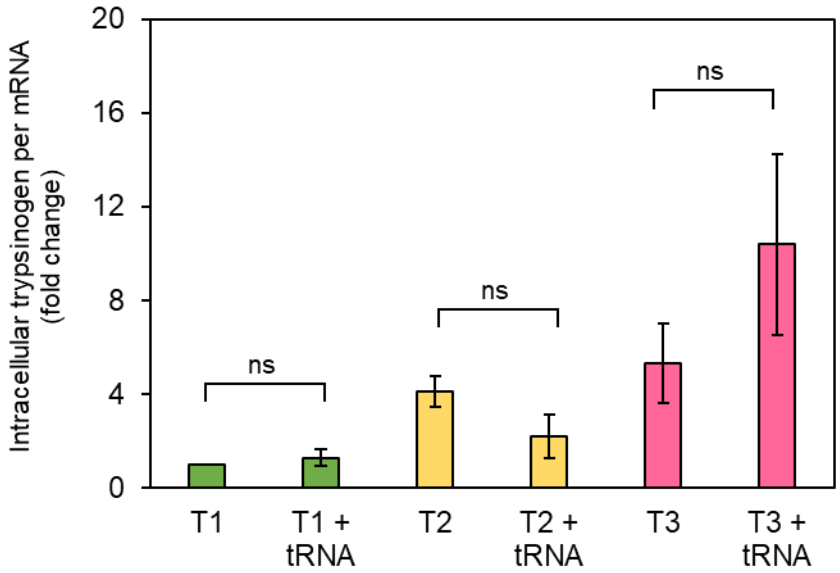


Figure 4.28: Intracellular trypsinogen per mRNA strains with and without co-expression of *S. cerevisiae* tRNA-Leu-GAG.

Intracellular protein for T1, T2, and T3 variants were assessed in the presence and absence of exogenous tRNA-Leu-GAG. Intracellular protein per mRNA was measured for each gene variant to assess the impact of the tRNA on translation efficiency. Student's T-tests were carried out to test for statistical differences between conditions. 'ns' denotes 'not significant'.

Figure 4.29 and Figure 4.30 show that for PHA-L production, addition of the tRNA had no effect on intracellular protein, nor intracellular protein per mRNA for any of the gene variants, suggesting that wobble decoding does not play a significant role in regulating PHA-L production. However, high levels of intracellular P1 were observed compared to P2 and P3, suggesting that P1 is the least efficiently secreted (Figure 4.29). Intracellular protein per mRNA suggests that P2, and to a lesser extent P3, is more efficiently produced compared to P1 (Figure 4.30). Therefore, the synonymous codon changes improved both production and secretion of PHA-L, but not through the introduction of wobble-mediated translation.

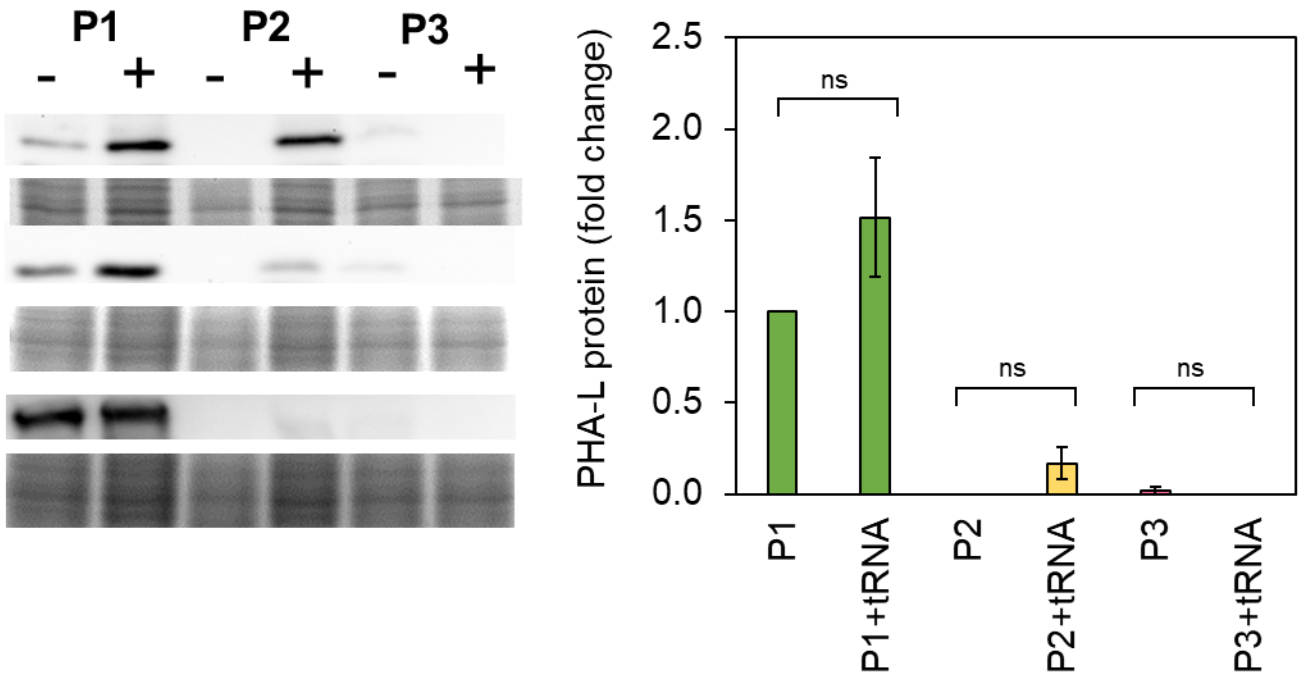


Figure 4.29: **Intracellular PHA-L levels strains with and without co-expression of *S. cerevisiae* tRNA-Leu-GAG.**

Crude lysates were obtained by resuspending cells in TWIRL buffer. Lysates were analysed by immunoblot. Signal was normalised to total protein, which was determined by SDS-PAGE and Coomassie stain. Intracellular protein was determined by western blotting and quantified using quantitative densitometry. Mean values are displayed relative to P1. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per strain. Student's T-tests were carried out to test for statistical differences between conditions. 'ns' denotes 'not significant'.

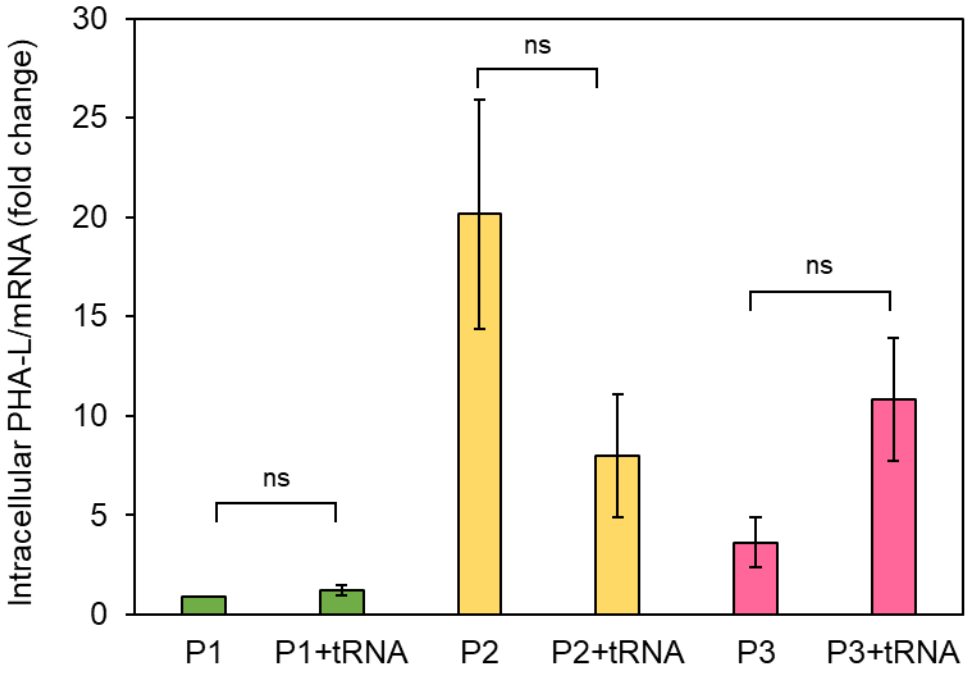


Figure 4.30: Intracellular PHA-L per mRNA strains with and without co-expression of *S. cerevisiae* tRNA-Leu-GAG.

Intracellular protein for P1, P2, and P3 variants were assessed in the presence and absence of exogenous tRNA-Leu-GAG. Intracellular protein per mRNA was measured for each gene variant to assess the impact of the tRNA on translation efficiency. Student's T-tests were carried out to test for statistical differences between conditions. 'ns' denotes 'not significant'.

4.2.11 Exogenous tRNA has no effect on endogenous tRNA expression

Investigations so far indicate that synonymous codon changes, but not wobble-mediated translation, can be beneficial for recombinant protein production. Introduction of the exogenous tRNA appears to have confounding effects depending on the transgene expressed, and likely has a role in regulating the expression of endogenous proteins. To further dissect the effect of introducing the tRNA, its effect on endogenous tRNA-Leucine expression was assessed. Indeed, there was no significant difference in tRNA-Leucine expression between the different gene variants for trypsinogen and PHA-L, meaning there is no sign of the cell adjusting for the increased demand of this anticodon (Figure 4.31, Figure 4.32). The primers designed were redundant for all leucine tRNAs and thus do not indicate changes in the proportions of each isoacceptors. No change was seen from the addition of the exogenous tRNA-Leucine, again, suggesting that tRNA leucine supply is constant.

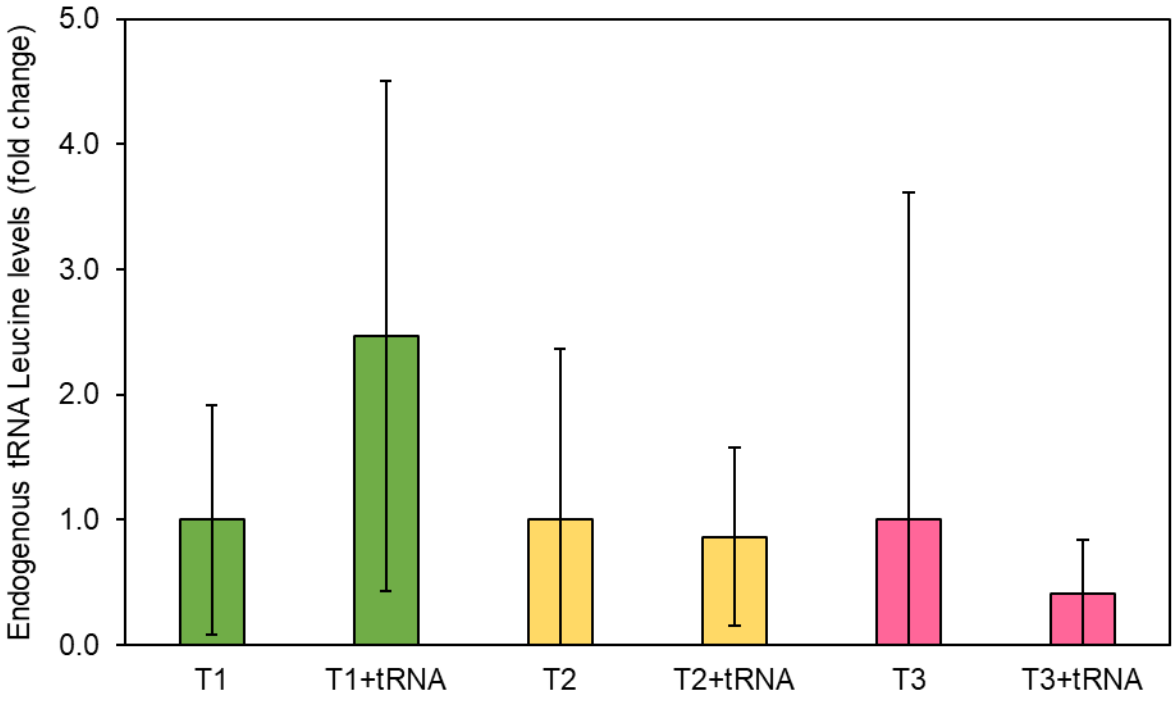


Figure 4.31 Endogenous *K. phaffii* tRNA-Leucine expression in trypsinogen-expressing strains with and without the exogenous tRNA-Leu-GAG gene.

Pre-tRNA levels were determined by RT-qPCR using TAF10 as an internal control and analysed using the delta-delta-Ct method. Transcript levels of each gene variant are expressed relative to T1. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per strain.

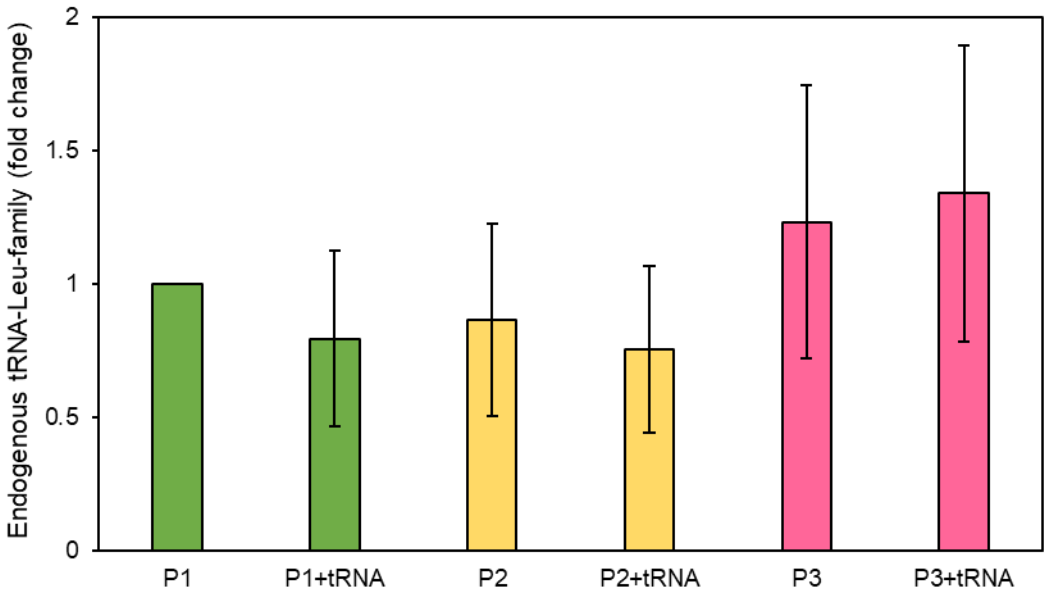


Figure 4.32: Endogenous *K. phaffii* tRNA-Leucine expression in PHA-L-expressing strains with and without the exogenous tRNA-Leu-GAG gene.

Pre-tRNA levels were determined by RT-qPCR using TAF10 as an internal control and analysed using the delta-delta-Ct method. Transcript levels of each gene variant are expressed relative to P1. Error bars depict the standard error of the mean (SEM), with n = 3 biological replicates per strain.

4.3. Discussion and conclusions

The work described in this chapter outlines the first attempts to investigate the effect of supplementing tRNAs to improve translation rates of recombinant proteins in *Komagataella phaffii*. Through assessing the production of genes with synonymous codon changes which introduce wobble – effectively 'de-codon-optimising' genes for production in *K. phaffii*, some of the assumptions made when transgenes are codon-optimised could be interrogated.

4.3.1 Synonymous codon changes can affect transgene transcript levels

First, it was found that synonymous codon changes to codon-optimised genes for trypsinogen and PHA-L can cause a decrease in transgene transcript levels. This was seen specifically for P2, whereby mRNA levels were reduced compared to P1. However, no change in transcript levels were observed for P3, nor for any of the trypsinogen variants. Considering the effect of the synonymous codon changes is gene-specific suggests that the CUC codon did not drastically affect transcription rates. Instead, it implies that unfavourable secondary structures could have formed for the P2 transcript which affected its stability.

4.3.2 Synonymous codon changes have no effect on recombinant titres

Non-optimal synonymous codon substitutions which introduce a reliance on wobble decoding, had no effect recombinant protein titres for both PHA-L and trypsinogen. This was surprising as it was expected that introducing wobble would be limiting for protein production and reduce titres. This suggests that wobble-mediated decoding to the extent used in this study does not have a significant impact on translation rates and protein production efficiencies in these two recombinant proteins. This finding undermines the concept that codon optimisation universally increases production, as the codon-optimised genes did not outperform the variants generated. This demonstrates that transgenes with non-optimal codons can be produced with similar efficiencies to codon-optimised transgenes.

Codon optimisation is routinely implemented when designing transgenes for recombinant protein production, so there could be benefits to reducing the emphasis of this stage and exploring alternative principles to decide transgene codon compositions, such as codon pair optimisation. Codon pair optimisation (CPO) is a strategy for improving recombinant protein production by considering the advantageous and deleterious effects of codon-pair contexts (283). It exploits the principle that there are not only patterns in codon usage which suggest optimal codons for increased production, but there are also favourable sequences of codons which promotes increased translational efficiencies. In *K. phaffii* production platforms, it has been shown that genes optimised based on codon pair bias outperformed genes optimised based on codon usage bias in terms of recombinant titres.

4.3.3 Introducing an exogenous tRNA can improves secreted protein levels

An interesting observation made in this chapter was that co-expression of the *S. cerevisiae* tRNA-Leucine-GAG gene with the trypsinogen gene variants T1 and T2 significantly increased their secreted titres. This was surprising as T1 was codon-optimised and therefore had no direct reliance on the tRNA for recombinant protein production. This implies a potential global effect of the tRNA which alleviated a bottleneck or improved the efficiency of another process in the protein production pipeline. Interestingly, it had no effect on T3, which can be explained by a reduction in secreted protein per mRNA. This suggests that the synonymous codon changes here introduced a bottleneck which countered the beneficial effect of introducing the tRNA-Leu-GAG gene. For example, T3 has an increased demand for the tRNA compared to T2, so perhaps supply became limiting for T3 production which countered the other effects of the tRNA. This hypothesis could be interrogated by adding additional copies of the tRNA would a) increase its supply and b) result in an increase in T3 titre.

For PHA-L, no net change in titre was observed upon co-expression of the *S. cerevisiae* tRNA-Leucine-GAG gene. However, the amount of secreted protein per mRNA for P2 was significantly reduced. Taking this together with the finding that there was no change in intracellular protein, a possible explanation for this is that addition of the tRNA reduced the efficiency of the P2 to be secreted. This supports the notion that introduction of an exogenous tRNA can have multiple, potentially confounding, effects on recombinant protein production. Dissecting these effects could be helpful in developing this concept in order to optimise the approach for maximum benefit.

4.3.4 Introducing an exogenous tRNA may have global impacts on protein production processes

A key finding in this work was that introduction of the exogenous tRNA caused an increase in secreted T2, combined with a reduction in intracellular T2. There are numerous mechanisms that could be contributing towards this finding. For example, the reduction in intracellular T2 could indicate reduced translational efficiency and recombinant protein production. However, no change in the production efficiency was observed as evidenced by there being no change in the amount of protein per mRNA. Therefore, the same amount of protein is being produced, but the amount retained in the cell is lower. This suggests that potentially the rate of secretion of the protein has increased with addition of the tRNA. Again, this suggests that the tRNA can not only directly affect translation of the recombinant protein but may have wider beneficial impacts on protein processing and secretion.

5. Chapter 5: Exploring the global effects of exogenous tRNAs on recombinant protein production

5.1. Introduction

The potential benefits of introducing an exogenous tRNA or increasing endogenous tRNA gene copy numbers are vast. Considering the global effects of this change provides a comprehensive view on whether the introduction of tRNAs and their genes is a promising strategy for improving recombinant protein production. This chapter aims to elucidate the effects of introducing a single anticodon in modulating the production of the vast array of endogenous proteins. Building upon the previous chapter, the GAG anticodon is explored here, which will provide insights into the global reliance on wobble-mediated translation.

5.1.1 Co-ordinated protein regulation

As described in the previous chapter, codon usage bias describes the phenomenon that despite synonymous codons having the same coding potential, some genes prefer certain codons over others (284). Patterns in codon usage can be linked to tRNA supply, as it has been found that overrepresented codons in highly expressed genes are typically those which have highly expressed tRNAs carrying the cognate anticodon (285). On the other hand, lowly expressed genes can show no preferences, but often actually prefer codons decoded by rare tRNA species (286). This provides evidence that codon selection is not random, and that natural selection has a force in incorporating codons for higher as well as lower translational efficiencies.

In addition to tRNA supply, patterns in codon usage indicate the potential for coordinated regulation of proteins sharing functional properties. For example, it was shown that in *S cerevisiae,* codon-biased translation occurs to enhance the cells ability to cope with oxidative damage (119). More specifically, there is an upregulation of tRNA-Leu-CAA with a methyl group is added to the fifth carbon of the cytosine base (m^5C) at the wobble position, enabling more wobble-mediated translation of genes containing the TTG codon. Indeed, an increase in protein expression was observed from the TTG-enriched ribosomal protein gene *RPL22A*, but not its unenriched paralog (119). This demonstrates a role of wobble-mediated translation in coordinating the upregulation of proteins required for cell survival.

Introducing an exogenous tRNA, such as the tRNA-Leu-GAG, may cause competition with endogenous tRNAs, which could disrupt the cell's natural regulatory mechanisms in translation. Specifically, this exogenous tRNA may preferentially engage in Watson-Crick base pairing with codons, reducing the opportunity for wobble base pairing. This chapter aims to address this by investigating the functions of genes with increased demand for the tRNA, to determine whether introducing this tRNA may disrupt any cellular processes relevant to recombinant protein production.

5.1.2 Ribosome mapping

Investigating the role of wobble-decoding of the CUC codon on translational efficiencies (TE) in endogenous *K. phaffii* proteins required the analysis of ribosome-profiling data in tandem with RNA sequencing data. Ribosome profiling quantifies the genome-wide ribosome occupancy of mRNA transcripts. Coupled with the integration of matched RNA sequencing data, the translation efficiency of genes can be calculated to reveal mechanisms of translational regulation (287).

RNA sequencing (RNA-seq) quantifies fragments of RNA molecules to assess the level of gene transcription. This is achieved by mapping the sequencing reads to the genome and counting them to quantify the absolute expression levels of each gene. Transcription rates can be inferred from this data, illustrating the regulatory landscape at the transcriptional level. However, it gives no indication of how much protein is produced from these transcripts. Proteomics data provides absolute protein levels, giving a fuller picture of the active components in the cell but, being subject to degradation and recycling, does not capture the full picture when it comes to protein production rates. Ribosome profiling (Ribo-seq) offers the opportunity to study translational regulation, which directly correlates to protein synthesis rates, in a quantitative manner. Like RNA-seq, it maps sequenced reads to the genome, but instead of total transcript levels, it provides a count of ribosome-protected RNA fragments. Therefore, it provides information on the positioning and number of ribosomes per read.

5.1.3 Translation efficiencies (TE)

One of the reasons why mRNA transcript levels can be poor indicators of protein levels is that there is the potential for many protein molecules to be produced from just one mRNA molecule (258). A major contributing factor to this is that an mRNA molecule is typically translated simultaneously by multiple ribosomes, forming a polyribosome or polysome, with each ribosome within the polysome independently synthesising its own polypeptide (258). This can be measured by coupling RNA-seq and Ribo-seq data to calculate translational efficiency (TE). This in essence describes the number of ribosomes per transcript, as it is a ratio of ribosome-protected reads over total reads (287). Translational efficiencies can be used to determine the effect of unfavourable codons on ribosomal transit and pausing.

Disruptive effects on ribosome transit are amplified due to the fact that a single mRNA molecule can be translated by more than one ribosome (258). The distribution of ribosomes within the polysome are typically moderately sparse as they are unable to cannot overtake each other when decoding an mRNA molecule (258), but ribosome collisions and queuing are still possible, and can reduce translation rates substantially (256).

Non-optimal decoding has been demonstrated to play a role in ribosome queuing. On one hand, a meta-analysis of ribosome profiling data in *S. cerevisiae* shows that ribosomes do discernibly pause at non-optimal codons (288). This suggests that codons where a cognate anticodon is not available can limit translational elongation by creating a pause whereby the ribosome waits for the correct anticodon to be available. It was found that loss of anticodon wobble uridine (U_{34}) modifications in a subset of tRNAs leads to ribosome pausing at their cognate codons in *S. cerevisiae* (289). This suggests that a reliance on wobble can prevent ribosomal pausing, by facilitating the less efficient form of decoding rather than waiting for an optimal anticodon to arrive. Ultimately, higher adaptation to the tRNA pool decreases ribosomal queuing (290), demonstrating that while wobble is helpful in facilitating translation when no better alternative is available, standard canonical decoding is required for optimal rates of translation.

5.1.4 Overview of the analytical approach

This chapter takes a bioinformatics approach to investigate the global role of wobble-mediated translation and generate hypotheses as to what processes are affected by expression of the *S. cerevisiae* tRNA-Leu-GAG in *K. phaffii* expression platforms as explored in the previous chapter of this thesis. Genomic data obtained from NCBI and coupled Ribo-seq and RNA-seq data obtained from Dalvie et al were harnessed to investigate the reliance of *K. phaffii* cells on wobble-mediated translation (291).

Data analysed in this chapter were obtained from *K. phaffii* strains expressing two recombinant protein variants: P[4] and P[6] (291). These variants are antigens of rotavirus, developed for the manufacture of subunit vaccines, a class of vaccines that are becoming increasingly prevalent due to their efficacy and safety (292-294). The production of subunit vaccines were investigated as a biologic rising in popularity due to their relatively simple structure, enabling their production to meet the exceedingly formidable global demand for vaccine supply (295). Despite P[4] and P[6] having largely homologous amino acid sequences, they presented unique manufacturing challenges which led the study to take an analytical approach towards optimising their production (291).

Transgenes encoding P[4] and P[6] were codon-optimised for production in *K. phaffii* and were tagged with a signal peptide to facilitate secretion into the culture supernatant. Strains expressing these proteins had comparable transgene transcript levels, yet they exhibited vastly different yields. Secreted P[4] production reached a titre of ~ 50 mg/L but exhibited several low-molecular weight variants, whereas P[6] titres were barely detectable, both intracellularly and extracellularly (291).

To address the low molecular weight variants in P[4], its codon usage was altered to remove secondary structures in the mRNA transcripts and alleviate potential ribosomal stalls and truncations during protein translation. This approach proved to be successful, substantially increasing the proportion of full length P[4] (291).

RNA-seq data revealed that during P[6] production, genes associated with cytoplasmic translation were particularly highly expressed, followed by gene sets related to cell wall integrity and DNA repair, which are reported signals of stress in *K. phaffii*. Utilising the ribosome-mapping data, they observed ribosomal stall sites with over fourfold higher occupancy on the P[6] transcript relative to P[4]. Additional stall sites present in P[6] which was not seen in P[4] were observed. The ribosome mapping data was used to inform modifications to the amino acid sequence of the protein, increasing its homology with P[4]. A single amino acid substitution was enough to see readily detectable levels of secreted P[6] via SDS PAGE (291).

Considering that P[4] and P[6] have different significant bottlenecks serve as an interesting model for assessing the demands of the *S. cerevisiae* tRNA-Leu-GAG under different conditions. Utilising the RNA-seq and ribosome profiling data produced from this study enables a bioinformatics approach towards assessing the reliance of endogenous proteins on wobble decoding.

5.1.5 Aims and objectives

This chapter will focus on understanding to what extent endogenous proteins in *Komagataella phaffii* depend on wobble decoding and examines the potential impacts of introducing exogenous tRNAs which may abolish global wobble decoding of the CTC codon.

The objectives of this chapter are as follows:

- To determine the extent to which *K. phaffii* cells rely on wobble when producing endogenous proteins
- To quantify global demand for the *S. cerevisiae* tRNA-Leu-GAG
- To identify genes with high demand for the *S. cerevisiae* tRNA-Leu-GAG
- To examine the cellular functions which rely on wobble decoding of the CTC codon
- To determine the effect of different gene variants on global demand for the tRNA-Leu-GAG
- To determine whether the CTC codon causes ribosomal pausing

5.2. Results

To approach the investigation of the effect of an anticodon, RNA-seq and Ribo-seq data were required. This allowed for an approximation of demand to be calculated, which is determined by how many mRNAs are there to decode and how many CTC codons they contain. Ribo-seq data can then be used to investigate the translation of these mRNA molecules, giving an insight into ribosome occupancy and distribution across the mRNA molecule. Datasets were obtained from the Gene Expression Omnibus (GEO) under the SuperSeries GSE159338.

5.2.1 Most endogenous *K. phaffii* genes rely on wobble-mediated translation

The first aim of this study was to look at the *K. phaffii* gene sequences and identify which genes contain the CTC codon. Interestingly, 93.5% of *K. phaffii* genes contain at least one CTC codon, demonstrating the significant reliance of the cell on wobble decoding (Figure 5.1). This suggests that the vast majority of genes could be somewhat affected by addition of the exogenous *S. cerevisiae* tRNA-Leu-GAG gene. The remaining 6.5% of *K. phaffii* genes which do not contain a CTC codon represent a minority of genes which do not rely on wobble to decode leucine codons and are therefore likely not directly affected by the exogenous tRNA.

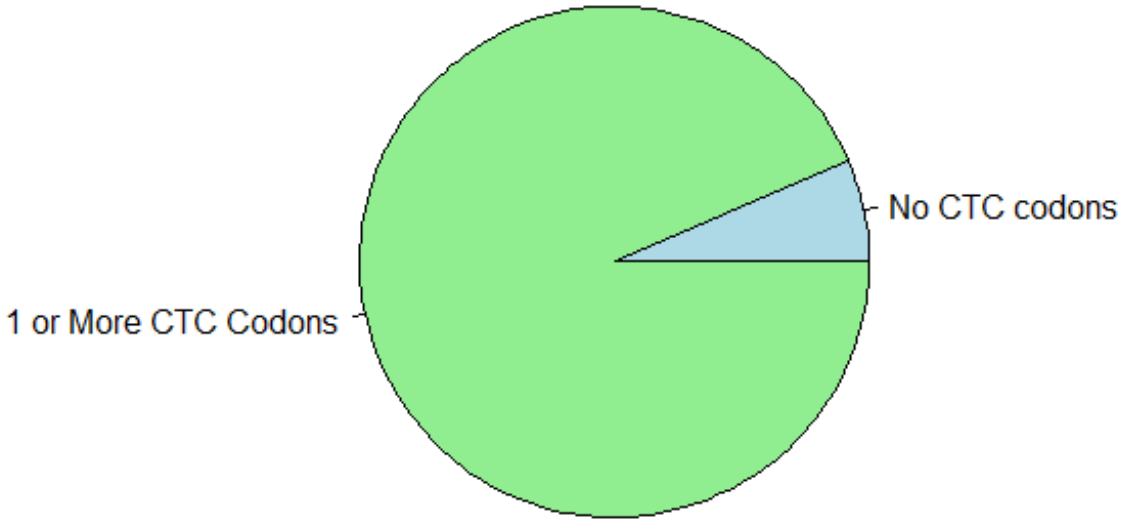


Figure 5.1: Proportion of genes which rely on tRNA-Leu-GAG.
CTC codon counts for each gene was obtained using R package. The vast majority of *K. phaffii* genes contain the CTC codon, thus indicating some reliance on wobble to decode leucine codons.

5.2.2 A small subset of genes drives significant demand for tRNA-Leu-GAG

Next, RNA-seq and ribosome-profiling data for P[4] was used to calculate demand for tRNA-Leu-GAG, and for each gene, percentage demand was calculated to give an indication of each gene's relative demand to total demand. Plotting the percentage demand in decreasing order reveals an elbow plot, demonstrating that most genes have a relatively low demand for the GAG anticodon, and a smaller proportion of genes have a significantly high demand (Figure 5.2). Amongst the genes with highest demand, the top 7 genes account for nearly 1% total demand, suggesting a disproportional amount of demand comes from a handful of genes.

A limitation which is regularly encountered upon studying *K. phaffii* is that many proteins are still uncharacterised, limiting the interpretation of these results. For example, PAS_chr2-2_0489 and PAS_chr2-2_0206 are uncharacterised and have no known function, while PAS_chr3_0626 is only vaguely annotated as an ER membrane protein.

CTA1 is the gene with the highest demand for tRNA-Leu-GAG that is characterised, boasting 20 CTC codons in its gene (Figure 5.2). CTA1 encodes Catalase A, an enzyme that plays a critical role in managing oxidative stress (296). Specifically, it catalyses the breakdown of hydrogen peroxide (H_2O_2) into water and oxygen, reducing the levels of this reactive oxygen species (ROS) within the cell (296). In *S. cerevisiae*, cellular responses to ionising radiation was examined, looking at its two catalases; Catalase A and Catalase T. Catalase A is encoded by the CTA1 gene and is located in peroxisomes, whereas Catalase T is encoded by the CTT1 gene and is widely distributed in the cytosolic matrix (296). Drawing parallels between the wobble-reliant genes involved in responses to oxidative damage in *S. cerevisiae* discussed previously, the dataset was examined for other catalases to determine if there is a more wobble-independent catalase gene which could participate in compensatory relationship with catalase A. Unfortunately, no other catalases could be identified.

GCW14 also exhibited relatively high demand for the tRNA-Leu-GAG, but only has 6 CTC codons (Figure 5.2). The GCW14 gene has been identified in screens for having a strong promoter, thus it may constitute high demand due to high expression levels (297, 298). This gene encodes a potential glycosyl phosphatidyl inositol (GPI)-anchored protein (termed GCW14p), which is predicted to be anchored to the cell wall via a glycosylphosphatidylinositol (GPI) anchor. Beyond characterisation of its promoter, little is understood about the protein itself.

The MIR1 gene also exhibited high demand, due to having 15 CTC codons. It encodes a mitochondrial phosphate carrier protein, responsible for transporting phosphate ions (Pi) across the inner mitochondrial membrane (299). Phosphates are essential for ATP synthesis, so MIR1 plays a critical role in cellular energy production by ensuring that phosphate is available for phosphorylation processes within the mitochondria (299). Therefore, its high demand for the tRNA-Leu-GAG could mean that supplementation of the tRNA improves the cells energy production efficiencies.

Finally, RPL17B has 8 CTC codons, creating a high demand for tRNA-Leu-GAG (Figure 5.2). It encodes the ribosomal protein L17B, which is a component of the large (60S) ribosomal subunit (300). Ribosomal proteins like L17B are essential for the assembly and function of ribosomes, which facilitate protein synthesis in cells (300). The finding that L17B can strongly utilise tRNA-Leu-GAG suggests is possible that supplementation of the tRNA could have increased the translation efficiency of RPL17B mRNA, potentially enhancing ribosome biogenesis and increasing the overall protein synthesis capacity. This could explain the increased recombinant protein production in strains expressing transgenes with no CTC codons.

However, it is also possible that RPL17B plays a more specific role in regulating translation, beyond general ribosome function. Recent studies have shown that ribosomal protein paralogs can influence the selective translation of specific mRNA subsets, contributing to what is known as ribosome heterogeneity (309, 310). While speculative, an RPL17B-enriched ribosome population may preferentially enhance translation of particular transcripts or support stress-responsive translational programs. Further experimental work would be needed to investigate whether RPL17B exerts such specialized effects in this context.

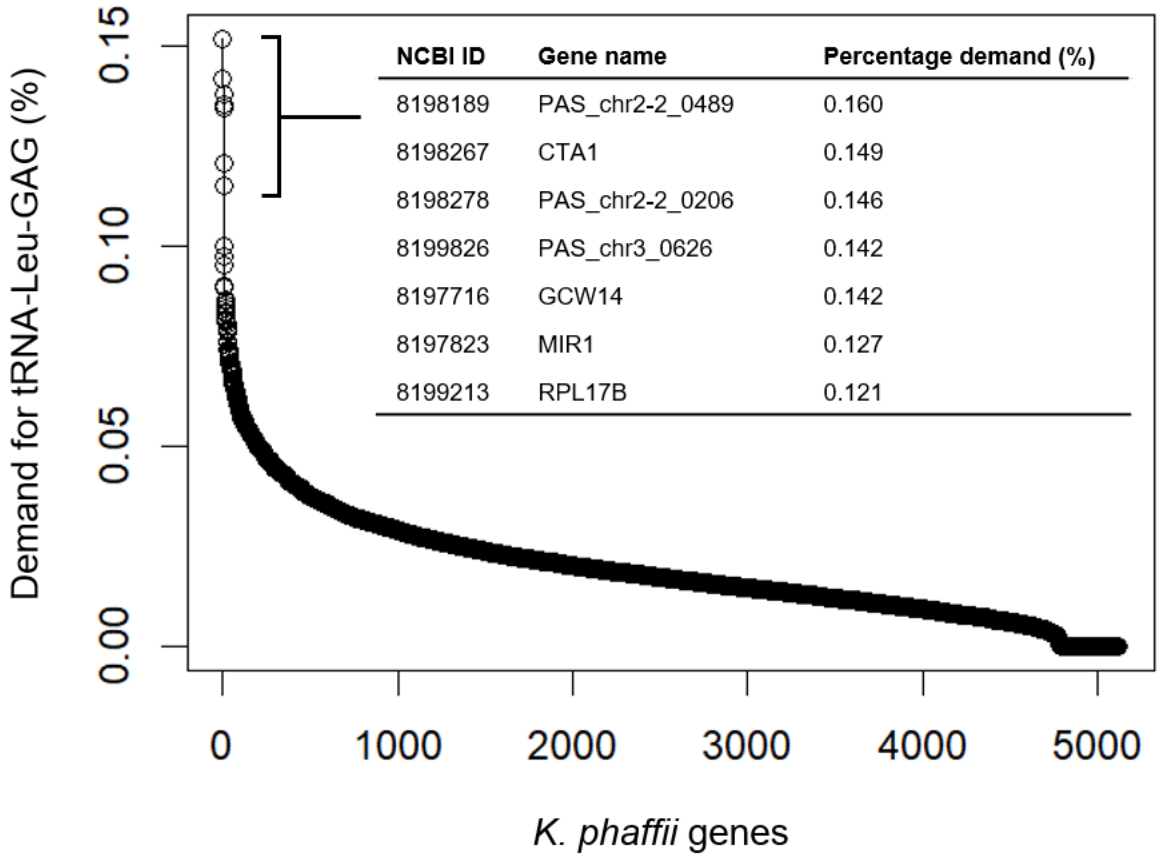


Figure 5.2: Percentage demand for tRNA-Leu-GAG in *K. phaffii* genes.

Demand for tRNA-Leu-GAG was calculated and organised in descending order to indicate the distribution of demand across the *K. phaffii* genome. An elbow plot is revealed, showing a small subset of genes with high demand and a large subset of genes with low demand.

5.2.3 Wobble-mediated translation is utilised in genes regulating metabolism

To determine the functions in which the genes with the highest demand have, genes constituting the top 80% genes were analysed using KEGG Enrichment analysis, constituting 2650 genes (Table 11). Four key KEGG Enrichment pathways were identified, all of which fall under metabolic pathways: Metabolism of co-factors and vitamins, Lipid metabolism, Glycan biosynthesis and metabolism and Amino acid metabolism (Table 11) (Figure 5.3). This suggests that metabolic genes may have a disproportionately high ability to be decoded by wobble. It could be that these genes evolved especially flexible codon structures to allow for their proteins to be produced even when tRNA pools are limited, to promote survival under resource-limiting conditions. Therefore, upon supplementing the tRNA-Leu-GAG gene may limit this adaptative mechanism.

KEGG pathway category	KEGG ID	Gene count
Amino acid metabolism	ppa00280, ppa00380	21
Glycan biosynthesis and metabolism	ppa00510	21
Lipid metabolism	ppa00071	5
Metabolism of cofactors and vitamins	ppa00760	13

Table 11: KEGG categories with enriched gene counts

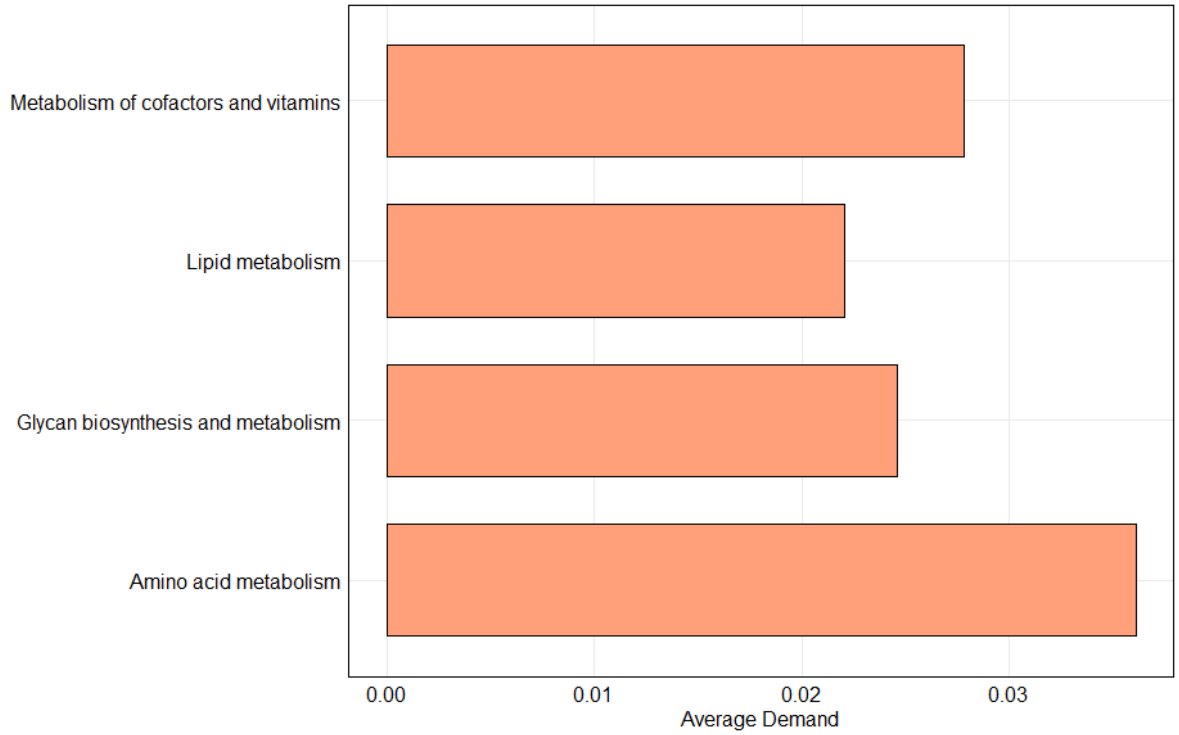


Figure 5.3: KEGG pathway gene set enrichment analysis of genes with high demand for tRNA-Leu-GAG.

KEGG pathway gene set enrichment analysis was conducted in genes constituting the top 80% demand.

5.2.4 ALG genes have a particularly high demand for tRNA-Leu-GAG

Of the enriched categories, 'glycan biosynthesis and metabolism' was flagged as a category of functions which may directly affect recombinant protein production. This is especially relevant for P[4] production as it is a glycosylated recombinant protein (291). Genes within the Glycan biosynthesis and metabolism category were found to be most prevalently involved in N-Glycan biosynthesis. This was of particular interest as findings in the previous chapter implicated a role of the *S. cerevisiae* tRNA-Leu-GAG gene in increasing secreted protein and decreasing intracellular protein for recombinant proteins which did not directly rely on the tRNA to decode the transgene transcripts.

N-glycan biosynthesis involves a series of complex enzymes that help in protein glycosylation, which is crucial for protein folding, stability, and secretion (301, 302), so it could be that introduction of the tRNA improved production of proteins regulating these processes . N-glycan biosynthesis is critical for the quality control of proteins, especially secreted proteins, which pass through the ER.

Indeed, looking at the specific genes which demonstrated high demand in this study were disproportionately represented by ALG proteins (Table 12). ALG proteins, such as ALG1 (Asparagine-Linked Glycosylation protein 1), play crucial roles in the process of N-linked glycosylation (303). ALG proteins work sequentially, each adding specific sugars to build the glycan precursor. ALG1, for example, functions at the ER membrane, catalysing the addition of the first mannose to the lipid-linked oligosaccharide precursor, which is a foundational step for subsequent modifications (304). Disruptions in ALG1 or other ALG proteins can impair glycosylation (304). Therefore, it is possible that supplementing the tRNA-Leu-GAG could improve processing at the ER, especially for glycosylated proteins such as P[4] and P[6]. Considering the significant differences in production rates of the two proteins, differences in demand for the tRNA-Leu-GAG could reveal whether supplementing this tRNA could be a strategy to resolve this discrepancy in yields.

NCBI IDs	Gene Name	Percentage demand (%)
8197083	STT3	0.04
8198377	ALG1	0.04
8200607	ALG3	0.03
8197036	ALG13	0.03
8198108	WBP1	0.03
8200875	OST3	0.03
8196480	SWP1	0.03
8198246	OST2	0.02
8200933	CAX4	0.02
8197521	DPM1	0.02
8199744	ALG8	0.02
8197860	ALG2	0.02
8200435	OST1	0.02
8198420	SEC59	0.02
8198055	ROT2	0.02
8197005	ALG11	0.02
8198371	MNS1	0.02
8197988	ALG6	0.02
8197063	DIE2	0.02
8200812	PAS_chr4_0544	0.02
8197925	CWH41	0.02

Table 12: N-Glycan biosynthesis genes

5.2.5 Recombinant protein with significant production bottleneck has increased demand for tRNA

Comparing total demand for the tRNA-Leu-GAG in strains expressing P[4] and P[6] recombinant proteins reveal that P[6] had a higher demand for the tRNA (Figure 5.4). Compared to P[4], P[6] exhibited significant production bottlenecks as despite exhibiting comparable transgene transcript levels, P[6] titres were barely detectable, both intracellularly and extracellularly (291). Alleviating bottlenecks in translation by mutating the transgene helped to increase P[6] yields but it still did not achieve P[4] titres (291). Therefore, it is possible that there are other bottlenecks in the production process that could be alleviated by supplementation of tRNA-Leu-GAG. In particular, bottlenecks in ER processing may be contributing to the limited production of P[6], as it was found that cells producing P[6] also exhibited increased demand for the tRNA in genes involved in glycan biosynthesis and metabolism (Figure 5.5).

Figure 5.6 shows the demand for tRNA-Leu-GAG for each ALG gene in P[4] and P[6] producing cells. It was observed that some genes displayed minimal differences in demand, including ALG6. Others exhibited more prominent differences, such as ALG2 and ALG3. Interestingly, ALG9 showed a decrease in demand.

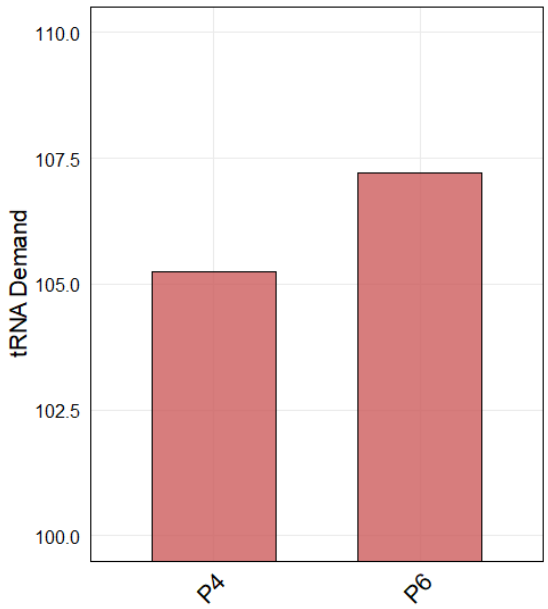


Figure 5.4: Total demand comparison between P[4]- and P[6]- producing cells.

Total demand for tRNA-Leu-GAG was quantified and compared in strains expression P[4] and P[6].

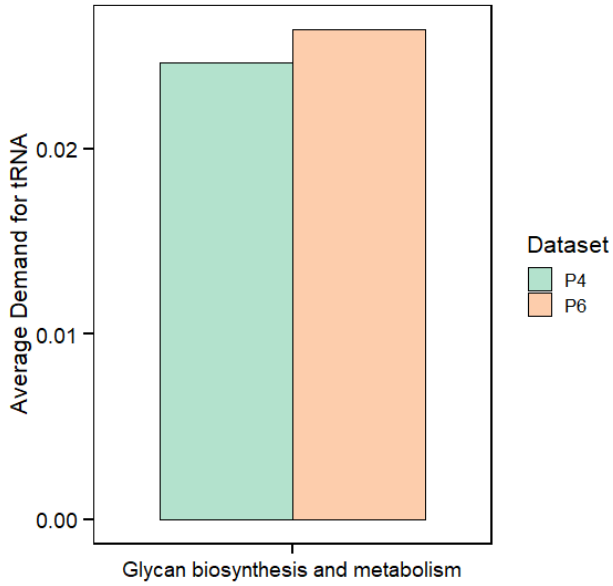


Figure 5.5: Comparison of demand according to KEGG categories

Average demand for tRNA-Leu-GAG per gene within the Glycan biosynthesis and metabolism subcategory was quantified and compared in strains expression P[4] and P[6].

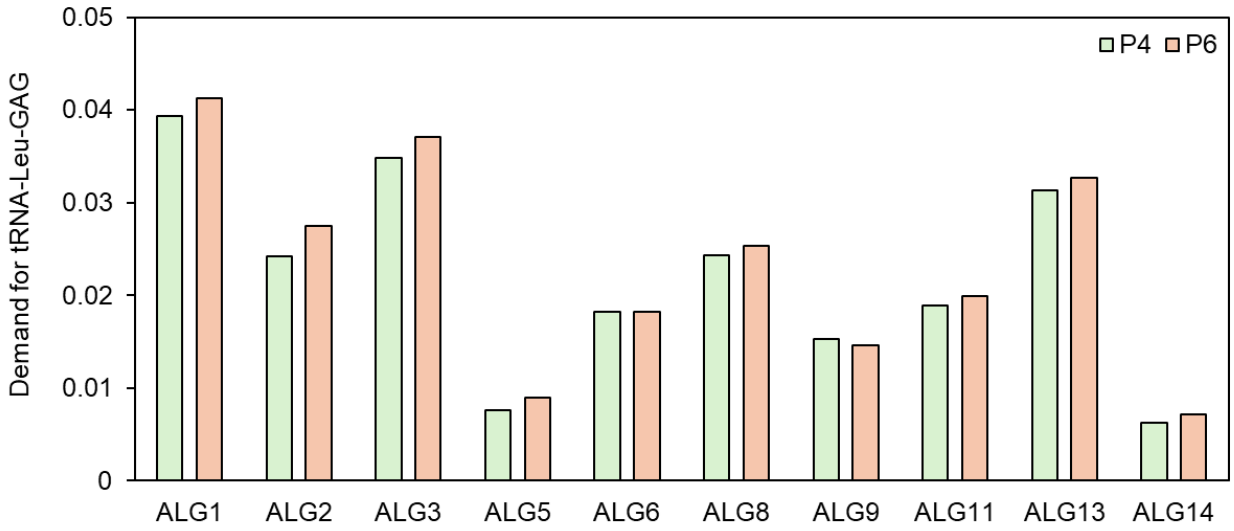


Figure 5.6: Comparison of demand for tRNA-Leu-GAG in ALG proteins in cells producing P[4] and P[6].

Demand for tRNA-Leu-GAG is displayed for each ALG gene in P[4] and P[6] producing cells.

5.2.6 There is no correlation between demand for tRNA-Leu-GAG and ribosome occupancy

In order to determine whether demand for tRNA-Leu-GAG correlates with translation efficiencies, translation efficiencies were calculated using ribosome profiling data. This was then plotted with tRNA demand to investigate the relationship between these two variables. It was hypothesised that an increase in demand for tRNA-Leu-GAG would be associated with reduced translational efficiencies, as the use of wobble would impede translation.

Interestingly, no correlation between tRNA demand and translational efficiencies was found (Figure 5.7). As translational efficiencies are a proxy for number of ribosomes per mRNA transcript, it can be seen that most genes had about one ribosome per transcript. The genes within this area had wide ranges of tRNA demands. It is possible that genes relying on wobble would not exhibit a reduction in ribosome occupancy, but instead experience slower ribosome transit times, due to ribosomal pausing. Therefore, the next step in this investigation involved looking at ribosomal pausing events at specific genes with high reliance on wobble.

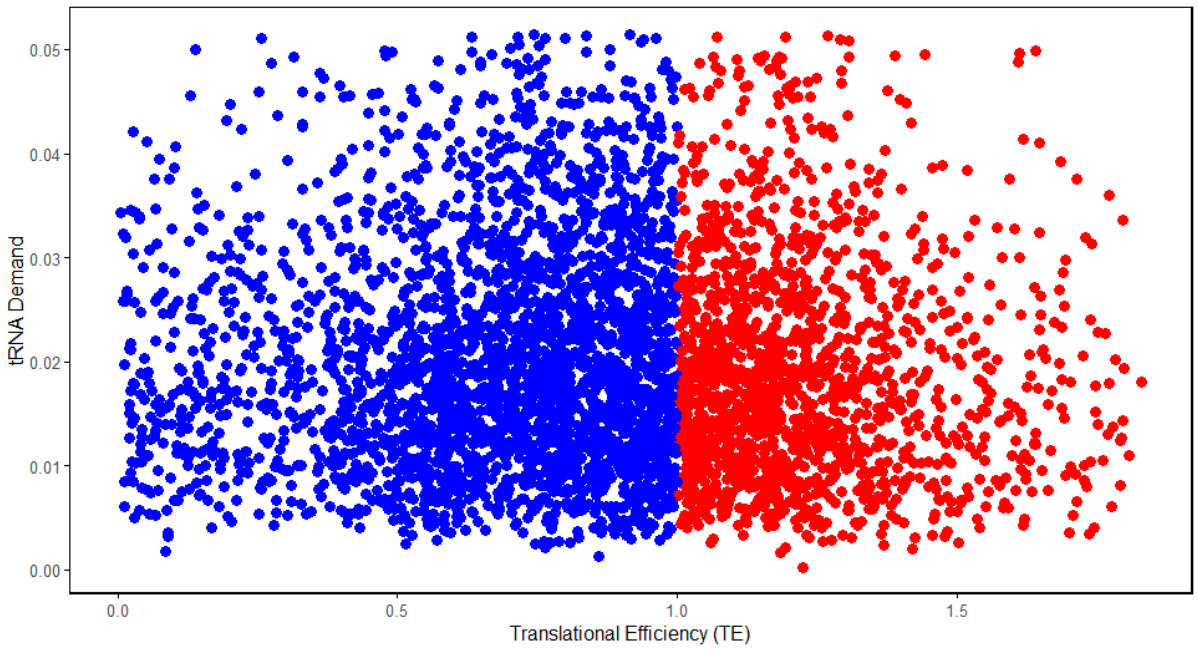


Figure 5.7: Relationship between tRNA demand and translational efficiency.
Translation efficiencies for each gene in *K. phaffii* were calculated using ribosome profiling data and mRNA-seq data. This was plotted against tRNA demand, which is a function of CTC codon count and transcript levels. Genes with translational efficiencies less than 1 were coloured in blue and genes with translational efficiencies more than 1 were coloured in red.

5.2.7 CUC codon on mRNA molecules cause ribosomal pausing

Finally, to determine whether the decoding of CTC leucine codons causes ribosomal pausing, ribosome profiling data for ALG genes was analysed (Figure 5.8). Within the genome browser tool, only ALG5 and ALG7 were annotated, facilitating the identification of their loci. The remaining genes were likely named under other aliases, demonstrating the challenges of data analysis in non-model organism species, which are more likely to use unstandardised gene and protein naming conventions.

Assessing ribosomal occupancy over ALG5 and ALG7 transcripts found that, interestingly, no ribosomal pause sites were observed at CTC codons, but significant pause sites were found at leucine codons TTG. This suggests that ribosomal transit is not impeded by wobble decoding of CTC codons but instead is impeded by TTG codons. This is surprising as the cognate anticodon for TTG, CAA, is the most prevalent in the *K. phaffii* tRNA gene library at three copies. Therefore, it would be predicted that this tRNA would be in highest supply. It is possible that the supply of this tRNA was sequestered by the production of the recombinant protein which was codon optimised for *K. phaffii*. This presents the issue that codon-optimising a transgene for production in *K. phaffii* requires adapting its codon composition to require the most highly used anticodon, thereby creating increased competition for this tRNA, thereby causing tRNA supply to become limiting. Utilising lesser used codons means that there is less competition with endogenous proteins for the tRNA pool and could be beneficial for recombinant protein production.

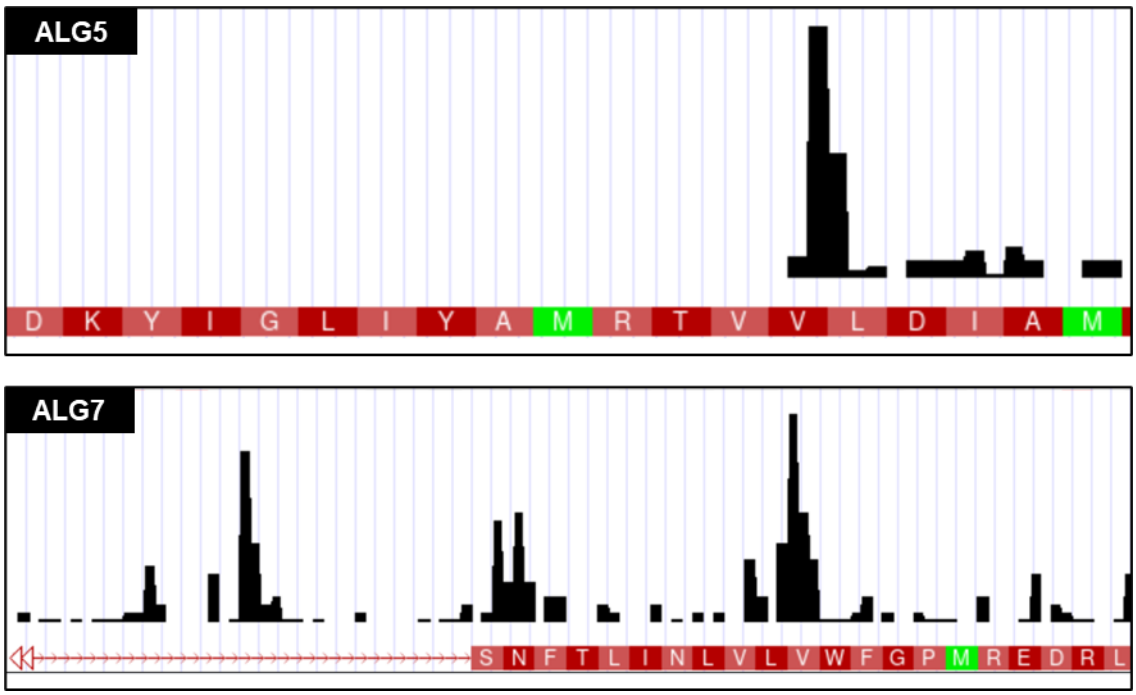


Figure 5.8: Ribosomal pausing at leucine codons in ALG5 and ALG7 genes.

Ribosome profiling data was viewed using the University of California Santa Cruz (UCSC) Genome Browser, against the *K.phaffii* (GS115) (GCF_000027005.1).

5.3. Discussion and conclusions

5.3.1 *K. phaffii* has a high reliance on wobble-decoding

To investigate the role that a single anticodon can play on regulating endogenous proteins and affecting cellular processes, the CUC codon as selected. As discussed in the previous chapter, this codon does not have its cognate anticodon, GAG, explicitly encoded in the *K. phaffii* genome and therefore is translated using wobble-mediated translation. This chapter demonstrates that *K. phaffii* cells heavily rely on wobble decoding, as the majority of its genes contain at least one CTC codon. With only 44 anticodons available to translate 61 sense codons, *K. phaffii* relies on wobble-mediated translation for 17 of these codons, underscoring the organism's dependence on flexible tRNA pairing to meet its translational demands. (68). The lack of understanding of the role of wobble decoding in *K. phaffii* cellular functioning, as well as recombinant protein production, represents a significant knowledge gap which, once filled, could reveal novel avenues for engineering more efficient *K. phaffii* production platforms.

5.3.2 Wobble may coordinate regulation of ALG protein production

This work demonstrates that metabolic genes have a strong reliance on wobble, possibly having evolved a greater flexibility in codon structure by disproportionally using a codon that is missing. This could promote cellular adaptability under fluctuating stress conditions. tRNA supply has been shown to be modulated in response to stress conditions (305-307), so in the context of *K. phaffii* production systems, a reliance on wobble might increase when nutrients become limiting, such as between methanol inductions. In these instances, wobble may play a role in mitigating tRNA scarcity.

Examining the groups of genes which demonstrated high demand in this study revealed that ALG genes comprise a significant proportion of genes which rely on wobble decoding (Table 12). These play crucial roles in the process of N-linked glycosylation (303). This shared pattern of codon usage implicates the potential for coordinated regulation of these proteins. As previously discussed, in *S cerevisiae,* codon-biased translation occurs to enhance the cells ability to cope with oxidative damage (119). More specifically, upregulation of tRNA-Leu-CAA with a methyl group is added to the fifth carbon of the cytosine base (m^5C) at the wobble position enables more wobble-mediated translation of genes containing the TTG codon and causes an increase in protein expression was observed from the TTG-enriched ribosomal protein gene *RPL22A* (119). The shared reliance on the CTC codon means that there could be an endogenous tRNA which can decode this codon and plays an important role in regulating ALG protein production. Implications of this would need to be assessed when considering introducing the cognate anticodon as this may disrupt the coordinated regulatory mechanism and impede the cells' ability to adapt to environmental stressors.

5.3.3 Interpreting ribosomal occupancies

To complete the assessment of the effect of wobble-mediated translation on the production of recombinant proteins, translational efficiencies was determined using ribosome profiling, which not only informs the ribosomal occupancies of the genes but also indicate potential stall sites during elongation that may be caused by wobble-mediated translation (308).

No relationship between tRNA-Leu-GAG demand and ribosomal occupancy on mRNA transcripts was found, potentially due to the confounding implications of ribosomal occupancy. In the analysis of Ribo-seq data, it can be difficult to discern whether high ribosome occupancy due to high translation efficiencies or increased pausing and thus low translational efficiency. In a study looking at ribosomal queuing in *S. cerevisiae*, it was found that up to 20% of the ribosomes are positioned close enough to another ribosome on the transcript to be collected as a single footprint in a ribosome profiling experiment (290). Therefore, ribosomal queues can cause an underestimation of the number of ribosomes on an mRNA molecule, requiring analysis of distributions of ribosomes on the mRNA transcript to be done in tandem.

Analysis of ribosomal stall sites in ALG5 and ALG7 genes revealed ribosome stall sites associated with leucine codons, but instead of being a codon decoded by wobble, it was caused by a codon which is theoretically highly efficiently decoded using Watson-Crick base pairing. This implies that the type of decoding may not be as limiting in protein translation as expected, and that tRNA supply may be a greater issue. This conflicts with the principles of codon-optimisation, whereby transgenes adopt highly used codons, assuming that these are more efficiently translated. TTG is a highly used codon in *K. phaffii*, yet still causes ribosomal stall sites in cells producing recombinant protein, implying that competition for the cognate tRNA pool may be an issue in these cells.

6. Conclusions

The work presented in this thesis explored the roles of tRNAs and their genes in modulating recombinant protein production in *K. phaffii* production hosts and aimed to determine whether they could be utilised as a powerful intervention to improve production efficiencies. To achieve this, the role of tRNA genes (tDNAs) in genome organisation was examined. Evidence of endogenous *K. phaffii* tDNAs with barrier activity was found; testing *K. phaffii* tDNAs for barrier activity in *S. cerevisiae* led to the first ever demonstration of barrier activity in *K. phaffii* genetic elements. It was especially interesting that some of the tDNAs tested outperformed the barrier activity of native tDNAs. Despite this, attempts to utilise a *K. phaffii* tDNA barrier in improving transgene expression saw an abolishment of transgene transcription and recombinant protein production. This discrepancy suggests that tDNA functions are context-dependent, with a number of variables which could influence whether a tDNA is capable of barrier activity, or a distinct but related property, tRNA-gene mediated (tgm) silencing.

The role of epigenetic variables in potentially limiting transgene expression in *K. phaffii* expression systems was also evaluated in this thesis. Chromatin accessibility was assessed using FAIRE-qPCR, which revealed stable accessibility over time. Addition of the tDNA caused an increase in chromatin accessibility which did not cause an increase in transcriptional activity, suggesting that it is not a limiting factor. Despite having a stable chromatin environment, transcriptional activity varies across the cultivation run. These results implied that the inducible *AOX1* promoter may have evolved to remain constitutively open and accessible to carbon source-sensitive transcription factors in order to facilitate a timely and efficient response to changes in nutrient availability. Other epigenetic mechanisms were explored, including position effect variegation (PEV), but the failed tDNA barrier prevented conclusions from being drawn from this. An interesting observation made during this work was that transcriptional responses to methanol inductions was reminiscent of epigenetic memory. This opens up avenues for future explorations of an epigenetic mechanism to exploit for increased transcriptional activity and improved recombinant protein production.

The role of tRNAs in decoding mRNAs was also examined in this thesis. Firstly, it was found that synonymous codon changes introducing a reliance on wobble-mediated decoding did not introduce a bottleneck in recombinant protein production. This undermines the principles of codon optimisation as genes which did not have the necessary tRNA available to efficiently decode its mRNA were not translated less efficiently than a completely optimised gene. Codon optimisation is routinely implemented when designing transgenes for recombinant protein production, so there could be benefits to reducing the emphasis of this stage and exploring alternative principles to decide transgene codon compositions, such as codon pair optimisation.

The addition of tRNAs with novel anticodons was found to be beneficial for trypsinogen recombinant protein titres despite data suggesting wobble-mediated translation is not a bottleneck in the production. This finding was not reflected for PHA-L, however, highlighting that this intervention can be beneficial, but is not yet universal. Further investigations dissecting the effects of the tRNA on trypsinogen production could be used to determine which factors need to align in order to realise the tRNAs full potential. This result further supports the notion that introduction of the tDNA has global impacts on endogenous processes which support recombinant protein production, and may affect processes which disproportionately supported trypsinogen production over PHA-L.

These findings prompted the final investigation of this thesis, where the demand for the tRNA-Leu-GAG gene amongst endogenous proteins were examined as part of initial characterisation of the global impacts of introducing novel tRNAs to *K. phaffii* cells. It is demonstrated that *K. phaffii* cells heavily rely on wobble decoding, as the majority of its genes contain at least one CTC codon and that metabolic genes in particular have a strong reliance on wobble, possibly having evolved a greater flexibility in codon structure by disproportionally using a codon that is missing. Proteins associated with protein processing in the ER and N-glycosylation were flagged as a group of proteins which particularly rely on wobble decoding. An increase in demand from these genes for tRNA-Leu-GAG in cells with a secretion bottleneck points towards a potential strategy for increasing secreted protein yields. Introducing tRNAs which collectively upregulate proteins which facilitate protein secretion could be a powerful application of this intervention.

Overall, this thesis presents numerous possible applications of tRNAs and their genes in improving recombinant protein production. After extensive optimisation and characterisation, it could be possible to introduce a tRNA gene which not only improves transgene transcription but also alleviates translation and secretion bottlenecks using the tRNA product. The combinatorial effect of all the different roles which tRNAs and their genes play in regulating protein production is an exciting avenue of research to pursue, which could not only expand our knowledge of a currently understudied species, but also improve the efficiency of biologic production, thereby increasing the accessibility of thousands of powerful therapeutics.

Abbreviations

DNA - deoxyribose nucleic acid

RNA - ribose nucleic acid

RPP - recombinant protein production

AOX1 - alcohol oxidase 1

pAOX1 – alcohol oxidase 1 promoter

GAP – glyceraldehyde-3-phosphate dehydrogenase

pGAP - glyceraldehyde-3-phosphate dehydrogenase promoter

mRNA – messenger RNA

FDBK – Fujifilm Diosynth Biotechnologies UK

IBioIC – Industrial Biotechnology Innovation Centre

ER – Endoplasmic reticulum

OD_{600} - Optical density measured at 600 nm

PCR – polymerase chain reaction

S. cerevisiae – Saccharomyces cerevisiae

S. pombe – Schizosaccharomyces pombe

E. coli – Escherichia coli

P. pastoris – Pichia pastoris

K. phaffii – Komagataella phaffii

Appendix

Boundary tDNAs in *K. phaffii*	
tRNA-Lys-CTT-2-1	chr1.trna1
tRNA-Trp-CCA-1-1	chr1.trna51
tRNA-Gly-TCC-2-1	chr1.trna3
tRNA-Gln-TTG-1-1	chr1.trna5
tRNA-Pro-TGG-1-1	chr1.trna50
tRNA-Pro-CGG-1-1	chr1.trna49
tRNA-Ser-TGA-1-2	chr1.trna48
tRNA-Ala-AGC-1-2	chr1.trna44
tRNA-Asn-GTT-1-1	chr1.trna43
tRNA-His-GTG-1-1	chr1.trna39
tRNA-Gly-GCC-2-3	chr1.trna14
tRNA-Val-CAC-1-1	chr1.trna15
tRNA-Ile-AAT-1-2	chr1.trna19
tRNA-Thr-AGT-2-1	ch1.tRNA37
tRNA-Arg-TCT-1-2	chr1.trna34
tRNA-Ser-AGA-1-1	chr1.trna29
tRNA-Tyr-GTA-3-1	chr2.trna14
tRNA-Asn-GTT-1-2	chr3.trna22
tRNA-Ser-GCT-1-2	chr3.trna1
tRNA-Val-AAC-1-2	chr3.trna30
tRNA-Arg-ACG-1-2	chr4.trna15
tRNA-Glu-TTC-2-1	chr4.trna18
tRNA-Cys-GCA-1-2	chr4.trna1

Table 13: Boundary tDNAs in K. phaffii

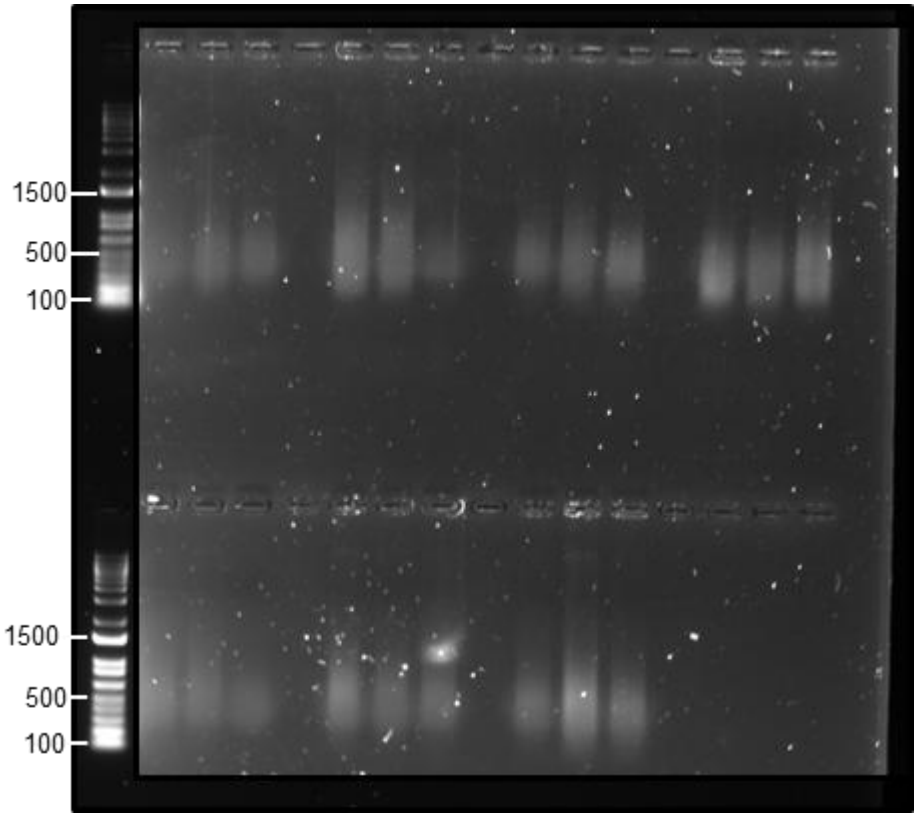


Figure 0.1. Chromatin sonicated to fragments ranging from 100bp to 500bp which is optimum for qPCR analysis.

References

1. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proceedings of the National Academy of Sciences. 1990;87(12):4576-9.
2. Crick FH, editor On protein synthesis. Symp Soc Exp Biol; 1958.
3. Ebrahimi SB, Samanta D. Engineering protein-based therapeutics through structural and chemical design. Nat Commun. 2023;14(1):2411.
4. Duman-Özdamar ZE, Binay B. Production of Industrial Enzymes via Pichia pastoris as a Cell Factory in Bioreactor: Current Status and Future Aspects. Protein J. 2021;40(3):367-76.
5. Mauro VP, Chappell SA. A critical analysis of codon optimization in human therapeutics. Trends Mol Med. 2014;20(11):604-13.
6. Vogl T, Gebbie L, Palfreyman RW, Speight R. Effect of Plasmid Design and Type of Integration Event on Recombinant Protein Expression in Pichia pastoris. Appl Environ Microbiol. 2018;84(6).
7. Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, et al. Protein production and purification. Nature Methods. 2008;5(2):135-46.
8. De Magistris P. The Great Escape: mRNA Export through the Nuclear Pore Complex. Int J Mol Sci. 2021;22(21).
9. Ramakrishnan V. Ribosome Structure and the Mechanism of Translation. Cell. 2002;108(4):557-72.
10. Ellgaard L, McCaul N, Chatsisvili A, Braakman I. Co- and Post-Translational Protein Folding in the ER. Traffic. 2016;17(6):615-38.
11. Kettel P, Karagöz GE. Endoplasmic reticulum: Monitoring and maintaining protein and membrane homeostasis in the endoplasmic reticulum by the unfolded protein response. The International Journal of Biochemistry & Cell Biology. 2024;172:106598.
12. Frappaolo A, Karimpour-Ghahnavieh A, Sechi S, Giansanti MG. The Close Relationship between the Golgi Trafficking Machinery and Protein Glycosylation. Cells. 2020;9(12).
13. Delic M, Valli M, Graf AB, Pfeffer M, Mattanovich D, Gasser B. The secretory pathway: exploring yeast diversity. FEMS Microbiology Reviews. 2013;37(6):872-914.
14. Cregg JM, Cereghino JL, Shi J, Higgins DR. Recombinant protein expression in Pichia pastoris. Mol Biotechnol. 2000;16(1):23-52.
15. Berkmen M. Production of disulfide-bonded proteins in Escherichia coli. Protein Expression and Purification. 2012;82(1):240-51.
16. Nevalainen H, Peterson R. Making recombinant proteins in filamentous fungi- are we expecting too much? Front Microbiol. 2014;5:75.
17. Damasceno LM, Huang CJ, Batt CA. Protein secretion in Pichia pastoris and advances in protein production. Appl Microbiol Biotechnol. 2012;93(1):31-9.
18. De Brabander P, Uitterhaegen E, Delmulle T, De Winter K, Soetaert W. Challenges and progress towards industrial recombinant protein production in yeasts: A review. Biotechnol Adv. 2023;64:108121.
19. Parthiban S, Vijeesh T, Gayathri T, Shanmugaraj B, Sharma A, Sathishkumar R. Artificial intelligence-driven systems engineering for next-generation plant-derived biopharmaceuticals. Front Plant Sci. 2023;14:1252166.
20. Nikolados E-M, Oyarzún DA. Deep learning for optimization of protein expression. Current Opinion in Biotechnology. 2023;81:102941.
21. Yuan H, Zhou Y, Lin Y, Tu R, Guo Y, Zhang Y, et al. Microfluidic screening and genomic mutation identification for enhancing cellulase production in Pichia pastoris. Biotechnology for Biofuels and Bioproducts. 2022;15(1):50.
22. Itakura K, Hirose T, Crea R, Riggs AD, Heyneker HL, Bolivar F, et al. Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. Science. 1977;198(4321):1056-63.

23. Baeshen NA, Baeshen MN, Sheikh A, Bora RS, Ahmed MMM, Ramadan HAI, et al. Cell factories for insulin production. Microbial Cell Factories. 2014;13(1):141.
24. Rezaei M, Zarkesh-Esfahani SH. Optimization of production of recombinant human growth hormone in Escherichia coli. J Res Med Sci. 2012;17(7):681-5.
25. Ouellette T, Destrau S, Ouellette T, Zhu J, Roach JM, Coffman JD, et al. Production and purification of refolded recombinant human IL-7 from inclusion bodies. Protein expression and purification. 2003;30(2):156-66.
26. Vallejo LF, Brokelmann M, Marten S, Trappe S, Cabrera-Crespo J, Hoffmann A, et al. Renaturation and purification of bone morphogenetic protein-2 produced as inclusion bodies in high-cell-density cultures of recombinant Escherichia coli. Journal of biotechnology. 2002;94(2):185-94.
27. Rosano GL, Ceccarelli EA. Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol. 2014;5:172.
28. Anyaogu DC, Mortensen UH. Manipulating the glycosylation pathway in bacterial and lower eukaryotes for production of therapeutic proteins. Curr Opin Biotechnol. 2015;36:122-8.
29. Diamant S, Ben-Zvi AP, Bukau B, Goloubinoff P. Size-dependent disaggregation of stable protein aggregates by the DnaK chaperone machinery. J Biol Chem. 2000;275(28):21107-13.
30. Dumont J, Euwart D, Mei B, Estes S, Kshirsagar R. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. Crit Rev Biotechnol. 2016;36(6):1110-22.
31. Durocher Y, Butler M. Expression systems for therapeutic glycoprotein production. Curr Opin Biotechnol. 2009;20(6):700-7.
32. Ghaderi D, Zhang M, Hurtado-Ziola N, Varki A. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. Biotechnol Genet Eng Rev. 2012;28:147-75.
33. Swiech K, Picanço-Castro V, Covas DT. Human cells: new platform for recombinant therapeutic protein production. Protein Expr Purif. 2012;84(1):147-53.
34. Summers MD. Milestones leading to the genetic engineering of baculoviruses as expression vector systems and viral pesticides. Adv Virus Res. 2006;68:3-73.
35. Nagels B, Weterings K, Callewaert N, Van Damme EJM. Production of Plant Made Pharmaceuticals: From Plant Host to Functional Protein. Critical Reviews in Plant Sciences. 2012;31(2):148-80.
36. Smith GE, Summers MD, Fraser MJ. Production of human beta interferon in insect cells infected with a baculovirus expression vector. Mol Cell Biol. 1983;3(12):2156-65.
37. Barta A, Sommergruber K, Thompson D, Hartmuth K, Matzke MA, Matzke AJ. The expression of a nopaline synthase - human growth hormone chimaeric gene in transformed tobacco and sunflower callus tissue. Plant Mol Biol. 1986;6(5):347-57.
38. Feng Z, Li X, Fan B, Zhu C, Chen Z. Maximizing the Production of Recombinant Proteins in Plants: From Transcription to Protein Stability. Int J Mol Sci. 2022;23(21).
39. Yee CM, Zak AJ, Hill BD, Wen F. The Coming Age of Insect Cells for Manufacturing and Development of Protein Therapeutics. Ind Eng Chem Res. 2018;57(31):10061-70.
40. Tan E, Chin CSH, Lim ZFS, Ng SK. HEK293 Cell Line as a Platform to Produce Recombinant Proteins and Viral Vectors. Front Bioeng Biotechnol. 2021;9:796991.
41. Rubino MJ. Experiences with HEK293: A Human Cell Line. PDA J Pharm Sci Technol. 2010;64(5):392-5.
42. Lin Y-C, Boone M, Meuris L, Lemmens I, Van Roy N, Soete A, et al. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. Nature Communications. 2014;5(1):4767.
43. Graham FL, Smiley J, Russell WC, Nairn R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. J Gen Virol. 1977;36(1):59-74.
44. Barone PW, Wiebe ME, Leung JC, Hussein ITM, Keumurian FJ, Bouressa J, et al. Viral contamination in biologic manufacture and implications for emerging therapies. Nat Biotechnol. 2020;38(5):563-72.

45. Kim SK, Jung I, Lee H, Kang K, Kim M, Jeong K, et al. Human histone H3K79 methyltransferase DOT1L protein binds actively transcribing RNA polymerase II to regulate gene expression. J Biol Chem. 2012;287:39698-709.
46. MacDonald MA, Nöbel M, Roche Recinos D, Martínez VS, Schulz BL, Howard CB, et al. Perfusion culture of Chinese Hamster Ovary cells for bioprocessing applications. Crit Rev Biotechnol. 2022;42(7):1099-115.
47. Jayapal KP, Wlaschin KF, Hu W, Yap MG. Recombinant protein therapeutics from CHO cells-20 years and counting. Chemical engineering progress. 2007;103(10):40.
48. Carver JH, Salazar EP, Knize MG. Chinese hamster ovary cells cultured in low concentrations of fetal bovine serum: cloning efficiency, growth in suspension, and selection of drug-resistant mutant phenotypes. In Vitro. 1983;19(9):699-706.
49. Onitsuka M, Kim WD, Ozaki H, Kawaguchi A, Honda K, Kajiura H, et al. Enhancement of sialylation on humanized IgG-like bispecific antibody by overexpression of α2,6-sialyltransferase derived from Chinese hamster ovary cells. Appl Microbiol Biotechnol. 2012;94(1):69-80.
50. Howard DR, Fukuda M, Fukuda MN, Stanley P. The GDP-fucose:N-acetylglucosaminide 3-alpha-L-fucosyltransferases of LEC11 and LEC12 Chinese hamster ovary mutants exhibit novel specificities for glycolipid substrates. J Biol Chem. 1987;262(35):16830-7.
51. Barnes LM, Bentley CM, Dickson AJ. Stability of protein production from recombinant mammalian cells. Biotechnol Bioeng. 2003;81(6):631-9.
52. Celik E, Calık P. Production of recombinant proteins by yeast cells. Biotechnol Adv. 2012;30(5):1108-18.
53. Duina AA, Miller ME, Keeney JB. Budding Yeast for Budding Geneticists: A Primer on the Saccharomyces cerevisiae Model System. Genetics. 2014;197(1):33-48.
54. Kulagina N, Besseau S, Godon C, Goldman GH, Papon N, Courdavault V. Yeasts as Biopharmaceutical Production Platforms. Front Fungal Biol. 2021;2:733492.
55. McAleer WJ, Buynak EB, Maigetter RZ, Wampler DE, Miller WJ, Hilleman MR. Human hepatitis B vaccine from recombinant yeast. Nature. 1984;307(5947):178-80.
56. Buckholz RG, Gleeson MA. Yeast systems for the commercial production of heterologous proteins. Biotechnology (N Y). 1991;9(11):1067-72.
57. Manfrão-Netto JHC, Gomes AMV, Parachin NS. Advances in Using Hansenula polymorpha as Chassis for Recombinant Protein Production. Front Bioeng Biotechnol. 2019;7:94.
58. Müller S, Sandal T, Kamp-Hansen P, Dalbøge H. Comparison of expression systems in the yeasts Saccharomyces cerevisiae, Hansenula polymorpha, Klyveromyces lactis, Schizosaccharomyces pombe and Yarrowia lipolytica. Cloning of two novel promoters from Yarrowia lipolytica. Yeast. 1998;14(14):1267-83.
59. Ahmad M, Hirz M, Pichler H, Schwab H. Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production. Appl Microbiol Biotechnol. 2014;98(12):5301-17.
60. Guilliermond A. Zygosaccharomyces pastori, nouvelle espèce de levures à copulation hétérogamique: Société mycologique de France; 1920 1920.
61. PHAFF HJ, MILLER MW, SHIFRINE M. The taxonomy of yeasts isolated from Drosophila in the Yosemite region of California. Antonie Van Leeuwenhoek. 1956;22(2):145-61.
62. Ogata K, Nishikawa H, Ohsugi M. A Yeast Capable of Utilizing Methanol. Agric Biol Chem. 1969;33(10):1519-20.
63. Yamada Y, Matsuda M, Maeda K, Mikata K. The Phylogenetic Relationships of Methanol-assimilating Yeasts Based on the Partial Sequences of 18S and 26S Ribosomal RNAs: The Proposal of Komagataella Gen. Nov. (Saccharomycetaceae). Bioscience, Biotechnology, and Biochemistry. 1995;59(3):439-44.
64. Love KR, Shah KA, Whittaker CA, Wu J, Bartlett MC, Ma D, et al. Comparative genomics and transcriptomics of Pichia pastoris. BMC Genomics. 2016;17:550.

65. Higgins DR, Cregg JM. Introduction to Pichia pastoris. Methods Mol Biol. 1998;103:1-15.
66. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 genes. Science. 1996;274:562-7.
67. Bernauer L, Radkohl A, Lehmayer LGK, Emmerstorfer-Augustin A. Komagataella phaffii as Emerging Model Organism in Fundamental Research. Front Microbiol. 2020;11:607028.
68. De Schutter K, Lin Y-C, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, et al. Genome sequence of the recombinant protein production host Pichia pastoris. Nature Biotechnology. 2009;27(6):561-6.
69. Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, et al. The genome sequence of Schizosaccharomyces pombe. Nature. 2002;415(6874):871-80.
70. Coughlan AY, Hanson SJ, Byrne KP, Wolfe KH. Centromeres of the Yeast Komagataella phaffii (Pichia pastoris) Have a Simple Inverted-Repeat Structure. Genome Biol Evol. 2016;8(8):2482-92.
71. Fitzgerald-Hayes M, Clarke L, Carbon J. Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs. Cell. 1982;29(1):235-44.
72. Logsdon GA, Rozanski AN, Ryabov F, Potapova T, Shepelev VA, Catacchio CR, et al. The variation and evolution of complete human centromeres. Nature. 2024;629(8010):136-45.
73. Rossanese OW, Soderholm J, Bevis BJ, Sears IB, O'Connor J, Williamson EK, et al. Golgi structure correlates with transitional endoplasmic reticulum organization in Pichia pastoris and Saccharomyces cerevisiae. J Cell Biol. 1999;145(1):69-81.
74. Gould SJ, McCollum D, Spong AP, Heyman JA, Subramani S. Development of the yeast Pichia pastoris as a model organism for a genetic and molecular analysis of peroxisome assembly. Yeast. 1992;8(8):613-28.
75. Preuss D, Mulholland J, Franzusoff A, Segev N, Botstein D. Characterization of the Saccharomyces Golgi complex through the cell cycle by immunoelectron microscopy. Mol Biol Cell. 1992;3(7):789-803.
76. Glick BS. Cell biology: alternatives to baker's yeast. Curr Biol. 1996;6(12):1570-2.
77. Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, et al. The fungi. Curr Biol. 2009;19(18):R840-5.
78. Ata Ö, Ergün BG, Fickers P, Heistinger L, Mattanovich D, Rebnegger C, et al. What makes Komagataella phaffii non-conventional? FEMS Yeast Research. 2021;21(8):foab059.
79. Waterham HR, Digan ME, Koutz PJ, Lair SV, Cregg JM. Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. Gene. 1997;186(1):37-44.
80. Gellissen G. Heterologous protein production in methylotrophic yeasts. Appl Microbiol Biotechnol. 2000;54(6):741-50.
81. Cos O, Ramón R, Montesinos JL, Valero F. Operational strategies, monitoring and control of heterologous protein production in the methylotrophic yeast Pichia pastoris under different promoters: A review. Microbial Cell Factories. 2006;5(1):17.
82. Hilt W, Wolf DH. Stress-induced proteolysis in yeast. Mol Microbiol. 1992;6(17):2437-42.
83. Karbalaei M, Rezaee SA, Farsiani H. Pichia pastoris: A highly successful expression system for optimal synthesis of heterologous proteins. J Cell Physiol. 2020;235(9):5867-81.
84. Shen W, Xue Y, Liu Y, Kong C, Wang X, Huang M, et al. A novel methanol-free Pichia pastoris system for recombinant protein expression. Microbial Cell Factories. 2016;15(1):178.
85. Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA. Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. Mol Cell Biol. 1989;9(3):1316-23.
86. Cregg JM, Tschopp JF, Stillman C, Siegel R, Akong M, Craig WS, et al. High–Level Expression and Efficient Assembly of Hepatitis B Surface Antigen in the Methylotrophic Yeast, Pichia Pastoris. Bio/Technology. 1987;5(5):479-85.

87. Orman MA, Calik P, Ozdamar TH. The influence of carbon sources on recombinant-human- growth-hormone production by Pichia pastoris is dependent on phenotype: a comparison of Muts and Mut+ strains. Biotechnol Appl Biochem. 2009;52(Pt 3):245-55.
88. Näätsaari L, Mistlberger B, Ruth C, Hajek T, Hartner FS, Glieder A. Deletion of the Pichia pastoris KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One. 2012;7(6):e39720.
89. Velastegui E, Quezada J, Guerrero K, Altamirano C, Martinez JA, Berrios J, et al. Is heterogeneity in large-scale bioreactors a real problem in recombinant protein synthesis by Pichia pastoris? Appl Microbiol Biotechnol. 2023;107(7-8):2223-33.
90. Charoenrat T, Khumruaengsri N, Promdonkoy P, Rattanaphan N, Eurwilaichitr L, Tanapongpipat S, et al. Improvement of recombinant endoglucanase produced in Pichia pastoris KM71 through the use of synthetic medium for inoculum and pH control of proteolysis. J Biosci Bioeng. 2013;116(2):193-8.
91. Van Landuyt L, Lonigro C, Meuris L, Callewaert N. Customized protein glycosylation to improve biopharmaceutical function and targeting. Current Opinion in Biotechnology. 2019;60:17-28.
92. Jacobs PP, Geysens S, Vervecken W, Contreras R, Callewaert N. Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology. Nature protocols. 2009;4(1):58-70.
93. Hamilton SR, Gerngross TU. Glycosylation engineering in yeast: the advent of fully humanized yeast. Curr Opin Biotechnol. 2007;18(5):387-92.
94. De Wachter C, Van Landuyt L, Callewaert N. Engineering of Yeast Glycoprotein Expression. Adv Biochem Eng Biotechnol. 2021;175:93-135.
95. Luong M, Lam JS, Chen J, Levitz SM. Effects of fungal N- and O-linked mannosylation on the immunogenicity of model vaccines. Vaccine. 2007;25(22):4340-4.
96. Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. Nat Rev Nephrol. 2019;15(6):346-66.
97. Gerngross TU. Advances in the production of human therapeutic proteins in yeasts and filamentous fungi. Nat Biotechnol. 2004;22(11):1409-14.
98. Krainer FW, Gmeiner C, Neutsch L, Windwarder M, Pletzenauer R, Herwig C, et al. Knockout of an endogenous mannosyltransferase increases the homogeneity of glycoproteins produced in Pichia pastoris. Sci Rep. 2013;3:3279.
99. Owji H, Nezafat N, Negahdaripour M, Hajiebrahimi A, Ghasemi Y. A comprehensive review of signal peptides: Structure, roles, and applications. Eur J Cell Biol. 2018;97(6):422-41.
100. Zou C, Lu L, Wang S, Zhang C, Chen X, Lin Y, et al. The α-mating factor secretion signals and endogenous signal peptides for recombinant protein secretion in Komagataella phaffii. Biotechnology for Biofuels and Bioproducts. 2022;15(1):140.
101. Duan G, Ding L, Wei D, Zhou H, Chu J, Zhang S, et al. Screening endogenous signal peptides and protein folding factors to promote the secretory expression of heterologous proteins in Pichia pastoris. Journal of biotechnology. 2019;306:193-202.
102. Liang S, Li C, Ye Y, Lin Y. Endogenous signal peptides efficiently mediate the secretion of recombinant proteins in Pichia pastoris. Biotechnology letters. 2013;35:97-105.
103. García-Ortega X, Cámara E, Ferrer P, Albiol J, Montesinos-Seguí JL, Valero F. Rational development of bioprocess engineering strategies for recombinant protein production in Pichia pastoris (Komagataella phaffii) using the methanol-free GAP promoter. Where do we stand? New biotechnology. 2019;53:24-34.
104. Gasset A, Garcia-Ortega X, Garrigós-Martínez J, Valero F, Montesinos-Seguí JL. Innovative bioprocess strategies combining physiological control and strain engineering of Pichia pastoris to improve recombinant protein production. Frontiers in bioengineering and biotechnology. 2022;10:818434.
105. Love KR, Politano TJ, Panagiotou V, Jiang B, Stadheim TA, Love JC. Systematic single-cell analysis of Pichia pastoris reveals secretory capacity limits productivity. PLoS One. 2012;7(6):e37915.

106. Berg MD, Brandl CJ. Transfer RNAs: diversity in form and function. RNA Biol. 2021;18(3):316-39.
107. Donze D, Kamakaka RT. RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in Saccharomyces cerevisiae. EMBO J. 2001;20(3):520-31.
108. Hanson SJ, Byrne KP, Wolfe KH. Mating-type switching by chromosomal inversion in methylotrophic yeasts suggests an origin for the three-locus Saccharomyces cerevisiae system. Proc Natl Acad Sci U S A. 2014;111(45):E4851-8.
109. Hicks JB, Strathern JN, Herskowitz I. The cassette model of mating-type interconversion [in the yeast Saccharomyces cerevisiae]. DNA Insertions Meeting Cold Spring Harbor, NY (USA) 1976. 1977.
110. Loo S, Rine J. Silencers and domains of generalized repression. Science. 1994;264(5166):1768-71.
111. Rine J, Strathern JN, Hicks JB, Herskowitz I. A suppressor of mating-type locus mutations in Saccharomyces cerevisiae: evidence for and identification of cryptic mating-type loci. Genetics. 1979;93(4):877-901.
112. Van Bortle K, Corces VG. tDNA insulators and the emerging role of TFIIIC in genome organization. Transcription. 2012;3(6):277-84.
113. Kirkland JG, Raab JR, Kamakaka RT. TFIIIC bound DNA elements in nuclear organization and insulation. Biochim Biophys Acta. 2013;1829:418-24.
114. Sizer RE, Chahid N, Butterfield SP, Donze D, Bryant NJ, White RJ. TFIIIC-based chromatin insulators through eukaryotic evolution. Gene. 2022;835:146533.
115. Sizer RE, Ingram RM, Swan C, Biggs EK, Pybus LP, White RJ. Use of tRNA gene barriers improves stability of transgene expression in CHO cells. Biotechnol J. 2024;19(8):e2400196.
116. Raab JR, Chiu J, Zhu J, Katzman S, Kurukuti S, Wade PA, et al. Human tRNA genes function as chromatin insulators. EMBO J. 2012;31:330-50.
117. Tegel H, Tourle S, Ottosson J, Persson A. Increased levels of recombinant human proteins with the Escherichia coli strain Rosetta(DE3). Protein Expr Purif. 2010;69(2):159-67.
118. Crick FHC. Codon-anticodon pairing: the wobble hypothesis. J Mol Biol. 1966;19:548-55.
119. Chan CT, Pang YL, Deng W, Babu IR, Dyavaiah M, Begley TJ, et al. Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nat Commun. 2012;3:937.
120. Angov E, Hillier CJ, Kincaid RL, Lyon JA. Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host. PLoS One. 2008;3(5):e2189.
121. Guy O, Lombardo D, Bartelt DC, Amic J, Figarella C. Two human trypsinogens. Purification, molecular properties, and N-terminal sequences. Biochemistry. 1978;17(9):1669-75.
122. Walsh KA, Kauffman DL, Kumar KS, Neurath H. On the structure and function of bovine trypsinogen and trypsin. Proceedings of the National Academy of Sciences. 1964;51(2):301-8.
123. MALLUCCI L, WELLS V, Young M. Effect of trypsin on cell volume and mass. Nature New Biology. 1972;239(89):53-5.
124. Huang H-L, Hsing H-W, Lai T-C, Chen Y-W, Lee T-R, Chan H-T, et al. Trypsin-induced proteome alteration during cell subculture in mammalian cells. Journal of biomedical science. 2010;17:1-10.
125. Kono T, Barham FW. The relationship between the insulin-binding capacity of fat cells and the cellular response to insulin: studies with intact and trypsin-treated fat cells. Journal of Biological Chemistry. 1971;246(20):6210-6.
126. Schilling RJ, Mitra AK. Degradation of insulin by trypsin and alpha-chymotrypsin. Pharmaceutical research. 1991;8:721-7.

127. Hanquier J, Sorlet Y, Desplancq D, Baroche L, Ebtinger M, Lefèvre JF, et al. A single mutation in the activation site of bovine trypsinogen enhances its accumulation in the fermentation broth of the yeast Pichia pastoris. Appl Environ Microbiol. 2003;69(2):1108-13.
128. Kossiakoff AA, Chambers JL, Kay LM, Stroud RM. Structure of bovine trypsinogen at 1.9 A resolution. Biochemistry. 1977;16(4):654-64.
129. Vestling MM, Murphy CM, Fenselau C. Recognition of trypsin autolysis products by high-performance liquid chromatography and mass spectrometry. Analytical chemistry. 1990;62(21):2391-4.
130. Wainfan E, Hess GP. Enzymatically active products of trypsin autolysis. Journal of the American Chemical Society. 1960;82(8):2069-73.
131. Liu H, May K. Disulfide bond structures of IgG molecules. mAbs. 2012;4(1):17-23.
132. Nyaruhucha CN, Kito M, Fukuoka SI. Identification and expression of the cDNA-encoding human mesotrypsin(ogen), an isoform of trypsin with inhibitor resistance. J Biol Chem. 1997;272(16):10573-8.
133. Betancur MO, Reis VCB, Nicola AM, De Marco JL, de Moraes LMP, Torres FAG. Multicopy plasmid integration in Komagataella phaffii mediated by a defective auxotrophic marker. Microbial Cell Factories. 2017;16(1):99.
134. Lis H, Sharon N. Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition. Chemical Reviews. 1998;98(2):637-74.
135. Nagae M, Soga K, Morita-Matsumoto K, Hanashima S, Ikeda A, Yamamoto K, et al. Phytohemagglutinin from Phaseolus vulgaris (PHA-E) displays a novel glycan recognition mode using a common legume lectin fold. Glycobiology. 2014;24(4):368-78.
136. Gerfen CR, Sawchenko PE. An anterograde neuroanatomical tracing method that shows the detailed morphology of neurons, their axons and terminals: immunohistochemical localization of an axonally transported plant lectin, Phaseolus vulgaris leucoagglutinin (PHA-L). Brain Res. 1984;290(2):219-38.
137. Wang P, Hu J, Min S, Chen C, Zhu Y, Pan Y, et al. Recombinant Phaseolus vulgaris phytohemagglutinin L-form expressed in the Bacillus brevis exerts in vitro and in vivo anti-tumor activity through potentiation of apoptosis and immunomodulation. International Immunopharmacology. 2023;120:110322.
138. Dao-Thi MH, Hamelryck TW, Poortmans F, Voelker TA, Chrispeels MJ, Wyns L. Crystallization of glycosylated and nonglycosylated phytohemagglutinin-L. Proteins: structure, function, and bioinformatics. 1996;24(1):134-7.
139. Hoffman LM, Donaldson DD. Synthesis of Mitogenic Phytohemagglutinin–L in Escherichia Coli. Bio/Technology. 1987;5(2):157-60.
140. Zhu K, Huesing JE, Shade RE, Bressan RA, Hasegawa PM, Murdock LL. An insecticidal N-acetylglucosamine-specific lectin gene from Griffonia simplicifolia (Leguminosae). Plant physiology. 1996;110(1):195-202.
141. Rodriguez-Arango E, Arango R, Adar R, Galili G, Sharon N. Cloning, sequence analysis and expression in Escherichia coli of the cDNA encoding a precursor of peanut agglutinin. FEBS letters. 1992;307(2):185-9.
142. Tague BW, Chrispeels MJ. The plant vacuolar protein, phytohemagglutinin, is transported to the vacuole of transgenic yeast. The Journal of cell biology. 1987;105(5):1971-9.
143. Raemaekers RJM, de Muro L, Gatehouse JA, Fordham-Skelton AP. Functional phytohemagglutinin (PHA) and Galanthus nivalis agglutinin (GNA) expressed in Pichia pastoris. European Journal of Biochemistry. 1999;265(1):394-403.
144. Wong C, Sridhara S, Bardwell JCA, Jakob U. Heating Greatly Speeds Coomassie Blue Staining and Destaining. BioTechniques. 2000;28(3):426-32.
145. Lerdrup M, Johansen JV, Agrawal-Singh S, Hansen K. An interactive environment for agile analysis and visualization of ChIP-sequencing data. Nat Struct Mol Biol. 2016;23(4):349-57.
146. Xi L, Fondufe-Mittendorf Y, Xia L, Flatow J, Widom J, Wang JP. Predicting nucleosome positioning using a duration Hidden Markov Model. BMC Bioinformatics. 2010;11:346.

147. O'Kane CJ, Hyland EM. Yeast epigenetics: the inheritance of histone modification states. Biosci Rep. 2019;39(5).
148. Chusainow J, Yang YS, Yeo JH, Toh PC, Asvadi P, Wong NS, et al. A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? Biotechnol Bioeng. 2009;102(4):1182-96.
149. Bailey LA, Hatton D, Field R, Dickson AJ. Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. Biotechnol Bioeng. 2012;109:2093-103.
150. Kim M, O'Callaghan PM, Droms KA, James DC. A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnol Bioeng. 2011;108(10):2434-46.
151. Hernandez I, Dhiman H, Klanert G, Jadhav V, Auer N, Hanscho M, et al. Epigenetic regulation of gene expression in Chinese Hamster Ovary cells in response to the changing environment of a batch culture. Biotechnol Bioeng. 2019;116(3):677-92.
152. Noma, Allis CD, Grewal SI. Transitions in distinct histone H3 methylation patterns at the heterochromatin domain boundaries. Science. 2001;293(5532):1150-5.
153. Haldar D, Kamakaka RT. tRNA genes as chromatin barriers. Nature Structural and Mol Biol. 2006;13:192-3.
154. Hampsey M. Molecular genetics of the RNA polymerase II general transcriptional machinery. Microbiology and Molecular Biology Reviews. 1998;62(2):465-503.
155. Todeschini A-L, Georges A, Veitia RA. Transcription factors: specific DNA binding and specific gene regulation. Trends in Genetics. 2014;30(6):211-9.
156. Allfrey VG, Faulkner R, Mirsky AE. ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. Proc Natl Acad Sci U S A. 1964;51:786-94.
157. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Biol. 2018;19(4):229-44.
158. Berger SL. The histone modification circus. Science. 2001;292:64-5.
159. Burlingame RW, Love WE, Wang B-C, Hamlin R, Xuong N-H, Moudrianakis EN. Crystallographic structure of the octameric histone core of the nucleosome at a resolution of 3.3A. 1985;228:546-53.
160. Allshire RC, Ekwall K. Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe. Cold Spring Harb Perspect Biol. 2015;7(7):a018770.
161. Hansen JC, Wolffe AP. Influence of chromatin folding on transcription initiation and elongation by RNA polymerase III. Biochemistry. 1992;31(34):7977-88.
162. Morse RH, Roth SY, Simpson RT. A transcriptionally active tRNA gene interferes with nucleosome positioning *in vivo*. Mol Cell Biol. 1992;12:4015-25.
163. Lamond AI, Earnshaw WC. Structure and function in the nucleus. Science. 1998;280(5363):547-53.
164. Allan J, Harborne N, Rau DC, Gould H. Participation of core histone "tails" in the stabilization of the chromatin solenoid. J Cell Biol. 1982;93(2):285-97.
165. Cairns BR, Lorch Y, Li Y, Zhang M, Lacomis L, Erdjument-Bromage H, et al. RSC, an Essential, Abundant Chromatin-Remodeling Complex. Cell. 1996;87(7):1249-60.
166. Oki M, Valenzuela L, Chiba T, Ito T, Kamakaka RT. Barrier proteins remodel and modify chromatin to restrict silenced domains. Mol Cell Biol. 2004;24(5):1956-67.
167. Eissenberg JC, James TC, Foster-Hartnett DM, Hartnett T, Ngan V, Elgin SC. Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. Proc Natl Acad Sci U S A. 1990;87(24):9923-7.
168. Bannister AJ, Zegerman P, Partridge JF, Miska EA, Thomas JO, Allshire RC, et al. Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature. 2001;410(6824):120-4.
169. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature. 2001;410(6824):116-20.

170. Bachman KE, Park BH, Rhee I, Rajagopalan H, Herman JG, Baylin SB, et al. Histone modifications and silencing prior to DNA methylation of a tumour suppressor gene. Cancer Cell. 2003;3:89-95.
171. Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications. Biochim Biophys Acta. 2014;1839(8):627-43.
172. Saha A, Wittmeyer J, Cairns BR. Chromatin remodelling: the industrial revolution of DNA around histones. Nat Rev Mol Cell Biol. 2006;7(6):437-47.
173. Yamada T, Fischle W, Sugiyama T, Allis CD, Grewal SI. The nucleation and maintenance of heterochromatin by a histone deacetylase in fission yeast. Mol Cell. 2005;20(2):173-85.
174. Lombardi PM, Cole KE, Dowling DP, Christianson DW. Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes. Curr Opin Struct Biol. 2011;21(6):735-43.
175. Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. Genes Dev. 1999;13(15):1924-35.
176. Carey M, Li B, Workman JL. RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. Molecular cell. 2006;24(3):481-7.
177. De S, Rebnegger C, Moser J, Tatto N, Graf AB, Mattanovich D, et al. Pseudohyphal differentiation in Komagataella phaffii: investigating the FLO gene family. FEMS Yeast Res. 2020;20(5).
178. Martire S, Banaszynski LA. The roles of histone variants in fine-tuning chromatin organization and function. Nat Rev Mol Cell Biol. 2020;21(9):522-41.
179. Wang J, Lunyak VV, Jordan IK. Genome-wide prediction and analysis of human chromatin boundary elements. Nucleic Acids Res. 2012;40(2):511-29.
180. Jia S, Noma K, Grewal SI. RNAi-independent heterochromatin nucleation by the stress-activated ATF/CREB family proteins. Science. 2004;304(5679):1971-6.
181. Shukla A, Bhargava P. Regulation of tRNA gene transcription by the chromatin structure and nucleosome dynamics. BBA - Gene Regul Mech. 2018;1861:295-309.
182. Hall IM, Shankaranarayana GD, Noma K, Ayoub N, Cohen A, Grewal SI. Establishment and maintenance of a heterochromatin domain. Science. 2002;297(5590):2232-7.
183. Dubey RN, Gartenberg MR. A tDNA establishes cohesion of a neighboring silent chromatin domain. Genes Dev. 2007;21(17):2150-60.
184. Pidoux AL, Allshire RC. The role of heterochromatin in centromere function. Philosophical Transactions of the Royal Society B: Biological Sciences. 2005;360(1455):569-79.
185. Bierhoff H, Postepska-Igielska A, Grummt I. Noisy silence: non-coding RNA and heterochromatin formation at repetitive elements. Epigenetics. 2014;9(1):53-61.
186. Nishibuchi G, Déjardin J. The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. Chromosome Research. 2017;25(1):77-87.
187. Gilbert N, Allan J. Distinctive higher-order chromatin structure at mammalian centromeres. Proc Natl Acad Sci USA. 2001;98:11949-54.
188. Ishii K, Arib G, Lin C, Van Houwe G, Laemmli UK. Chromatin boundaries in budding yeast: the nuclear pore connection. Cell. 2002;109(5):551-62.
189. Stutzman AV, Liang AS, Beilinson V, Ikegami K. Transcription-independent TFIIIC-bound sites cluster near heterochromatin boundaries within lamina-associated domains in C. elegans. Epigenetics & Chromatin. 2020;13(1):1.
190. Tartof KD. Position effect variegation in yeast. BioEssays. 1994;16(10):713-4.
191. Hicks J, Strathern JN, Klar AJ. Transposable mating type genes in Saccharomyces cerevisiae. Nature. 1979;282(5738):478-3.
192. Strathern JN, Klar AJS, Hicks JB, Abraham JA, Ivy JM, Nasmyth KA, et al. Homothallic switching of yeast mating type cassettes is initiated by a double-stranded cut in the MAT locus. Cell. 1982;31(1):183-92.

193. Biswas M, Maqani N, Rai R, Kumaran SP, Iyer KR, Sendinc E, et al. Limiting the extent of the RDN1 heterochromatin domain by a silencing barrier and Sir2 protein levels in Saccharomyces cerevisiae. Mol Cell Biol. 2009;29(10):2889-98.
194. Blander G, Guarente L. The Sir2 family of protein deacetylases. Annu Rev Biochem. 2004;73:417-35.
195. Hecht A, Laroche T, Strahl-Bolsinger S, Gasser SM, Grunstein M. Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. Cell. 1995;80(4):583-92.
196. Donze D, Adams CR, Rine J, Kamakaka RT. The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*. Genes Dev. 1999;13:698-708.
197. Sharp SJ, Schaack J, Cooley L, Burke DJ, Soll D. Structure and transcription of eukaryotic tRNA genes. CRC Crit Rev Biochem. 1984;19:107-44.
198. Schramm L, Hernandez N. Recruitment of RNA polymerase III to its target promoters. Genes Dev. 2002;16:2593-620.
199. Ramsay EP, Vannini A. Structural rearrangements of the RNA polymerase III machinery during tRNA transcription initiation. Biochim Biophys Acta Gene Regul Mech. 2018;1861:285-94.
200. Baker RE, Gabrielsen O, Hall BD. Effects of $tRNA^{Tyr}$ point mutations on the binding of yeast RNA polymerase III transcription factor C. J Biol Chem. 1986;261:5275-82.
201. Dhillon N, Raab JR, Guzzo SJ, Szyjka S, Gangadharan S, Aparicio OM, et al. DNA polymerase epsilon, acetylases and remodellers cooperate to form a specialized chromatin structure at a tRNA insulator. EMBO J. 2009;28:2583-600.
202. Jansen A, Verstrepen KJ. Nucleosome positioning in Saccharomyces cerevisiae. Microbiol Mol Biol Rev. 2011;75(2):301-20.
203. Harismendy O, Gendrel C-G, Soularue P, Gidrol X, Sentenac A, Werner M, et al. Genome-wide location of yeast RNA polymerase III transcription machinery. EMBO J. 2003;22:4738-47.
204. Roberts DN, Stewart AJ, Huff JT, Cairns BR. The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. Proc Natl Acad Sci USA. 2003;100:14695-700.
205. Moqtaderi Z, Struhl K. Genome-wide occupancy profile of the RNA polymerase III machinery in Saccharomyces cerevisiae reveals loci with incomplete transcription complexes. Mol Cell Biol. 2004;24(10):4118-27.
206. Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, et al. Genome-scale identification of nucleosome positions in S. cerevisiae. Science. 2005;309(5734):626-30.
207. Dion MF, Kaplan T, Kim M, Buratowski S, Friedman N, Rando OJ. Dynamics of replication-independent histone turnover in budding yeast. Science. 2007;315:1405-8.
208. Nagarajavel V, Iben JR, Howard BH, Maraia RJ, Clark DJ. Global 'bootprinting' reveals the elastic architecture of the yeast TFIIIB-TFIIIC transcription complex in vivo. Nucleic Acids Res. 2013;41(17):8135-43.
209. Hamdani O, Dhillon N, Hsieh TS, Fujita T, Ocampo J, Kirkland JG, et al. tRNA Genes Affect Chromosome Structure and Function via Local Effects. Mol Cell Biol. 2019;39(8).
210. Ng HH, Robert F, Young RA, Struhl K. Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex. Genes Dev. 2002;16:806-19.
211. Mahapatra S, Dewari PS, Bhardwaj A, Bhargava P. Yeast H2A.Z, FACT complex and RSC regulate transcription of tRNA gene through differential dynamics of flanking nucleosomes. Nucleic Acids Res. 2011;39:4023-34.
212. Kumar Y, Bhargava P. A unique nucleosome arrangement, maintained actively by chromatin remodelers facilitates transcription of yeast tRNA genes. BMC Genomics. 2013;14:402.
213. Jambunathan N, Martinez AW, Robert EC, Agochukwu NB, Ibos ME, Dugas SL, et al. Multiple bromodomain genes are involved in restricting the spread of heterochromatic silencing at the Saccharomyces cerevisiae HMR-tRNA boundary. Genetics. 2005;171(3):913-22.

214. Simms TA, Dugas SL, Gremillion JC, Ibos ME, Dandurand MN, Toliver TT, et al. TFIIIC binding sites function as both heterochromatin barriers and chromatin insulators in Saccharomyces cerevisiae. Eukaryot Cell. 2008;7(12):2078-86.
215. Haeusler RA, Pratt-Hyatt M, Good PD, Gipson TA, Engelke DR. Clustering of yeast tRNA genes is mediated by specific association of condensin with tRNA gene transcription complexes. Genes Dev. 2008;22:2204-14.
216. Thompson M, Haeusler RA, Good PD, Engelke DR. Nucleolar clustering of dispersed tRNA genes. Science. 2003;302:1399-401.
217. Noma K-I, Cam HP, Maraia RJ, Grewal SIS. A role for TFIIIC transcription factor complex in genome organization. Cell. 2006;125(5):859-72.
218. Iwasaki O, Tanaka A, Tanizawa H, Grewal SI, Noma K. Centromeric localization of dispersed Pol III genes in fission yeast. Mol Biol Cell. 2010;21(2):254-65.
219. Hiraga S-I, Botsios S, Donze D, Donaldson AD. TFIIIC localizes budding yeast ETC sites to the nuclear periphery. Mol Biol Cell. 2012;23(14):2741-54.
220. Bachleitner S, Severinsen MM, Lutz G, Mattanovich D. Overexpression of the transcriptional activators Mxr1 and Mit1 enhances lactic acid production on methanol in Komagataella phaffii. Metabolic Engineering. 2024;85:133-44.
221. Chang C-H, Hsiung H-A, Hong K-L, Huang C-T. Enhancing the efficiency of the Pichia pastoris AOX1 promoter via the synthetic positive feedback circuit of transcription factor Mxr1. BMC Biotechnol. 2018;18(1):81.
222. Wang X, Wang Q, Wang J, Bai P, Shi L, Shen W, et al. Mit1 Transcription Factor Mediates Methanol Signaling and Regulates the Alcohol Oxidase 1 (AOX1) Promoter in Pichia pastoris*. Journal of Biological Chemistry. 2016;291(12):6245-61.
223. Wang X, Cai M, Shi L, Wang Q, Zhu J, Wang J, et al. PpNrg1 is a transcriptional repressor for glucose and glycerol repression of AOX1 promoter in methylotrophic yeast Pichia pastoris. Biotechnol Lett. 2016;38(2):291-8.
224. Shi L, Wang X, Wang J, Zhang P, Qi F, Cai M, et al. Transcriptome analysis of Δ mig1 Δ mig2 mutant reveals their roles in methanol catabolism, peroxisome biogenesis and autophagy in methylotrophic yeast Pichia pastoris. Genes & genomics. 2018;40:399-412.
225. Polupanov AS, Nazarko VY, Sibirny AA. Gss1 protein of the methylotrophic yeast Pichia pastoris is involved in glucose sensing, pexophagy and catabolite repression. The international journal of biochemistry & cell biology. 2012;44(11):1906-18.
226. Zhang P, Zhang W, Zhou X, Bai P, Cregg JM, Zhang Y. Catabolite repression of Aox in Pichia pastoris is dependent on hexose transporter PpHxt1 and pexophagy. Applied and environmental microbiology. 2010;76(18):6108-18.
227. Tschopp JF, Brust PF, Cregg JM, Stillman CA, Gingeras TR. Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris. Nucleic Acids Res. 1987;15(9):3859-76.
228. Yang J, Cai H, Liu J, Zeng M, Chen J, Cheng Q, et al. Controlling AOX1 promoter strength in Pichia pastoris by manipulating poly (dA:dT) tracts. Sci Rep. 2018;8(1):1401.
229. Struhl K. Naturally occurring poly (dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. Proceedings of the National Academy of Sciences. 1985;82(24):8419-23.
230. Koch KA, Thiele DJ. Functional analysis of a homopolymeric (dA-dT) element that provides nucleosomal access to yeast and mammalian transcription factors. Journal of Biological Chemistry. 1999;274(34):23752-60.
231. Teves SS, Weber CM, Henikoff S. Transcribing through the nucleosome. Trends in biochemical sciences. 2014;39(12):577-86.
232. Park PJ. ChIP–seq: advantages and challenges of a maturing technology. Nature reviews genetics. 2009;10(10):669-80.
233. Kamman J, Tatini S. Optimal conditions for assay of staphylococcal nuclease. Journal of Food Science. 1977;42(2):421-4.
234. Chereji RV, Ocampo J, Clark DJ. MNase-Sensitive Complexes in Yeast: Nucleosomes and Non-histone Barriers. Molecular Cell. 2017;65(3):565-77.e3.

235. Fukushima Y, Hatazawa S, Hirai S, Kujirai T, Ehara H, Sekine S-I, et al. Structural and biochemical analyses of the nucleosome containing Komagataella pastoris histones. J Biochem. 2022.
236. Brady JR, Tan MC, Whittaker CA, Colant NA, Dalvie NC, Love KR, et al. Identifying Improved Sites for Heterologous Gene Integration Using ATAC-seq. ACS Synth Biol. 2020;9(9):2515-24.
237. Tsompana M, Buck MJ. Chromatin accessibility: a window into the genome. Epigenetics Chromatin. 2014;7(1):33.
238. Marck C, Kachouri-Lafond R, Lafontaine I, Westhof E, Dujon B, Grosjean H. The RNA polymerase III-dependent family of genes in hemiascomycetes: comparative RNomics, decoding strategies, transcription and evolutionary implications. Nucleic Acids Res. 2006;34(6):1816-35.
239. Oki M, Kamakaka RT. Barrier Function at HMR. Molecular Cell. 2005;19(5):707-16.
240. Hamdini O, Dhillon N, Hsieh T-HS, Fujita T, Ocampo J, Kirkland JG, et al. tRNA genes affect chromosome structure and function via local effects. Mol Cell Biol. 2019;39:e00432-18.
241. Brickner DG, Cajigas I, Fondufe-Mittendorf Y, Ahmed S, Lee P-C, Widom J, et al. H2A.Z-Mediated Localization of Genes at the Nuclear Periphery Confers Epigenetic Memory of Previous Transcriptional State. PLoS Biology. 2007;5(4):e81.
242. Sood V, Cajigas I, D'Urso A, Light WH, Brickner JH. Epigenetic Transcriptional Memory of GAL Genes Depends on Growth in Glucose and the Tup1 Transcription Factor in Saccharomyces cerevisiae. Genetics. 2017;206(4):1895-907.
243. Light WH, Brickner DG, Brand VR, Brickner JH. Interaction of a DNA Zip Code with the Nuclear Pore Complex Promotes H2A.Z Incorporation and INO1 Transcriptional Memory. Molecular Cell. 2010;40(1):112-25.
244. Chang C-R, Wu C-S, Hom Y, Gartenberg MR. Targeting of cohesin by transcriptionally silent chromatin. Genes & development. 2005;19(24):3031-42.
245. Glynn EF, Megee PC, Yu HG, Mistrot C, Unal E, Koshland DE, et al. Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae. PLoS Biol. 2004;2(9):E259.
246. Filipski J, Mucha M. Structure, function and DNA composition of Saccharomyces cerevisiae chromatin loops. Gene. 2002;300(1):63-8.
247. Veith N, Ziehr H, MacLeod RAF, Reamon-Buettner SM. Mechanisms underlying epigenetic and transcriptional heterogeneity in Chinese hamster ovary (CHO) cell lines. BMC Biotechnology. 2016;16(1):6.
248. Gerasimova TI, Gdula DA, Gerasimov DV, Simonova O, Corces VG. A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. Cell. 1995;82(4):587-97.
249. Kundu S, Horn PJ, Peterson CL. SWI/SNF is required for transcriptional memory at the yeast GAL gene cluster. Genes Dev. 2007;21(8):997-1004.
250. Vihervaara A, Mahat DB, Himanen SV, Blom MAH, Lis JT, Sistonen L. Stress-induced transcriptional memory accelerates promoter-proximal pause release and decelerates termination over mitotic divisions. Mol Cell. 2021;81(8):1715-31.e6.
251. Acar M, Becskei A, van Oudenaarden A. Enhancement of cellular memory by reducing stochastic transitions. Nature. 2005;435(7039):228-32.
252. Sump B, Brickner J. Establishment and inheritance of epigenetic transcriptional memory. Frontiers in Molecular Biosciences. 2022;9:977653.
253. Hull MW, Erickson J, Johnston M, Engelke DR. tRNA genes as transcriptional repressor elements. Mol Cell Biol. 1994;14:1266-77.
254. Good PD, Kendall A, Ignatz-Hoover J, Miller EL, Pai DA, Rivera SR, et al. Silencing near tRNA genes is nucleosome-mediated and distinct from boundary element function. Gene. 2013;526(1):7-15.
255. Zhang D, Gao Y, Zhu L, Wang Y, Li P. Advances and opportunities in methods to study protein translation - A review. International Journal of Biological Macromolecules. 2024;259:129150.

256. Ciandrini L, Stansfield I, Romano MC. Ribosome traffic on mRNAs maps to gene ontology: genome-wide quantification of translation initiation rates and polysome size regulation. PLoS Comput Biol. 2013;9(1):e1002866.
257. Hinnebusch AG. Molecular mechanism of scanning and start codon selection in eukaryotes. Microbiol Mol Biol Rev. 2011;75(3):434-67, first page of table of contents.
258. Alberts B, AlexanderLewis, JulianRaff, MartinRoberts, KeithWalter, Peter. Molecular Biology of the Cell2002.
259. Watson JD, Crick FHC. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. Nature. 1953;171(4356):737-8.
260. Crick FH. Codon--anticodon pairing: the wobble hypothesis. J Mol Biol. 1966;19(2):548-55.
261. Ou X, Cao J, Cheng A, Peppelenbosch MP, Pan Q. Errors in translational decoding: tRNA wobbling or misincorporation? PLoS Genet. 2019;15(3):e1008017.
262. Hani J, Feldmann H. tRNA genes and retroelements in the yeast genome. Nucleic Acids Research. 1998;26(3):689-96.
263. Zhang Z, Shah B, Bondarenko PV. G/U and certain wobble position mismatches as possible main causes of amino acid misincorporations. Biochemistry. 2013;52:8165-76.
264. Crombie T, Swaffield JC, Brown AJ. Protein folding within the cell is influenced by controlled rates of polypeptide elongation. J Mol Biol. 1992;228(1):7-12.
265. Kyogoku Y, Lord RC, Rich A. Hydrogen Bonding Specificity of Nucleic Acid Purines and Pyrimidines in Solution. Science. 1966;154(3748):518-20.
266. Eisenberg E, Levanon EY. A-to-I RNA editing—immune protector and transcriptome diversifier. Nature Reviews Genetics. 2018;19(8):473-90.
267. Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression. Trends Biotechnol. 2004;22:346-53.
268. Ikemura T. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. J Mol Biol. 1982;158:573-97.
269. Dong H, Nilsson L, Kurland CG. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. J Mol Biol. 1996;260:649-63.
270. Duret L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. Trends Genet. 2000;16:287-9.
271. Krüger MK, Sørensen MA. Aminoacylation of hypomodified tRNAGlu in vivo. Journal of molecular biology. 1998;284(3):609-20.
272. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, et al. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell. 2010;141(2):344-54.
273. Brackley CA, Romano MC, Thiel M. The dynamics of supply and demand in mRNA translation. PLoS Comput Biol. 2011;7(10):e1002203.
274. Soragni E, Kassavetis GA. Absolute gene occupancies by RNA polymerase III, TFIIIB, and TFIIIC in Saccharomyces cerevisiae. J Biol Chem. 2008;283(39):26568-76.
275. Moir RD, Willis IM. Regulation of pol III transcription by nutrient and stress signaling pathways. Biochim Biophys Acta. 2013;1829:361-75.
276. Torrent M, Chalancon G, de Groot NS, Wuster A, Madan Babu M. Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. Sci Signal. 2018;11(546).
277. Wang H, McManus J, Kingsford C. Accurate Recovery of Ribosome Positions Reveals Slow Translation of Wobble-Pairing Codons in Yeast. Journal of Computational Biology. 2016;24(6):486-500.
278. Percudani R, Pavesi A, Ottonello S. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. J Mol Biol. 1997;268:322-30.
279. Bennetzen JL, Hall BD. Codon selection in yeast. J Biol Chem. 1982;257:3026-31.
280. Xu Y, Liu K, Han Y, Xing Y, Zhang Y, Yang Q, et al. Codon usage bias regulates gene expression and protein conformation in yeast expression system P. pastoris. Microbial Cell Factories. 2021;20(1):91.

281. Hanson G, Coller J. Codon optimality, bias and usage in translation and mRNA decay. Nature Rev Mol Cell Biol. 2018;19:20-30.
282. Huang Q, Yao P, Eriani G, Wang ED. In vivo identification of essential nucleotides in tRNALeu to its functions by using a constructed yeast tRNALeu knockout strain. Nucleic Acids Res. 2012;40(20):10463-77.
283. Huang Y, Lin T, Lu L, Cai F, Lin J, Jiang Ye, et al. Codon pair optimization (CPO): a software tool for synthetic gene design based on codon pair bias to improve the expression of recombinant proteins in Pichia pastoris. Microbial Cell Factories. 2021;20(1):209.
284. Hershberg R, Petrov DA. Selection on codon bias. Annu Rev Genet. 2008;42:287-99.
285. Ikemura T. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes: Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of isoaccepting transfer RNAs. Journal of Molecular Biology. 1982;158(4):573-97.
286. Neafsey DE, Galagan JE. Positive selection for unpreferred codon usage in eukaryotic genomes. BMC Evol Biol. 2007;7:119.
287. Chothani S, Adami E, Ouyang JF, Viswanathan S, Hubner N, Cook SA, et al. deltaTE: Detection of translationally regulated genes by integrative analysis of Ribo-seq and RNA-seq data. Current protocols in molecular biology. 2019;129(1):e108.
288. Hussmann JA, Patchett S, Johnson A, Sawyer S, Press WH. Understanding biases in ribosome profiling experiments reveals signatures of translation dynamics in yeast. PLoS genetics. 2015;11(12):e1005732.
289. Nedialkova DD, Leidel SA. Optimization of codon translation rates via tRNA modifications maintains proteome integrity. Cell. 2015;161:1606-18.
290. Diament A, Feldman A, Schochet E, Kupiec M, Arava Y, Tuller T. The extent of ribosome queuing in budding yeast. PLoS computational biology. 2018;14(1):e1005951.
291. Dalvie NC, Brady JR, Crowell LE, Tracey MK, Biedermann AM, Kaur K, et al. Molecular engineering improves antigen quality and enables integrated manufacturing of a trivalent subunit vaccine candidate for rotavirus. Microb Cell Fact. 2021;20(1):94.
292. Baay M, Bollaerts K, Verstraeten T. A systematic review and meta-analysis on the safety of newly adjuvanted vaccines among older adults. Vaccine. 2018;36(29):4207-14.
293. Rappuoli R. Reverse vaccinology. Current opinion in microbiology. 2000;3(5):445-50.
294. Lal H, Cunningham AL, Godeaux O, Chlibek R, Diez-Domingo J, Hwang S-J, et al. Efficacy of an adjuvanted herpes zoster subunit vaccine in older adults. New England Journal of Medicine. 2015;372(22):2087-96.
295. Ford AQ, Touchette N, Hall BF, Hwang A, Hombach J. Meeting report: global vaccine and immunization research forum. Vaccine. 2018;36(7):915-20.
296. Nishimoto T, Furuta M, Kataoka M, Kishida M. Important Role of Catalase in the Cellular Response of the Budding Yeast Saccharomyces cerevisiae Exposed to Ionizing Radiation.
297. Liang S, Zou C, Lin Y, Zhang X, Ye Y. Identification and characterization of P GCW14 : a novel, strong constitutive promoter of Pichia pastoris. Biotechnol Lett. 2013;35(11):1865-71.
298. Zhang X, Zhang X, Liang S, Ye Y, Lin Y. Key regulatory elements of a strong constitutive promoter, PGCW14, from Pichiapastoris. Biotechnology Letters. 2013;35(12):2113-9.
299. Murakami H, Blobel G, Pain D. Isolation and characterization of the gene for a yeast mitochondrial import receptor. Nature. 1990;347(6292):488-91.
300. Berroteran RW, Hampsey M. Sequence, map position and genome organization of the RPL17B gene, encoding ribosomal protein L17b in Saccharomyces cerevisiae. Yeast. 1995;11(8):761-6.
301. Aebi M, Bernasconi R, Clerc S, Molinari M. N-glycan structures: recognition and processing in the ER. Trends in biochemical sciences. 2010;35(2):74-82.
302. Viinikangas T, Khosrowabadi E, Kellokumpu S. N-glycan biosynthesis: basic principles and factors affecting its outcome. Antibody Glycosylation. 2021:237-57.

303. Kukuruzinska MA, Lennon-Hopkins K. ALG gene expression and cell cycle progression. Biochimica et Biophysica Acta (BBA) - General Subjects. 1999;1426(2):359-72.
304. Burda P, Aebi M. The dolichol pathway of N-linked glycosylation. Biochimica et Biophysica Acta (BBA) - General Subjects. 1999;1426(2):239-57.
305. Boguta M. Maf1, a general negative regulator of RNA polymerase III in yeast. Biochim Biophys Acta. 2013;1829:376-84.
306. Cai Y, Wei Y-H. Stress resistance and lifespan are increased in C. elegans but decreased in S.cerevisia by mafr-1/maf1 deletion. Oncotarget. 2016;7:10812-26.
307. Michels AA. MAF1: a new target of mTORC1. Biochem Soc Trans. 2011;39(2):487-91.
308. Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science. 2009;324(5924):218-23.
309. Xue S, Barna M. Specialized ribosomes: a new frontier in gene regulation and organismal biology. Nat Rev Mol Cell Biol. 2012 May 23;13(6):355-69. doi: 10.1038/nrm3359. PMID: 22617470; PMCID: PMC4039366.
310. Shi Z, Fujii K, Kovary KM, Genuth NR, Röst HL, Teruel MN, Barna M. Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. Mol Cell. 2017 Jul 6;67(1):71-83.e7. doi: 10.1016/j.molcel.2017.05.021. Epub 2017 Jun 15. PMID: 28625553; PMCID: PMC5548184.