Universitat Autònoma de Barcelona

Overcoming the secretory limitations in Pichia pastoris for recombinant protein production

Juan José Barrero Peña

Universitat Autònoma de Barcelona

Escola d'Enginyeria

Departament d'Enginyeria Química, Biològica i Ambiental

Overcoming the secretory limitations in *Pichia pastoris* for recombinant protein production.

Memòria per a optar al grau de Doctor

per la Universitat Autònoma de Barcelona,

dins del programa de doctorat de Biotecnología

sota la direcció del Dr. Pau Ferrer i el Dr. Francisco Valero

per

Juan José Barrero Peña

Bellaterra, Juliol 2020

El Dr. Francisco Valero Barranco i el Dr. Pau Ferrer Alegre, catedràtic i professor agregat del Departament d'Enginyeria Química, Biològica i Ambiental, i membres del grup de recerca d'Enginyeria de Bioprocessos i Biocatàlisi Aplicada de la Universitat Autònoma de Barcelona.

Certifiquem:

Que el biotecnòleg Juan José Barrero ha dut a terme al Departament d'Enginyeria Química, Biològica i Ambiental de la Universitat Autònoma de Barcelona i amb la nostra direcció la tesi doctoral titulada **"Overcoming the secretory limitations in *Pichia pastoris* for recombinant protein production"**. La mateixa es presenta en aquesta memòria i constitueix el manuscrit per optar al Grau de Doctor en Biotecnologia per la Universitat Autònoma de Barcelona.

I per tal que se'n prengui coneixement i consti als efectes oportuns, signem la present declaració a Bellaterra, a 30 de maig de 2020.

FRANCISCO VALERO BARRANCO
Firmado digitalmente por FRANCISCO VALERO BARRANCO
Fecha: 2020.06.15 10:24:37 +02'00'

Dr. Francisco Valero

(co-director)

PAU FERRER ALEGRE - DNI 46123768M
Firmado digitalmente por PAU FERRER ALEGRE - DNI 46123768M
Fecha: 2020.06.16 09:51:24 +02'00'

Dr. Pau Ferrer Alegre

(co-director)

Juan José Barrero Peña

(Autor)

Acknowledgements

Primer de tot, m'agradaria agrair als meus directors Francisco Valero i Pau Ferrer per la seva ajuda durant l'etapa de la meva tesis. Tot i que al principi va ser força complicat tot plegat, al final tot ha sortit bé i he pogut enllestir una tesis amb la que estic molt content! Sense vosaltres això no hauria estat possible! I would also like to thank Ben Glick. The first part of the project would not have been possible without all your help. Thank you! También me gustaría agradecer a Jose Luis Montesinos por la ayuda con el sensor de metanol. Cuando el sensor estaba a puntito de fallar siempre ha estado allí y lo ha arreglado en nada. ¡La verdad que alguna que otra fermentación se ha salvado gracias a tu ayuda!

Por otra parte, ¡la tesis habría sido muchísimo más aburrida sin los chavales!! De los seniors primero Luismi, Kirian, Jordi y Sergi. De los que empezaron como yo Miquel, Javi y Manito. Y los que han empezado después Albert F, Josu y Arnau. La veritat que ha estat genial conviure aquesta etapa amb tots vosaltres. A ver si pronto podemos hacer algún Merlinete de nuevo, ¡pero sin bautizos! También quiero agradecer a mi máster Alejandro la ayuda que me distes durante tu proyecto de tesis. Fueron un par de noches duras, pero bastante amenas contigo.

A la gente del despacho les doy las gracias por todas las risas y crisis vividas juntos. Jesi, Alejandra, Xavi, Jordi y Xènia! La veritat que ha estat un plaer poder compartir tantes hores junts. Espero que puguem fer un SushiDespatch ben aviat! També agrair als laborants i secretaria per tota l'ajuda que m'heu donat en aquests anys en la ETSE.

Fuera de la ETSE, agradezco a todos los amigos que han estado en mayor a menor medida dándome apoyo en momentos malos y divirtiéndonos juntos en los momentos buenos. ¡Sois los mejores!

La verdad que de pocas cosas puedo estar más agradecido y orgulloso en esta vida que de tener la familia más maravillosa del mundo. A mis padres Juan y Rosa, a mi hermana Natàlia y a mi perrita Xula! En los tiempos malos siempre habéis estado junto a mí en todo momento y me habéis apoyado absolutamente en todo. ¡Os lo estaré agradecido toda mi vida! La mitad de esta tesis es absolutamente vuestra. También agradecer a las

“nuevas” incorporaciones de la familia, Javi y Anna. Aunque para ti Anna ahora hay un apartado!

Només tu Anna saps les ganes que tenia de poder enllestir ja la tesis… Sempre has estat incondicionalment al meu costat fos el què fos i m’has recolzat en tot moment, sobretot en els temps dolents. Ets la persona més meravellosa que he conegut en la meva vida i l’altre meitat de la tesis te la dec a tu. T’estimo amb tot el meu cor!

Abstract

The methylotrophic yeast *Pichia pastoris* (*Komagataella* spp.) has become one of the most popular cellular platforms to produce industrially relevant proteins, most of which are secreted in the extracellular media to facilitate subsequent downstream processing. The secretion of heterologous proteins out of the cell is mediated by the secretory pathway, a route that encompasses several steps and has different organelles interconnected to achieve an efficient secretion. However, when *P. pastoris* is forced to produce heterologous proteins, the proper function of the secretion pathway might be compromised. In this context, different studies have been carried out to identify specific bottlenecks and maximize protein production.

We identified the translocation of proteins from the cytoplasm to the Endoplasmic Reticulum (ER) as a major early bottleneck present in the secretory pathway. The translocation step is usually mediated by a signal sequence present at the N-terminal of the protein to be translocated. Depending on the signal sequence, this step can be performed while the protein is being translated (co-translational translocation) or after the translation (post-translational translocation). In addition, the secretion signal has to contain a so-called pro-region to mediate a fast export of the recombinant protein from the ER to the Golgi for subsequent secretion.

In the first part of this study, we focused on the engineering and characterization of the secretion signal peptide as a strategy to improve recombinant protein secretion. For comparison, we used the secretion signal from the alpha mating factor (α-MF) from *Saccharomyces cerevisiae*, which is the most common secretion signal used by the scientific community to secrete recombinant proteins in *P. pastoris*. This secretion signal is reported to drive proteins through the post-translational translocation. As a result, if the α-MF is fused to a protein that can fold in the yeast cytosol, the protein may be unable to traverse the ER translocon and enter the secretory pathway. The solution was to replace the pre-α-MF signal sequence with the pre-Ost1 signal sequence, which directs co-translational translocation across the ER membrane, thereby ensuring that heterologous proteins fold only after reaching the ER lumen. Additionally, the pro-region from the α-MF contained a region prone to aggregation, which was easily

removed by exchanging a threonine by a serine at position 42 (Ser42). The resulting hybrid secretion signal drastically increased the production of three different model proteins used in this thesis: a fluorescent model protein called E2-Crimson and two different lipases of industrial interest, namely the lipase 2 from *Bacillus thermocatenulatus* (BTL2) and a lipase from *Rhizopus oryzae* (ROL). More importantly, these findings were then tested at bioreactor scale, thereby obtaining similar results to those observed at shake flask scale. Notably, strains with the improved secretion signal had a better cell performance in comparison with the α-MF.

Secondly, after releasing the bottleneck present in early stages of the secretory pathway, we decided to perform a transcriptional analysis to further pinpoint other possible limitations that might be present downstream the secretory pathway. To this end, we used the hybrid secretion signal and the α-MF secretion signal to drive secretion of the same model protein (E2-crimson) for comparative purposes. The study was carried out in fed-batch operated bioreactor cultures. Despite the increased protein secretion levels in the initial stages of the induction phase when using the pre-Ost1 signal, our results pointed that an accumulation of protein is occurring at the ER lumen. Thereby, increased transcriptional levels of unfolded protein response related genes and, consequently, of the ER-Associated Degradation pathway, resulted in a reduction of intracellular protein levels as well as reduced production rates at the later stages of the fermentation.

Overall, a new hybrid secretion signal has been proposed to replace the α-factor secretion signal as the default standard for producing heterologous proteins in *P. pastoris*. However, to fully exploit the power of the improved secretion signal, additional cellular engineering is needed to overcome bottlenecks that appear downstream of the translocation event, particularly under bioprocess (fed-batch) conditions.

Ressenya

El llevat metilotròfic *Pichia pastoris* (genere *Komagataella*) s'ha convertit en una de les plataformes cel·lulars més populars per produir proteïnes d'interès industrial, la majoria de les quals es secreten extracel·lularment per facilitar el posterior procés de purificació. La secreció de proteïnes heteròlogues fora de la cèl·lula està mediada per la via secretora, una ruta que engloba diversos passos i té diferents orgànuls interconnectats per aconseguir una secreció eficient. Tanmateix, quan *P. pastoris* es veu obligat a produir proteïnes heteròlogues, el correcte funcionament de la via de secreció podria veure's compromesa. En aquest context, s'han realitzat diferents estudis per identificar colls d'ampolla específics i maximitzar la producció de proteïnes.

Es va identificar la translocació de proteïnes del citoplasma al reticle endoplasmàtic (ER) com un coll d'ampolla precoç important que es troba a la via secretora. El pas de translocació sol estar mediat per un pèptid senyal que es troba al N-terminal de la proteïna a traduir. Depenent de la seqüència senyal, aquest pas es pot realitzar mentre la proteïna es tradueix (translocació co-traduccional) o després de la traducció (post-traducció). A més, la senyal de secreció ha de contenir l'anomenada regió "pro" per mediar una ràpida exportació de la proteïna recombinant des del ER al Golgi per tal que s'acabi secretant.

A la primera part d'aquest estudi, ens vam centrar en l'enginyeria i la caracterització del pèptid senyal de secreció com a estratègia per millorar la secreció de proteïnes recombinants. Per tal de comparar-les, hem utilitzat el senyal de secreció de *Saccharomyces cerevisiae* anomenat "alfa mating factor" (α-MF) que és el senyal de secreció més comú utilitzat per la comunitat científica per secretar proteïnes recombinants en *P. pastoris*. Es diu que aquest senyal de secreció condueix proteïnes mitjançant la traducció post-traduccional. Com a resultat, si el α-MF es fusiona amb una proteïna que es pot plegar en el citosol del llevat, és possible que la proteïna no pugui travessar el translocó del ER i entrar a la via secretora. La solució va ser substituir la seqüència senyal pre-α-MF per la seqüència senyal pre-Ost1, que dirigeix la translocació de forma co-traduccional a través del ER, garantint així que les proteïnes heteròlogues només es pleguin després d'arribar al lumen del ER. Addicionalment, la pro-regió del α-

MF contenia una regió propensa a l'agregació, que va ser fàcilment eliminada mitjançant l'intercanvi d'una treonina per una serina en la posició 42 (Ser42). Aquest senyal de secreció híbrid resultant va augmentar dràsticament la producció de tres proteïnes models diferents utilitzades en aquesta tesi: una proteïna model fluorescent anomenada E2-Crimson i dues lipases diferents d'interès industrial anomenades Lipasa 2 de *Bacillus thermocatenulatus* (BTL2) i Lipasa de *Rhizopus oryzae* (ROL). Més important encara, es van provar aquestes troballes a escala bioreactor, obtenint així resultats similars als observats a escala matràs. En particular, les soques amb el senyal de secreció millorat van tenir un millor rendiment cel·lular en comparació amb l'α-MF.

En segon lloc, després d'alliberar el coll d'ampolla present en les primeres etapes de la via secretora, vam decidir realitzar una anàlisi transcripcional per identificar altres possibles limitacions que puguin estar presents a la via secretora. Per aquest motiu, es va utilitzar el senyal de secreció híbrid i el senyal de secreció α-MF per estudiar la secreció de la mateixa proteïna model (E2-Crimson) amb finalitats comparatives. L'estudi es va realitzar a escala bioreactor amb una estratègia d'alimentació del tipus "fed-batch". Malgrat l'augment dels nivells de secreció de proteïnes en la fase inicial d'inducció quan s'utilitza el senyal pre-Ost1, els nostres resultats van assenyalar que s'està produint una acumulació de proteïna al lumen del ER. Per aquest motiu, es va donar també un augment dels nivells transcripcionals de gens relacionats amb UPR i, en conseqüència, de la via de degradació associada a l'ER (ERAD), que va donar lloc a una reducció dels nivells de proteïnes intracel·lulars així com uns valors de producció reduïts en les etapes posteriors de la fermentació.

En resum, s'ha proposat un nou senyal de secreció híbrid per substituir el senyal de secreció α-MF com a estàndard predeterminat per produir proteïnes heteròlogues en *P. pastoris*. No obstant això, per aprofitar plenament el potencial del senyal de secreció millorat, es necessita enginyeria cel·lular addicional per superar els colls d'ampolla que apareixen posteriorment de la translocació, particularment en condicions de bioprocessos com en un fed-batch.

Table of contents

1

General Introduction

Table of Contents

1. Yeast as an industrial workhorse for recombinant protein production

One of the manifold application areas of biotechnology is the production of metabolites and proteins with a wide range of applications. The area of recombinant protein production (RPP) emerged in the early 1980s as an alternative to overcome the limitations imposed by the obtention of proteins from natural resources. At that time, the insulin for the treatment of diabetes was obtained from pigs and cows with its obvious complex purification processes, together with the potential immunogenicity reactions and risk of unknown infectious agents. This problem was circumvented in 1982 by producing the first human recombinant insulin in microbial systems. Since then, by improving and creating new pharmaceutical and other industrial products, biotechnology is becoming one of the most important sectors to cover human needs.

Heterologous proteins are biological complex products that have to be created by using a host organism capable of synthesize the protein biologically active. To do so, a high variety of host organisms are available. Nevertheless, not all microbial cell factories have the same capacities [1]. In Table 1 the three most important groups of host organisms (bacteria, yeast and mammalian cells) are represented with their advantages and disadvantages.

The most common model organism among bacteria is *Escherichia coli*, whose genetic versatility allows the generation and improvement of industrial strains in an easy way. The big advantage of *E. coli* as host organism is its growth capacity and high productivity. Moreover, the costs related to media culture are relatively low and, consequently, this host organism is attractive in terms of efficiency. However, the inability to perform post-translational modifications restricts the use of *E. coli* - especially for complex human proteins for pharmaceutical use. Furthermore, it is generally unable to secrete heterologous proteins which usually remain as cytoplasmic aggregates inside the cell, forming so-called inclusion bodies [2]. Although the most extensively used bacterial system is generally unable to secrete proteins, some gram-positive bacteria such as the *Bacillus* genus, which are able to secrete proteins, have also been developed as host organisms for both academic and industrial scale RPP [3].

On the other hand, there is the group of mammalian cells, which contains elevated species diversity. Since these cells are originated from multi-cellular, higher eukaryotic organisms, most post-translational modifications are done correctly (i.e. similarly to human cells) and proteins can be secreted. Therefore, complex biopharmaceutical

proteins are usually produced in this host organism [4] However, the high production costs and lower productivities of mammalian systems make them less economically competitive for industrial scale production [5].

The protein production platform with increasing importance in this field is the group of yeasts [6,7]. As unicellular eukaryotic microorganisms, yeasts have the ability to perform eukaryotic post-translational modifications of the expressed proteins, which confers on this group the capability to synthesize and secrete functional foreign proteins [8]. Moreover, their rapid growth and high-cell density fermentation capacity in minimal medium have a particularly high impact on the large-scale industrial production of heterologous proteins [9].

Table 1. Main characteristics of the three most important biological systems available for recombinant protein production.

Group organism	Advantages	Disadvantages
Bacteria	Cheap cultivation media	No posttranslational modifications
	High growth rates	
	Grow at high cell densities	
	Easy to manipulate genetically	
	High productivity	
Yeast	Protein folding and secretion	Different protein glycosylation pattern than human glycoproteins
	Posttranslational modifications	
	Cheap cultivation media	
	High growth rates	
	Grow at high cell densities	
	Easy to manipulate genetically	
	High productivity	
Mammalian cells	Posttranslational modifications	Expensive cultivation media
	Protein folding and secretion	More difficult to scale up
	Expression of complex proteins	Slow growth rates

Yeasts combine the ease of genetic manipulations and fermentation of a microorganism with the capability to secrete and modify foreign proteins according to a general eukaryotic scheme [10]. Among the group of yeast, the best-known organism is the baking yeast *Saccharomyces cerevisiae.* This yeast has been used over the past centuries for the baking and brewery industry and has been a model organism for the scientific community. Therefore, a lot of knowledge has been gathered over the years. However, the yeast phylogenetic group has a huge genetic and physiological diversity [11]. For this reason, some non-convention yeasts have emerged as alternative host organisms for RPP. Among this group of so-called non-conventional yeasts, *Pichia pastoris,* the model organism used for this thesis, has emerged as an important system for RPP both for academic and industrial applications.

2. *Pichia pastoris* as a host system for RPP

2.1. Historical overview

Pichia pastoris (syn *Komagataella* spp.) is a methylotrophic yeast widely used for recombinant protein production. *P. pastoris* was first discovered by Koichi Ogata [12] 40 years ago and was rapidly implemented in the early 70s by the Phillips Petroleum Company to consume the methanol derived from the oil generation and produce yeast biomass (single cell protein) that would be marketed as protein animal feed. At the end of the 60s and beginning of the 70s, Phillips Petroleum developed media and strategies to grow *P. pastoris* in methanol as sole carbon source. However, in the same period, the cost of methanol increased drastically and, thus, the production of single cell protein based on this substrate was not economically competitive with the use of the soybean as a source of animal feed protein.

In the early 80s, with the start of genetic engineering tools and the beginning of the age of biotechnology, Philips Petroleum contracted a biotechnology company called the Salk Institute Biotechnology/ Industrial Associate Inc. (SIBIA) to develop *P. pastori*s as a host organism to express different heterologous proteins with industrial interest. At this time, the researchers in SIBIA isolated the *AOX1* gene and developed a set of different plasmids containing the strong and tightly regulated *AOX1* promoter (P_{AOX1}) [13–15]. In

addition, different strains and molecular genetic methods for the proper manipulation of this host organism were developed [16]. The combination of fermentation media and methods created by Philips Petroleum as well as the different vectors and strains developed by SIBIA resulted in high protein productions for a wide range of different proteins in the 90s [17]. Philips Petroleum sold its patent to Research Corporation Technologies (RCT), the current holder of the license, and also licensed Invitrogen (currently part of ThermoFisher) to sell components of the system such as plasmids and strains. Both, strains and plasmids, are still being widely used worldwide.

Analysis of ribosomal RNA sequences from subunits 18S and 26S derived into a relocation of the genus of *Pichia pastoris* to *Komagataella* spp. and the previous strains were split into three different species: *Komagataella phaffi*, *Komagataella pastoris* and *Komagataella pseudopastoris* [18,19]. However, the scientific community is still referring all these different organisms as *Pichia pastoris*.

More recently, with the increasing advances in the sequencing methods, the genome of some strains of *P. pastoris* were fully sequenced, which allowed a better understanding of this host organism and facilitated the genetic engineering modifications [20,21].

2.2. Advantages of *P. pastoris* over *S. cerevisiae*

Although *P. pastoris* has received an increasing interest over the past 20 years for RPP, *S. cerevisiae* is still a frequently used yeast cell factory to produce heterologous proteins due to the vast knowledge that *S. cerevisiae* has obtained in all the areas of research as a model organism for several years.

The advantages that *P. pastoris* can provide to the research community over *S. cerevisiae* to produce heterologous proteins can be summarised as follows:

- **Protein glycosylation pattern:** Many heterologous proteins have to undergo through post-translation modifications to have a proper activity. Among all the post-translational modifications, disulfide bonds and glycosylations are highlighted as key elements to confer the proteins a proper structure and functionality. For therapeutic glycoproteins implemented in human treatments,

defective glycosylation patterns can also induce immunogenic reactions and reduce the circulatory life span of the drug. Although both *S. cerevisiae* and *P. pastoris* are capable of glycosylate proteins, their glycosylation pattern is different [22]. *S. cerevisiae* tends to hyperglycosylate proteins introducing 50-150 residues length of N-linked oligosaccharides, while *P. pastoris* introduces around 20 residues. This reduction of glycan residues makes *P. pastoris* a suitable organism to produce glycoproteins with potentially reduced immunogenicity. In fact, in the early 2000s, GlycoFi stood out as a company with a technology capable of creating glycoproteins with a fully human glycosylation pattern [23,24]. This company was bought by Merck but was finally closed down years later. In parallel, the group of Nico Callewaert developed the so-called Pichia GlycoSwitch system, a set of engineered strains capable of producing proteins with a human-like glycosylation pattern. These strains were licensed to the company RCT for their commercialisation [25–27]. These strains are capable of producing glycoproteins with no immunogenic reactions.

- **Space-time product yields:** Protein production yield is usually correlated with a higher amount of biomass in the culture media [28]. Unlike *S. cerevisiae*, *P. pastoris* is a Crabtree negative yeast and has the capacity to grow at high-cell densities reaching a final biomass of up to 150 g/L of dry cell weight (DCW) [29–31]. Nevertheless, it is important to mention that such biomass concentration needs a robust monitoring and a reliable control of process parameters at large volume scales to maximize space-time product yields, especially in those processes based on methanol utilization [28].
- **Fermentative phenotype:** Yeast has the capacity to generate ethanol as a response to different concentrations of substrate or oxygen. This is called the Crabtree effect. *S. cerevisiae* is defined as a Crabtree positive because it uses glycolysis as the terminal electron acceptor, despite sufficiently high dissolved oxygen concentration in the culture, and this effect is magnified when the carbon source concentration exceed a certain limit [32]. As stated in the previous point, *P. pastoris* is defined a Crabtree negative and, therefore, it doesn't have a fermentative phenotype under aerobic conditions [33]. Since ethanol is a toxic by-product derived from a fermentation, a fermentative-negative phenotype

reduces the risk of cellular stress derived from ethanol toxicity during the fermentation, thereby increasing both process robustness and carbon source efficiency when using a Crabtree-negative yeast.

- **Reduced secretome:** Secretion of heterologous proteins is usually of great interest to minimize the downstream purification steps and reach a proper purity of the expressed protein. In *S. cerevisiae*, the final amount of endogenous proteins secreted are much larger than in *P. pastoris*, where around 350 proteins are secreted in *S. cerevisiae* in comparison with around 50 secreted proteins in *P. pastoris* [34–38]. This makes *P. pastoris* an attractive alternative to express proteins with much higher purity as a starting point for the downstream purification.
- **Grows on glycerol at a high growth rate:** Glycerol is an abundant by-product obtained from the production of biodiesel and, as a non-fermentable C-source, has been used to minimize the Crabtree effect mentioned above. Therefore, the use of glycerol as a C-source generates even less amount of ethanol in comparison with glucose, obtaining higher biomass yields [39]. Notably, the maximum specific growth rate reported on glycerol of *P. pastoris* is 0.28 h^{-1} [40], being between 0.20 h^{-1} and 0.25 h^{-1} a value commonly obtained in lab-scale bioreactors [40]. These values are significantly higher than those observed for *S. cerevisiae* using glycerol, ranging from 0.05 h^{-1} to 0.13h^{-1} depending on the strain [41,42].
- **The strong and inducible promoter P_{AOX1}:** As explained above, *P. pastoris* has the capacity to consume methanol as a sole carbon source in organelles called peroxisomes. The first step of the methanol metabolic pathway is catalysed by the alcohol oxidase 1 and 2 (Aox1, Aox2). *AOX1* is strongly expressed in the presence of methanol and repressed in presence of other carbon sources such as glycerol or glucose [14,43]. Therefore, the *AOX1* promoter (P_{AOX1}) soon became of great interest for the scientific community, as it is one of the strongest eukaryotic promoters able to express heterologous proteins with industrial interest and enables the decoupling of growth and protein production.

2.3. *P. pastoris* strains for RPP

In the context of recombinant protein production, different *Pichia* strains have been engineered to cover a wide range of needs, which can be essentially categorised as follows: Strains better adapted to process constraints related to oxygen and heat transfer demands for gene expression under the P_{AOX1}. Strains engineered to meet product constraints e.g. protease-susceptibility or humanized glycosylation patterns. Need of different auxotrophic markers to support strain engineering [9]. Each category of strains is further explained in the next sections below. Furthermore, different strains combining two or more categories of traits are also available. As an example, the classic series of "best-sellers" strains commercialised by Invitrogen (Thermofisher) are shown in Table 2. Nevertheless, additional strains aside from the Invitrogen strains have been generated and successfully used for protein expression [44].

2.3.1. Strains with different methanol utilisation (Mut) phenotypes

In those recombinant protein production processes based on the use of the *AOX1* promoter, different strains have been engineered to consume methanol at different rates. *P. pastoris* has two alcohol oxidase encoding genes, which are called *AOX1* and *AOX2*. Both isoenzymes catalyze the first step of the methanol assimilation pathway, i.e. the oxidation of methanol into formaldehyde. However, *AOX1* is expressed 10 times more than *AOX2*. The so-called MUT (**M**ethanol **Ut**ilization) variants, are engineered strains with different AOX genes deletions [14]. Thus, these strains can minimize the methanol consumption to different extent, thereby reducing the oxygen and heat transfer limitations, decreasing the growth rate on methanol while maintaining the protein induction.

- **Mut^+:** This strain contains both of the AOX genes and can grow optimally on methanol. With this strain, methanol is typically used as gene expression inducer as well as carbon source.
- **Mut^S:** This strain contains just the *AOX2* gene functional. Therefore, the methanol consumption rate is 10 times slower than the MUT^+ strain. Methanol is usually used as inducer and as a co-substrate with another carbon source (e.g glycerol or sorbitol) [15].

- **Mut$^-$:** This strain does not contain any functional AOX gene and, therefore methanol is not metabolised by any of the AOX genes and P_{AOX1} can be used as the heterologous gene expression inducer during a fermentation process. Although a recent report shows that a MUT$^-$ strain can still grow on methanol, very little growth is observed in comparison with a MUT^S phenotype [45]. Therefore, another carbon source is necessary to maintain the growth during the fermentation. A first example of a strain with the MUT$^-$ phenotype is the MC100-3 [14].

2.3.2. Mutant strains for genetic engineering

To introduce any genetic element in *P. pastoris*, it is necessary a selection marker capable of differentiate the new genetic variant to the wild type. More importantly, this marker must differentiate the new strain unequivocally without obtaining any false positive or background. In general, the amount of selection markers available will limit the amount of genetic modifications that will be possible to introduce in the strain and, therefore, different markers have been obtained based on antibiotic resistance or auxotrophy. To use an auxotrophic marker, the *P. pastoris* strain must lack a gene necessary to generate important organic compounds such as amino acids, so the plasmid to integrate can have the gene and grow on a minimal media. In general, wild-type strains can easily grow on minimal media, so different strains have been generated lacking important genes and use them as auxotrophic markers. Among all the new generated strains, the GS115 and *Pichia*Pink [46] have been widely used as industrially relevant strains for protein production.

- **GS115:** This strain lacks the *HIS4* gene (ΔHis4) necessary to produce the amino acid Histidine. Therefore, the plasmid to be integrated can contain the *HIS4* to generate auxotrophy and grow on plates lacking histidine[21,47,48].

- ***Pichia*Pink:** This strain is an adenine auxotrophic strain with a precise single gene deletion in the *ADE2* gene (Δade2). The absence of the *ADE2* gene leads to the accumulation of a red metabolic intermediate, phosphoribosylaminoimidazole,

giving a dark pink colour to the cells. Transformation of *P. pastoris* ade2 mutant cells with the pPink-HC vector re-introduces the knocked-out gene (*ADE2*) into the *P. pastoris* chromosome. Re-establishing the adenine biosynthetic pathway reverts the pink cell colour to wild-type white, allowing visual selection of transformants [46,49].

2.3.3. Strains for specific protein requirements

Each protein to be produced has its requirements and particularities to be fully functional and many *P. pastoris* strains have been created to cover them. Among all the different parameters to be considered, the glycosylation pattern and protease sensitivity are two of the top considerations to take into account when producing proteins.

As mentioned in the previous section, for therapeutic glycoproteins implemented in human treatments, defective glycosylations can also induce immunogenic reactions and reduce the circulatory life span of the drug. Therefore, for therapeutic purposes, a humanized glycosylation pattern is necessary to properly produce glycoproteins while avoiding immunoreactivity issues. The Glycoswitch strain was indeed the solution for this problem as this is a genetic engineered strain with the glycosylation pattern from humans [22,26].

Additionally, *P. pastoris* secretes proteases to the extracellular medium [34]. Among all the proteases, Pep4 and Prb1 have been shown to degrade heterologous proteins produced in *P. pastoris* [30]. Fortunately, knocking out these genes does not affect the viability of the strain and significantly reduce protein degradation, creating a protease-deficient set of strains [50]. However, these strains are usually recommended as the last resort because of their deficient growth behaviour and lower transformation efficiencies [51].

Table 2. Strains commercialised by Invitrogen (Thermofisher) for the production of recombinant proteins.

Different parameters	Phenotype	Strains	Reference
Methanol related strains	MUT^+	X33, GS115	[16]
	MUT^S	KM71, KM71H	[15]
	MUT^-	MC100-3	[14]
Auxotrophic strains	Adenine auxotrophy	PichiaPink	[46,49]
	Histidine auxotrophy	GS115, KM71, SMD1168	[15,16]
Protein related strains	Modified –glycosilation pattern	SuperMan strains (PichiaSwitch)	[22,25–27]
	Protease deficiency	Pep4: SMD1168 Prb1: SMD1165	[50]

2.4. Genetic engineering tools for protein production in *P. pastoris*

P. pastoris needs all the necessary genetic elements to properly encode the protein of interest. For this purpose, over the past years several vectors (plasmids) with all the genetic elements have been created. In *S. cerevisiae*, these set of plasmids can be either integrative in the genome or episomal. However, *P. pastoris* does not have stable episomal plasmids as in the case of *S. cerevisiae* where there are origins of replications for multicopy plasmids (2µ) and low copy plasmids (CEN). For this reason, plasmids available for *P. pastoris* are generally vectors that have to be integrated in the genome [16]. The best known integrative plasmids used for protein production and secretion are the Invitrogen set of plasmids, specially the pPICZαA and pGAPZαA, which are exactly the same plasmids but containing the P_{AOX1} or P_{GAP}, respectively, to either induce or constitutively produce the protein of interest.

The necessary genetic elements that a plasmid has to contain to properly express a gene and secrete the protein of interest are summarized as follows:

- **Selection marker:** To properly identify those positive clones containing the plasmid, a genetic element able to properly select positive clones is necessary. The classical markers available to select positive clones are

through antibiotic resistance or auxotrophy. For the antibiotic resistance, the media is usually supplemented with a specific antibiotic and the plasmid contains a gene which encodes a protein that confers the strain antibiotic resistance. In *P. pastoris*, the commonly used antibiotics are Zeocin, Hygromycin, Geneticin (G418) and Nourseothricin (NAT). As for the auxotrophic markers, the selected strain has to be defective to produce a necessary metabolite. This strain cannot grow on minimum media unless the plasmid is integrated and restores the capacity of the strain to produce the metabolite. As explained in section 2.3.2, the best-known auxotrophic strains are PichiaPink and GS115, which are defective to metabolize Adenine (*ADE1* gene) and Histidine (*HIS4* gene), respectively. Although these are the most common auxotrophic strains, strains defective for the synthesis of other essential metabolites such as arginine (*ARG4* gene) or uracil (*URA3* gene) are also available [9,30]. However, these strains are usually used for basic research instead of being used as recombinant protein producer strains.

- **Integration locus:** The capacity of *P. pastoris* to have homologous recombination events occur in a lesser extent than *S. cerevisiae*. In fact, for a positive recombination in *P. pastoris* at least 500bp with homology in the genome are usually necessary, while in *S. cerevisiae* just 50bp are necessary for a positive recombination [52]. Therefore, the integration locus has to be longer than in other yeasts to confer an efficient plasmid integration. The integration locus can be the endogenous promoter or terminator used to express the gene that encodes the protein of interest. However, some plasmids have the locus of integration separated from the transcription unit and serves as an independent genetic element [53].
- **Origin of replication (ORI):** To efficiently obtain enough genetic material to produce all the genetic modifications and cloning, the plasmids are first created in *Escherichia coli*. The ORI element is necessary to allow *E. coli* to replicate the plasmid. This plasmid is then extracted from *E. coli* and purified to be subsequently transformed in *P. pastoris*.
- **Promoter:** This genetic element is necessary to express the gene that encodes the protein of interest. In the plasmids pPICZαA and pGAPZαA the

promoters are P_{AOX1} or P_{GAP}, which allow the strain to either induce protein production or constitutively produce the protein of interest, respectively. Many efforts have been done to find new promoters with different strengths and regulatory attributes [54,55].

- **Secretion signal:** This genetic element is present at the very beginning of the coding region from the gene of interest. This part commences just after the promoter and, once the protein is encoded, gives the cell the information to transport the encoded protein to the secretory pathway and its posterior secretion. The best-known secretion signal is the alpha mating factor from *S. cerevisiae* [56] and it is present in the plasmids pPICZαA and pGAPZαA mentioned before. Further information about this genetic element is explained in sections below.
- **CDS:** The coding region of a gene (CDS) includes the secretion signal mentioned before as well as the gene that encodes the protein of interest.
- **Terminator:** After the stop codon in the gene of interest, there is the so-called terminator. This genetic element contains all the necessary information to stop the translation of the protein as well as some elements that confer stability to the messenger RNA.

2.4.1 New genetic engineering tools for synthetic biology

Since the start of the genetic engineering, the construction of expression vectors has been of special interest to create strains capable of producing heterologous proteins or to confer a strain new ability. All the plasmids were first created by using a classic restriction-ligation approach. The so-called conventional cloning consists in the use of restriction enzymes to select a genetic part and combine this region in a single plasmid using a ligase. This cloning strategy is really time-consuming and does not allow the assemble of multiple fragments in a single backbone simultaneously. Fortunately, during the last years, more advanced cloning techniques have stood out as alternative cloning strategies to speed up the process of creating different expression plasmids. The most relevant cloning techniques are Gibson assembly and Golden Gate assembly, which are based on the use of a mix of polymerase, exonuclease and ligase, and type IIs restriction enzymes, respectively. More importantly, both techniques are able to assemble multiple

fragments in a single reaction, and without any scar (i.e. extra base pairs derived from the cloning) [57,58].

Taking advantage in these new biological approaches, a Modular Cloning (MoCLo) strategy was first implemented by the group of Sylvestre Marillonet to use the Golden Gate strategy in plants and assemble up to six fragments in a single reaction. In this strategy, the type IIs restriction enzymes BbsI and BsaI were used with predefined overhangs that mediated the position of each fragment [57]. The group of Diethard Mattanovich was the first to adapt the system created by Marillonet for *P. pastoris* and create a Golden Gate-derived modular cloning system called GoldenPiCS [59]. This modular system includes 20 promoters, 10 transcription units, 4 locus of integration and 4 antibiotic resistance cassettes, and can be combined to create plasmid with different characteristics. More importantly, this system allows the addition of up to 8 transcriptional units simultaneously, which really speeds up the cloning process and allows complex genetic engineering strategies such as creation of new pathways where several genes have to be cloned. A similar approach was also implemented with Gibson assembly [60].

In addition, the new genome editing strategy CRISPR was also implemented in *P. pastoris* as a new system to knock-out genes instead of the conventional knock-out strategies such as the split marker. However, this system still needs some fine-tuning in *P. pastoris* to increase its efficiency [61–63].

2.5. Pichia pastoris in the Market

P. pastoris was recognized as "Generally Recognized as Safe" (GRAS) organism by the Food and Drug Administration (FDA) [64]. This status enabled the use of this organism and its derivatives for human use. After this recognition, in 2009 there was the first market approval of a recombinant biopharmaceutical product produced in *P. pastoris* called Kalbitor (a recombinant kallikrein inhibitor protein) [65]. Since then, several products produced in *P. pastoris* have been commercialized for therapeutic use. Several of these products are summarized in Table 3.

Table 3. Different commercialized products produced in *P. pastoris*. (source: www.pichia.com, visited on 03/03/2020.

PRODUCT	COMPANY	USE
Kalbitor® (DX-88 ecallantide, a recombinant kallikrein inhibitor protein)	Dyax (Cambridge, MA)	Hereditary angioedema treatment
Insugen® (recombinant human insulin)	Biocon (India)	Diabetes therapy
Medway (recombinant human serum albumin)	Mitsubishi Tanabe Pharma (Japan)	Blood volume expansion
Shanvac ™ (recombinant hepatitis B vaccine)	Shantha/Sanofi (India)	Hepatitis B prevention
Shanferon™ (recombinant interferon-alpha 2b)	Shantha/Sanofi (India)	Hepatitis C & Cancer treatment
Ocriplasmin (recombinant microplasmin)	ThromboGenics (Belgium)	Vitreomacular adhesion (VMA) treatment
Nanobody® ALX-0061 (recombinant anti-IL6 receptor single domain antibody fragment)	Ablynx (Belgium)	Rheumatoid arthritis treatment
Nanobody® ALX00171 (recombinant anti-RSV single domain antibody fragment)	Ablynx (Belgium)	Respiratory syncytial virus (RSV) infection treatment
Heparin-binding EGF-like growth factor (HB-EGF)	Trillium (Canada)	Treatment of interstitial cystitis/bladder pain syndrome (IC/BPS) treatment
Purifine (recombinant phospholipase C)	Verenium/DSM (San Diego, CA/Netherlands)	Degumming of high phosphorus oils
Recombinant trypsin	Roche Applied Science (Germany)	Digestion of proteins
Recombinant collagen	Fibrogen (San Francisco, CA)	Medical research reagents/dermal filler
AQUAVAC IPN (recombinant infectious pancreatic necrosis virus capsid proteins)	Merck/Schering Plough Animal Health (Summit, NJ)	Vaccines for infectious pancreatic necrosis in salmon
Recombinant phytase	Phytex, LLC (Sheridan, IN)	Animal feed additive
Superior Stock recombinant nitrate reductase	The Nitrate Elimination Co. (Lake Linden, MI)	Enzyme-based products for water testing and water treatment
Recombinant human cystatin C	Scipac (United Kingdom)	Research reagent

3. The secretory pathway

Protein secretion is an essential process for all eukaryotes and this pathway encompasses several steps mediated by a big amount of proteins located in different organelles creating a protein network [11]. In yeast, protein folding and secretion have been identified as a major bottlenecks for recombinant protein production [66]. The minimum requirement that a secreted protein needs to contain in its precursor state is a secretion signal peptide at the N-terminus, which enables the cellular transport machinery to correctly target a given protein to a specific cell compartment or outside the cell [67].

The secretory pathway is then initiated by the translocation of the protein through the Endoplasmatic Reticulum (ER) membrane. Once the protein has been pulled inside this organelle, a series of post-translational modifications and protein folding steps are performed. Additionally, the ER also serves as a major protein quality control checkpoint, a process that is mediated by resident proteins. Misfolded or defective proteins are ubiquitinated and redirected to the cytoplasm, where they are recognized and degraded by the proteasome complex. Only those proteins that pass the ER quality control mechanism are directed to the Golgi compartment by vesicular transport. In the Golgi apparatus, the proteins undergo further modifications and are also rechecked by a quality control machinery that resides in the Golgi compartment. Finally, proteins that have passed the quality control are secreted by exocytosis in their native, biologically functional form [68]. All the organelles involved in the secretion pathway are shown in Figure 1 and are further explained in the next sections below.

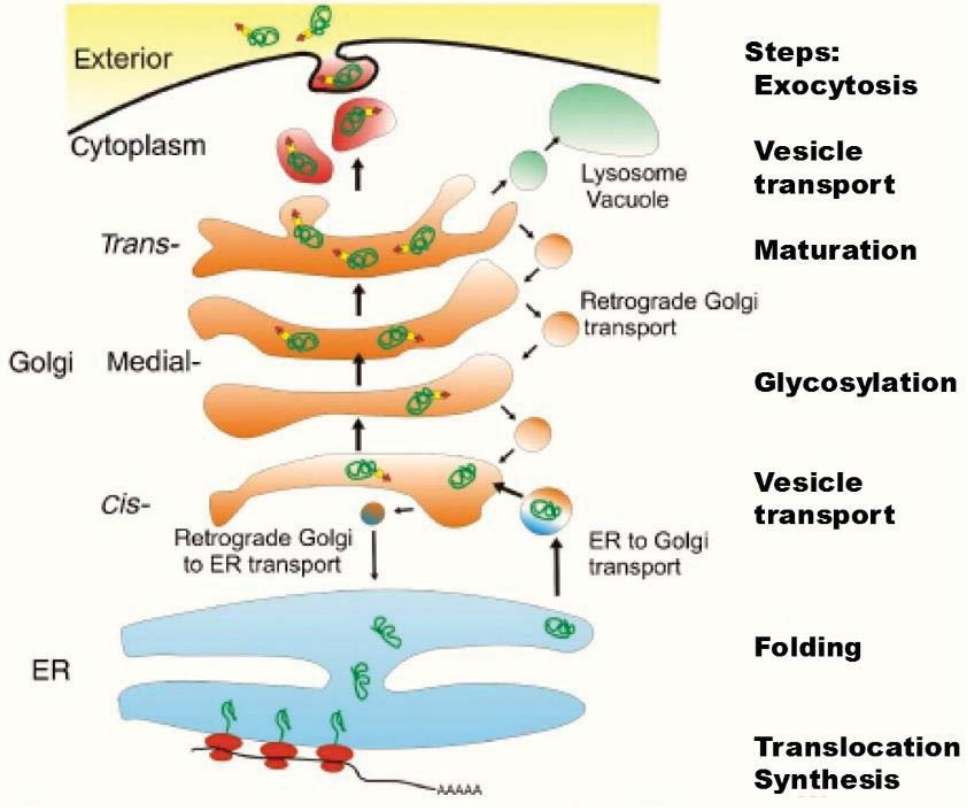

Figure 1. Different steps from the secretory pathway. Figure taken from [2] and reproduced with the permission of the publisher.

3.1. The start of the secretory pathway: The protein translocation into the ER

Translocation of proteins across ER can occur either co-translationally or post-translationally. In the first case, translation and translocation occur simultaneously, which means that the ribosome is linked to the translocation channel and the nascent polypeptide is inserted at the same time as it is synthesized. In contrast, post-translational translocation occurs when the ribosome is uncoupled and, consequently, translation and translocation occur in different steps. Nevertheless, both routes use the same protein channel that makes the translocation possible. In yeasts, this channel is composed of a highly conserved, heterotrimeric core called Sec61 complex [69].

The hydrophobicity of the signal sequence (also known as the pre sequence) located at the N-terminus of the polypeptide determines whether a polypeptide is targeted by the co-translational or post-translational pathway. In case the precursor protein contains a very hydrophobic signal sequence, it is preferentially targeted to the co-translational

pathway, while less hydrophobic signal peptides tend to be linked with the post-translational pathway [70].

The co-translational pathway starts when the Signal Recognition Particle (SRP) complex binds both to the signal sequence in the nascent polypeptide and the translating ribosome. In this case, the SRP binds to the SRP receptor (SR) and the ribosome binds to the translocon site. This mechanism induces the nascent polypeptide to be transferred into the aqueous channel of the translocon Sec61 (Figure 2.A).

The post-translational pathway is SRP independent. After the release of the polypeptide from the ribosome, it has to remain unfolded and without aggregation in order to be translocated. This is accomplished by binding to the cytosolic chaperones Ssa1 and Ydj1, which also transport the polypeptide to the translocon pore (Figure 2.B).

Once the polypeptide is translocated, both pathways recruit SPase (Signal Peptidase), which cleaves the signal peptides after the protein has entered the ER lumen.

Despite some exceptions [71], in eukaryotic cells most proteins that enter in the ER are translocated by co-translational translocation. Coupling the translocation of nascent polypeptides with their synthesis has the advantage of introducing proteins into the ER while they are still unfolded. This increases the efficiency of translocation across the membrane through the narrow Sec61 translocation pore [72]. Therefore, a plausible strategy to increase secretion of heterologous proteins in yeast would be targeting with a signal peptide that triggers co-translational translocation.

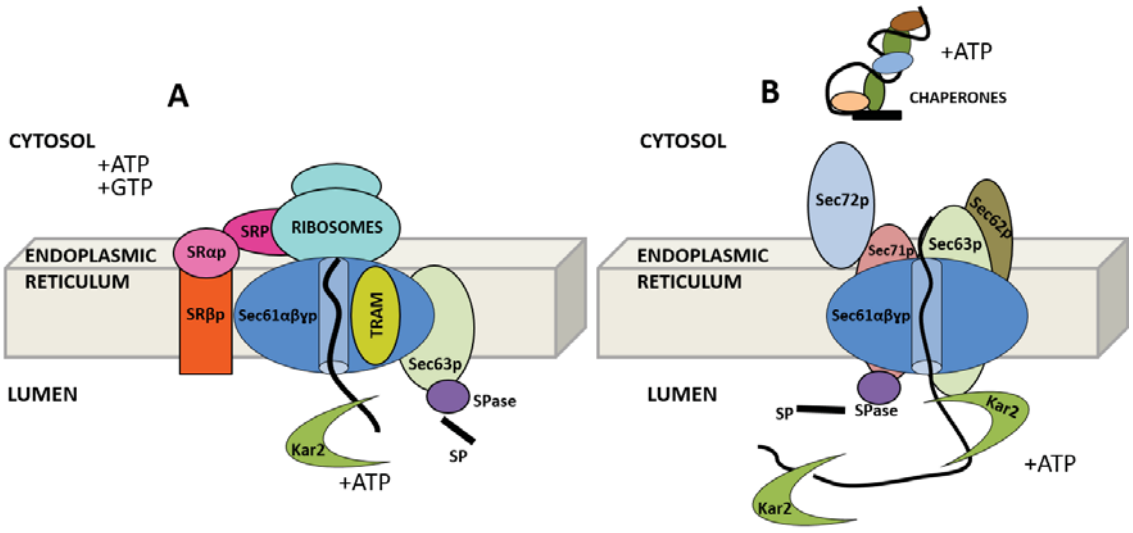

Figure 2. Different translocation pathways in eukaryotes. In Fig. 2A there is the pathway of co-translational translocation. In Fig. 2B there is the post-translational translocation pathway.

3.2. Endoplasmic reticulum

The endoplasmic reticulum (ER) is the largest membrane-bound organelle in the eukaryotic cell [73,74]. In yeast, the ER is located contiguous with the nuclear envelope as well as the plasma membrane [75]. It is the first organelle present in the secretion pathway and several functions are present in this organelle including calcium homeostasis, lipid synthesis and nascent protein maturation. Conditions are oxidizing, ATP is available as an energy source, and the ionic environment is pH neutral with a high concentration of calcium. Therefore, the environment in the ER is optimized for the folding and maturation of proteins for export and after the protein translocation into the ER lumen, heterologous proteins start to fold and obtain a tertiary structure with the aid of ER-resident chaperones, oxidoreductases and glycosylation enzymes [76,77].

The first protein interactor in the ER with the nascent protein to be matured is Kar2 (BiP in mammals). Kar2 interacts with the protein being translocated by pulling in the nascent protein into the ER. Then, Kar2 still interacts with the nascent protein as a chaperone together with other chaperones such as LHS1 and SIL1 to maintain the proper conformation of the protein to be produced [78]. In addition, other proteins as PDI1 interacts with the nascent protein to introduce the disulfide bonds if necessary.

Within this organelle, an elaborate quality control system regulates ER homeostasis by ensuring the fidelity of protein synthesis and maturation. The two main regulators necessary to maintain correct protein folding and the protein quality control are the ER-Associated Degradation (ERAD) and the Unfolded Protein Response (UPR) [79].

- **UPR:** In the context of recombinant protein production, the bulk of protein to be expressed or the conditions in which cells are being grown can create a metabolic burden and generate an increased number of unfolded or misfolded proteins. Additionally, this cellular stress can produce a high number of reactive oxygen species (ROS) such as superoxide anion and hydrogen peroxide which derives an oxidative stress and creates a REDOX imbalance [80]. In yeast, this cellular stress activates the UPR by the action of the transmembrane signal transducer Ire1 and the transcription factor Hac1. This transmembrane protein has a lumenal sensor

domain that can interact with Kar2 and unfolded proteins to sense protein folding status [81,82]. Under stress conditions, Ire1 is activated and interacts with Hac1 mRNA. In normal conditions, Hac1 mRNA has an intron which prevents Hac1 to be active as a transcription activator [83]. However, the activated Ire1 facilitates the splicing of Hac1 mRNA and allows the translation of an active form of Hac1 that is finally located in the nucleus and serves as an transcription factor with the capacity to activate or upregulate at least 381 genes such as chaperones, foldases and even proteins present in the ERAD [79].

- **ERAD:** The ER-Associated Degradation maintains the quality of the secretory proteome by degrading those proteins that fail to achieve the proper conformation due to translational or transcriptional mutations, or an inefficient assembly in their native structure [75]. There are three different ERAD pathways (ERAD-L, -M and -C) and exist to remove proteins located in different regions [83]. ERAD-M removes those transmembrane proteins present in the ER, ERAD-C removes proteins in the cytosol and ERAD-L removes proteins that are located in the ER lumen. Most of the produced proteins are soluble proteins present in the ER lumen, therefore the ERAD-L pathway is of great interest for the recombinant protein production [84]. In the ERAD-L, defective proteins are retro-translocated back to the cytosol, where they are ubiquitinated and degraded by the proteasome. The key player for the ERAD-L is the Hrd1 complex, a retrotranslocation channel consisting in proteins Yos9 and Hrd3, and are involved in the protein selection together with protein Der1 which is the protein responsible to insert defective proteins into the Hrd1 complex [85].
  ERAD is also stimulated by the UPR and is known to be one of the major systems to release the protein cargo into the ER during protein production, especially when the secretory capacity of the cell is overloaded [84].

Those proteins that succeeded into obtaining a native structure are then sent to the Golgi through the ER Exit Sites (ERES). The ERES are compartments inside the ER in which small vesicles called COPII are created [86]. These COPII vesicles are small carriers where cargo proteins are packed and sent to the Golgi and this phenomenon is called

anterograde transport. In addition, when proteins are sent to the Golgi erroneously, they are recognized in the Golgi and sent back to the ER through COPI vesicles by a phenomenon called retrograde transport [87].

To gather all the necessary proteins to be sent to the Golgi, several sorting receptors are present in the ER [87]. One of the transmembrane receptors with increasing interest in the field of protein production is the Erv29 transmembrane protein [88,89]. This protein is the specific receptor for the mating alpha factor (MF-α) in yeast and recognizes the alpha factor by the pro-region from its secretion signal. Heterologous proteins are usually tagged with the same secretion signal and, consequently, all these proteins are recognized by the Erv29 receptor to have a rapid export to the Golgi [90].

3.3. The Golgi network

After leaving the ER, secretory cargo proteins arrive to the Golgi, where they still suffer post-translational modifications such as glycosylations and are finally secreted to the extracellular media [91].

The Golgi apparatus has three different stages depending on its maturation state. The first one is called the early or cis Golgi and is the stage in which the Golgi is first formed by the constant assembly of COPII vesicles. Then, this early Golgi is matured until the formation of the trans-Golgi, a stage in which the glycosylations usually start forming [92]. Finally, this trans-Golgi is finally converted to the Trans-Golgi- Network (TGN) where the vesicles with the protein cargo are segregated and ultimately fused with the cytoplasm with its consequent release of the cargo inside the vesicles in the extracellular media [93].

Even though the Golgi might seem a final exit organelle to send the protein out of the cell, the compartment is rather dynamic. This organelle interacts with endosomes and the vacuole to recycle misfolded proteins [94]. The Vps10 transmembrane receptor is located at the trans-Golgi and plays a crucial role to sort misfolded proteins to the vacuole. Therefore, even if an unfolded protein can escape the ERAD machinery, in the Golgi can still be sent to the vacuole and be degraded [95].

For heterologous protein expression, the trans-Golgi is also crucial for the final produced protein as it contains Kex2 and Ste13 recognition sites that allow the cleavage of the secretion signal [96].

3.4. Reported bottlenecks appearing during the secretion pathway

In the context of recombinant protein production, many efforts have been made to overcome all the bottlenecks present in the secretion pathway. In *P. pastoris*, many of the improvements obtained by overexpressing different proteins involved in the secretion pathway for recombinant protein production are nicely reviewed in bibliography [97].

Several improvements are reported when overexpressing genes encoding for chaperones and foldases such as Kar2, Lhs1, Ero1 and PDI for both, *S. cerevisiae* and *P. pastoris* [80,98–101]. In addition, there has also been reported improvements in terms of protein secretion when overexpressing genes encoding for cytosolic chaperones such as Ssa1 and important components from the Sec61 translocon pore such as Sbh1 or Sec61, which seemed to facilitate the proper protein translocation into the ER [102,103]. The overexpression of the active form (or spliced form) of the gene encoding for the transcription factor Hac1 has been of great help to activate the UPR and increase the production of several heterologous proteins. Interestingly, in *P. pastoris* the endogenous *HAC1* as well as the *HAC1* from *S. cerevisiae* were overexpressed in different studies, conferring both *HAC1* variants improvement in protein secretion [98,99,104–106]. In the case of the ERAD, only a few genes related to ERAD can be knocked-out as cell viability was usually perturbed or a significant decrease of cellular growth was observed [84]. Just deletion of genes *DER1* or *DOA1* have been of help to increase protein production in budding yeast [84,90].

In *S. cerevisiae*, the overexpression of the rapid-export transmembrane *ERV29* has been implemented, resulting in an improvement of protein production. Presumably, the overexpression of *ERV29* might increase the levels of Erv29 which would target heterologous protein to the ERES with a fast export to the Golgi and, consequently, the ER might get less overloaded [90]. Another strategy to increase protein production

tested in *S. cerevisiae* was the deletion of genes involved in the lipid regulation, resulting in an ER expansion. The knocked-out genes that reported an improvement of protein productions were *PAH1* and *OPI1* [90,107]. At the Golgi level, the deletion of the gene encoding for the vacuolar sorter Vps10 also increased protein secretion. However, one of the functions of Vps10 is to sort vacuolar hydrolases into the vacuole and without Vps10, such proteins were also secreted, resulting in a higher degradation rate of secreted proteins. The solution was to delete part of the gene encoding for one of the domains from the Vps10 to allow the proper location of the vacuolar hydrolases while eliminating the targeting of the heterologous protein to be secreted. The reported domain deletion was called Vps10-104 [95]. In addition, overexpression of the gene encoding the Kex2 endoprotease also increased final protein production in *P. pastoris* [108].

Something worth mentioning is that, when overexpressing different genes in yeast, the selection of an adequate promoter is key to fine-tune expression of the gene [107]. An excess of a protein helper factor could even yield in a decrease of protein expression instead of obtaining any improvement. In addition, not all the reported gene deletions or genes overexpression work the same way for different heterologous proteins or even with the same protein but using a different promoters or gene dosage. All the proteins have different characteristics and, therefore, different bottlenecks might occur at the secretion pathway. In conclusion, for each expressed protein an exhaust tailoring of the secretion pathway is necessary to overcome all the limitations present in this system [97].

4. The importance of the secretion signal to start protein secretion

To introduce any protein into the secretory pathway, it is necessary to add at the N-terminal a so-called secretion signal. The secretion signal usually contains two main regions: The first region, also called the pre-region, is a signal sequence that gives the cell the information to introduce the protein inside the ER. The second region is called the pro-region and consists in a rapid ER export that target the proteins to the ERES to be sent ultimately to the Golgi [109]. Reports suggest that just the first part of the secretion signal could be enough to obtain some protein secretion. However, the lack of the pro region might significantly reduce the amount of secreted protein and end up overloading the ER [95]. This is because the protein does not have the yeast HDEL ER retention signal as in the case of Kar2. Therefore, there is some protein leakage throughout the secretion pathway. In conclusion, for an optimized secretion leader, both parts, the pre and pro regions, must be present in the secretion signal [95,109].

The most common secretion sequence used for heterologous protein production in *P. pastoris* is the alpha-mating factor (α-MF) secretion signal from *S. cerevisiae*. This secretion signal sequence was first discovered more than 30 years ago and was then rapidly implemented in the area of recombinant protein production for *S. cerevisiae* [110,111]. *P. pastoris* is a budding yeast with many similarities with *S. cerevisiae*, therefore this signal sequence was the first choice to start secreting the expressed heterologous proteins in this yeast system [112]. Since then, the α-MF signal has been broadly used for both, the scientific community and industry.

As most of the secretion signals, the α-MF secretion signal has two main regions the pre and pro region, and the entire secretion signal is shown in Figure 3.

- **pre-α-MF**: This signal sequence has 19 amino acidic residues that directs translocation into the ER. This pre-region is SRP independent and, therefore, the signal drives the tagged proteins through the post-translational translocation [113]. The relatively low hydrophobicity of this signal sequence is the major reason for the signal to drive the protein through the post-translational translocation [114]. At the end of the signal sequence, there is a cleavage site

that cleaves the pre-sequence at the position 19 just after the proper translocation of the tagged protein into the ER [115] (Figure 3).

- **pro-α-MF**: This pro region is formed by 66 amino acids decisive to mediate the receptor-dependent packaging into the ER-derived COPII vesicles. The pro-α-MF has been predicted to contain one alpha helix and 5 beta strands [116] as well as 3 glycosylation sites [113]. The pro-α-MF sequence recognition is carried out by the transmembrane protein Erv29 and permits the rapid export of the protein to the Golgi [88]. Then, at the Golgi level, the two amino acids present at the end of the pro-region (Lysine and Arginine) are recognized by the protease Kex2 and cleaved. Additionally, Glutamate and Alanine, also present at the end of the pro-region, are recognized by the Ste13 peptidase and cleaved as well [115] (Figure 3).

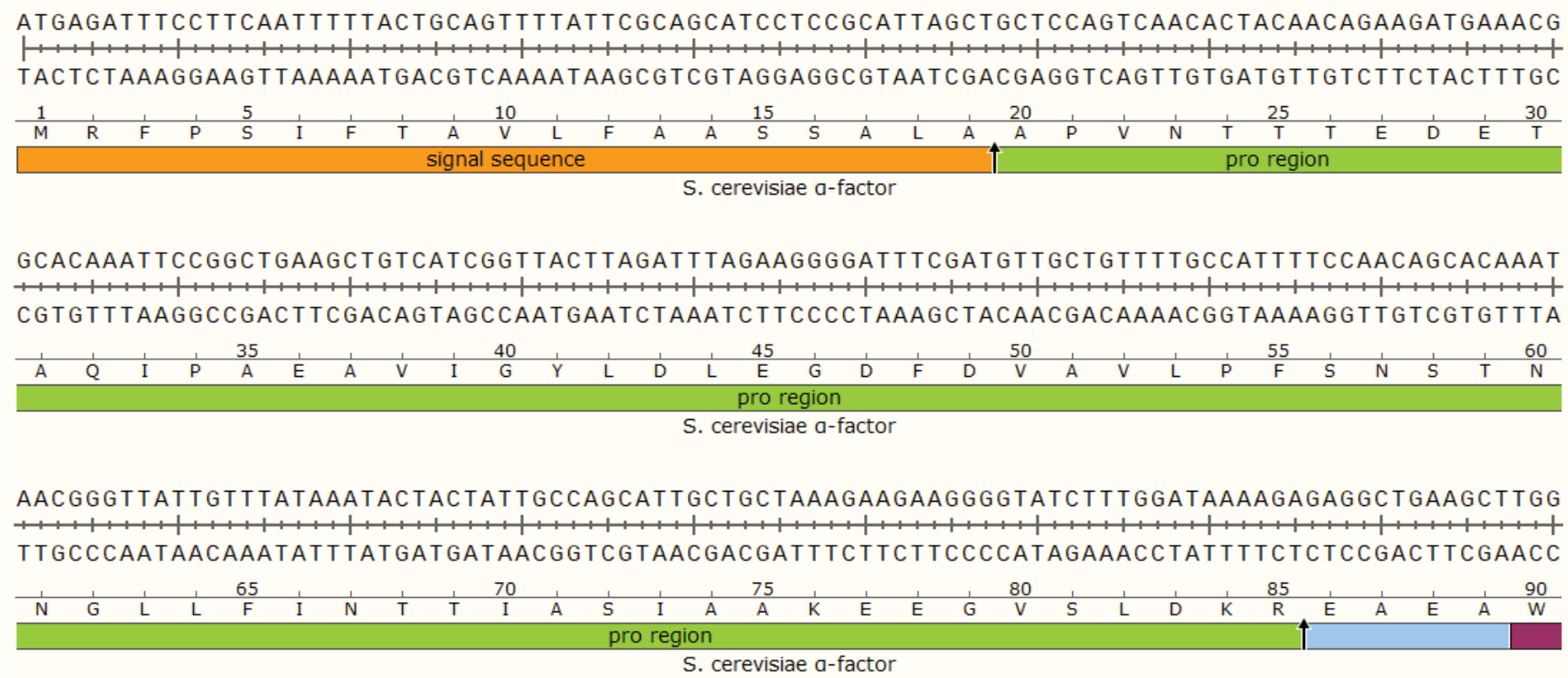

Figure 3. The *S. cerevisiae* α-MF secretion signal. The 5' portion of the α-MF gene encoding pre-pro-α-factor is depicted. pre-α-MF (signal sequence) is shown in orange and pro-α-MF is shown in green. Cleavage sites for signal peptidase and Kex2 are marked as arrows, and cleavage site for Ste13 is shown in blue. This image was generated using SnapGene software.

4.1. Reported strategies to overcome the limitations of the secretion signal

Although the α-MF has been the preferred secretion signal for budding yeast, some reports suggest that this secretion signal might not be the best choice to express heterologous proteins. For instance, a recent paper has shown that the α-MF signal might not translocate all the expressed protein into the ER [84]. In addition, the pro-α-MF could have a tendency to aggregate due to some hydrophobic regions that might reduce the secretion efficiency [117].

To overcome the translocation limitation from the pre-α-MF, some researchers have tried to modify the pre-region by increasing the hydrophobicity of the signal sequence. However, although the hydrophobicity of the signal sequence is key for a proper translocation, the length of the signal and the structure also plays an important role to properly translocate the proteins [114]. In addition, it has also been studied that a different pre region that drives the proteins through the co-translational translocation might be beneficial for recombinant protein production [95].

Conversely, there are studies in which the pro-region has been modified to increase the secretion efficiency. Some researchers have tried to rationally delete or double some parts of the pro regions, obtaining some secretion improvements [116,117] and also there are studies in which the researchers modified the secretion sequence through direct evolution and obtained secretion improvements as well [109,118]. A more recent study reports that just the first 6 amino acids from the pro-region is enough to mediate proper secretion in *S. cerevisiae*. The first three amino acids (APV) are enough to be recognized by Erv29 and the next three amino acids (NTT) are necessary for the asparagine-linked glycosylation and can enhance proper secretion as well [89,93]. However, further experiments have to be done to show that just these 6 amino acids are enough to outperform the full-length of the α-MF secretion signal. Another strategy to increase protein secretion with the α-MF secretion signal has been to codon optimize the secretion sequence [119]. Additionally, in *S. cerevisiae* a spacer after the Kex2 restriction site efficiently improved production of insulin or by fine-tuning the first amino acid after the Kex2 restriction site also improved protein production in *P. pastoris* [108,120].

An additional strategy has been to test other secretion signals from both, *S. cerevisae* [121] as well as *P. pastoris* [122,123]. For *P. pastoris*, an *in silico* determination of different endogenous secretion signals have been done using signal sequence prediction programs such as SignalP [124]. These new endogenous secretion signals could also be of great interest to improve the secretion of heterologous proteins.

5. Different strategies to produce heterologous proteins in a fermentor level based on the P_{AOX1} system

Once a potential cell protein producer has been obtained and tested in a small culture scale, the next step is to test whether this strain behaves properly in an industrial-like fermentation process. In this context, the bioprocess engineering is a key element to study and find the best bioprocess strategy to obtain the maximum amount of product in the minimum amount of time (space-time yield) while maintaining the complexity of the process as low as possible [125].

Depending on the selected promoter, the protein expression could be constitutive or inducible and, thus, might totally change the bioprocess strategy or the parameters to be used [125]. Since the promoter used in this thesis has been the inducible P_{AOX1}, the explained cultivation strategies are focused on the heterologous protein production in MUT^+ strains with the P_{AOX1}-based system.

In *P. pastoris*, some operational parameters such as temperature, pH and the dissolved oxygen in the culture are important to the proper growth of this yeast:

- **Temperature:** The optimum temperature to grow *P. pastoris* is usually 30°C, above 30°C growth gets reduced due to cellular stress [126]. However, there are some reports in which lower temperatures (between 20°C and 25°C) have been used to maximize the yield of soluble recombinant proteins, reduce the metabolic burden derived from oxidative stress and cell lysis, and to reduce the proteolytic activity [127,128].

- **pH:** *P. pastoris* can grow in a pH ranging from 3 to 7 [129], but the pH can change the specific growth rate and activate host proteases [130]. The best pH to work

is between 5 and 6, and is the preferred pH used in most of the studies [40]. Nevertheless, something important to check when producing any protein is the isoelectric point (pI) of the expressed protein and maintain the pH of the culture distant to avoid protein precipitation [131].

- **Dissolved oxygen:** In fermentations where methanol is the main carbon source, dissolved oxygen is a key element for the methanol oxidation to formaldehyde and its consequent metabolization. An excess of oxygen is crucial to have the cells growing in fully aerobic methanol metabolism. Typically, dissolved oxygen is maintained at least above 20-30% of oxygen saturation to maximize protein production [132,133].

Three different fermentation strategies are used to produce recombinant proteins and these systems are explained in the following sections.

5.1. Batch cultivation

A batch fermentation is usually the easiest process to produce heterologous proteins. The reactor contains all the necessary elements to grow and the fermentation is kept until the substrate is fully consumed. In this kind of fermentation process, the strain grows at a theoretical maximum growth rate (μ_{MAX}) unless the substrate produces toxicity to the cells, or the cells are inhibited by side-products. This process is also useful when you want to do a direct comparison between strains or if you want to have an approximate idea on how the strain behaves and even to get some basic process kinetics such as μ, q_s, q_p and yields [134]. However, the results obtained in this process could differ from processes in which the cells grow at high cell densities.

5.2. Fed-batch cultivation

The Fed-batch cultivation strategy is usually the best choice to maximize the production of recombinant protein in terms of space-time yield as cells grow at high cell densities and there is usually a correlation between biomass concentration and produced protein.

In this process, the reactor does not start with all the necessary substrate, so the rest of the carbon-source is added during the fermentation. Therefore, the volume of the culture is increased as well. When heterologous proteins are expressed via the P_{AOX1}-based system, there are three different stages during the fed-batch fermentation [135]. The first stage consists in a glycerol batch phase in which the objective is the fast generation of biomass before the methanol induction. When the glycerol is fully consumed, the transition phase starts. In this second phase, the goal is to increase the cell density of the culture and the derepression of the AOX1 promoter. In the transition phase, limiting glycerol is added at the same time that methanol is also added, so the cell metabolism can adapt to the consumption of methanol. Although the transition phase could be skipped [136], some researchers fully recommend the transition phase to improve the cell adaptation and maximize final protein production [137]. Once *P. pastoris* can consume methanol, the induction phase starts. In the third and last phase, the methanol is added and acts as both, carbon source and protein inducer.

In the induction phase, the methanol feeding strategy is one of the most important factors to maximize heterologous protein production as it can directly control bioprocess parameters such as specific growth and production rates. The methanol feeding can be added in pulses, which each pulse is usually no more than 10 g/L of methanol and the next pulse is added after a sudden peak in the oxygen concentration. This is the easiest addition strategy in comparison with other options such as the addition profiles. In *P. pastoris*, the most important feeding addition profiles reported in literature are constant addition, methanol limited addition and methanol non-limited addition, and all of them are explained below:

- **Constant methanol addition:** The constant addition is the easiest strategy to use in a fed-batch. In the induction phase, there is a constant methanol flow that is kept the same during the entire fermentation. As the cells start growing, the

amount of methanol will not be the same and the growth rate will gradually decrease in the fermentation. This addition strategy does not exert any control on the fermentation parameters and cannot maximize the production rate.

- **Methanol Limited addition (MLFB):** The MLFB is an open-loop addition profile strategy with a μ-dependent exponential feeding. The carbon source is limited during the entire fermentation trying to maintain a constant specific growth rate during the entire fed-batch. The addition profile is based on the mass balance equation and allows the system to keep a specific growth rate during the entire induction process. This addition profile is one of the most used profiles for a fed-batch due to a high reproducibility and feasibility to monitor. However, the system does not respond to perturbations of the bioprocess. To avoid this problem, the set-point of μ is fixed far from the maximum growth rate μ_{MAX}, so the methanol doesn't get accumulated [135]. There are some new papers that show how an open-loop with an specific growth rate could maximize the final production of a protein of interest [138,139].

- **Methanol non-Limited addition (MNLFB):** This closed-loop addition profile system allows the fermentation to maintain a specific concentration of methanol during the entire fermentation process. To do so, there are commercially available e.g. on-line methanol sensors (e.g. from Raven Biotech) that can measure the concentration of methanol in the reactor. There are studies confirming that a constant concentration of methanol can increase the final production of certain heterologous proteins [140,141]. The reported optimum concentration of methanol is 3 g/L and this is probably due to a higher induction of the P_{AOX1}, while this concentration does not exert any inhibition issue as it could happen at higher concentrations of methanol where the strain could decrease the growth rate [141]. To perform such addition control, a Proportional-Integrative Derivative (PID) control algorithm can be used, allowing the system to efficiently predict the necessary amount of methanol to be added to maintain the methanol concentration setpoint. For an implementation of this system in industry, robust and sophisticated control algorithms based on a

predictive control have been developed and optimized to increase the adaptation capacity [142].

5.3. Continuous cultivation

In a continuous cultivation, there is a constant feeding flow in the reactor system while the same constant flow is removing part of the culture from the reactor, maintaining the same volume. Thus, force the cells to grow at a constant growth rate (dilution rate), reaching a physiological steady-state of all the cells. In general, this system has a low volumetric productivity due to the low biomass reached constant dilution of the culture in comparison with a fed-batch system where there is an accumulation of cells and heterologous protein produced inside the reactor. However, this system is really stable and could be maintained much longer than systems such as fed-batches [143]. However, when comparing productivities between both operational modes over a long period time (months), continuous operation could be better due to reduced downtimes of cleaning in place and re-start of fed-batch operated plants [144].

Currently, the main applicability of this operating system in *P. pastoris* is for physiological studies in lab-scale bioreactors, as the operational conditions can easily be controlled, and cells are kept in a steady-state. Thus, continuous cultures can allow scientists to study the effect of process parameters, either individually or in combination, on production, and select the best operational condition to maximize the production of a given strain. In addition, with this system it is possible to compare the behaviour of different strains. Several studies have been carried out in our group using this experimental approach [125,138,139,145].

6. Model proteins used in this study

To properly show the extent of the results obtained in this thesis, proteins with different structural properties were selected. As a starter point, a fluorescent protein called E2-Crimson was selected to track down the protein inside the cell and investigate potential bottlenecks appearing during the secretion pathway. Then, the results obtained were verified using two different lipases of industrial interest, namely the lipase 2 from *Bacillus thermocatenulatus* (BTL2) and a lipase from *Rhizopus oryzae* (ROL).

6.1. E2-Crimson

E2-Crimson is a far-red derivative of the tetrameric fluorescent protein DsRed-Express2 [146]. Its optimum excitation and emission wavelength are 611 nm and 646 nm respectively. In contrast with other fluorescent proteins, E2-Crimson matures faster due to a lack of post-translational modifications (e.g. disulfide bonds) and a fast self-folding capacity. Furthermore, this protein is not cytotoxic, has a high photostability and a good brightness [147]. Therefore, it is possible to overexpress it at high levels without having problems with cell viability. In addition, its photostability together with its brightness, substantially decrease photobleaching, making E2-Crimson well suited to live-cell imaging [93,148]. Furthermore, the excitation and emission from E2-Crimson does not interfere with any molecule as in the case of riboflavin with the Green Fluorescent Protein (GFP) [149]. Thus, makes E2-Crimson easy to study its fluorescence without any interference from other components.

We envisioned that this protein could have problems to enter the ER due to its capacity to fold in a rapid way, thus making this protein a good model for implementing different engineering strategies to increase translocation into the ER. Then, by visualizing cellular fluorescence under the microscope, we would be able to find other possible compartments where E2-Crimson could get stuck.

6.2. Lipases: BTL2 and ROL

Lipases (EC 3.1.1.3) are hydrolytic enzymes that catalyze the hydrolysis of the ester bonds of triacylglycerols. Their application is wide open as they can catalyze many types of reactions in aqueous and non-aqueous media [150]. In industry, among all the applications, production of detergents, food and pharma industry and production of

biodiesel are highlighted. In this thesis, the selected proteins with industrial relevance were two different microbial lipases: the mature lipase 2 from *Bacillus thermocatenulatus* (BTL2) and the mature *Rhizopus oryzae* lipase (ROL):

- **ROL:** Its structure and function have already been well characterized [151]. Besides its industrial applications, the mature form of ROL has also been used as a model protein to investigate potential bottlenecks from the host *P. pastoris* during secretion of heterologous proteins, since simple and robust enzymatic assays are available for enzyme quantification [152]. It is known that overexpression of the mature form of ROL triggers the unfolded protein stress response in *P. pastoris* [153]. However, nobody has investigated whether ROL has complications entering into the ER or not.

- **BTL2:** This lipase is a promising lipase with a high thermostability and catalytically active at high pH and in the presence of organic solvents [154–156]. *P. pastoris* has already been engineered to secrete BTL2 which ensures that this protein can be an effective model protein capable of being secreted by *P. pastoris* [157].

Overall, these two structurally different lipases can give us a wide spectrum on how the improvements obtained in E2-Crimson can be translated in proteins with industrial relevance.

7. References

[1] Berlec A, Štrukelj B. Current state and recent advances in biopharmaceutical production in Escherichia coli, yeasts and mammalian cells. J Ind Microbiol Biotechnol 2013;40:257–74. https://doi.org/10.1007/s10295-013-1235-0.

[2] Graf A, Dragosits M, Gasser B, Mattanovich D. Yeast systems biotechnology for the production of heterologous proteins. FEMS Yeast Res 2009;9:335–48. https://doi.org/10.1111/j.1567-1364.2009.00507.x.

[3] Anné J, Vrancken K, Van Mellaert L, Van Impe J, Bernaerts K. Protein secretion biotechnology in Gram-positive bacteria with special emphasis on Streptomyces lividans. Biochim Biophys Acta - Mol Cell Res 2014;1843:1750–61. https://doi.org/10.1016/j.bbamcr.2013.12.023.

[4] Walsh G. Biopharmaceutical benchmarks 2018. Nat Biotechnol 2018;36:1136–45. https://doi.org/10.1038/nbt.4305.

[5] Pourmir A, Johannes TW. Directed evolution: selection of the host organism. Comput Struct Biotechnol J 2012;2:e201209012. https://doi.org/10.5936/csbj.201209012.

[6] Corchero JL, Gasser B, Resina D, Smith W, Parrilli E, Vázquez F, et al. Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics. Biotechnol Adv 2013;31:140–53. https://doi.org/10.1016/j.biotechadv.2012.09.001.

[7] Baghban R, Farajnia S, Rajabibazl M, Ghasemi Y, Mafi A, Hoseinpoor R, et al. Yeast Expression Systems: Overview and Recent Advances. Mol Biotechnol 2033;61:365–84. https://doi.org/10.1007/s12033-019-00164-8.

[8] Ellis T, Adie T, Baldwin GS. DNA assembly for synthetic biology: from parts to pathways and beyond. Integr Biol (Camb) 2011;3:109–18. https://doi.org/10.1039/c0ib00070a.

[9] 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;98:5301–17. https://doi.org/10.1007/s00253-014-5732-5.

[10] Idiris A, Tohda H, Kumagai H, Takegawa K. Engineering of protein secretion in yeast: Strategies and impact on protein production. Appl Microbiol Biotechnol 2010;86:403–17. https://doi.org/10.1007/s00253-010-2447-0.

[11] Delic M, Valli M, Graf AB, Pfeffer M, Mattanovich D, Gasser B. The secretory pathway: Exploring yeast diversity. FEMS Microbiol Rev 2013;37:872–914. https://doi.org/10.1111/1574-6976.12020.

[12] Ogata K, Ohsugi M, Nishikawa H. A Yeast Capable of Utilizing Methanol. Agric Biol Chem 1969;33:1519–20. https://doi.org/10.1271/bbb1961.33.1519.

[13] Ellis SB, Brust PF, Koutz PJ, Waters AF, Harpold MM, Gingeras TR. Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast Pichia pastoris. Mol Cell Biol 1985;5:1111–21. https://doi.org/10.1128/mcb.5.5.1111.

[14] 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;9:1316–23. https://doi.org/10.1128/mcb.9.3.1316.

[15] Tschopp J., Brust P., Cregg J., Stillman C., Gingeras T. Expression of the lacZ gene from two methanol-regulated promoters in Pichia pastoris. Nucleic Acids Res 1987;15:3859–76. https://doi.org/10.1093/nar/15.9.3859.

[16] Cregg JM, Barringer KJ, Hessler AY, Madden KR. Pichia pastoris as a host system for transformations. Mol Cell Biol 1985;5:3376–85. https://doi.org/10.1128/mcb.5.12.3376.

[17] Siegel RS, Brierley RA. Methylotrophic yeast Pichia pastoris produced in high-cell-density fermentations with high cell yields as vehicle for recombinant protein production. Biotechnol Bioeng 1989;34:403–4. https://doi.org/10.1002/bit.260340315.

[18] Kurtzman CP. Description of Komagataella phaffii sp. nov. and the transfer of

Pichia pseudopastoris to the methylotrophic yeast genus Komagataella. Int J Syst Evol Microbiol 2005;55:973–6. https://doi.org/10.1099/ijs.0.63491-0.

[19] 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. (Sacchar omy cet aceae). Biosci Biotechnol Biochem 1995;59:439–44. https://doi.org/10.1271/bbb.59.439.

[20] Sturmberger L, Chappell T, Geier M, Krainer F, Day KJ, Vide U, et al. Refined Pichia pastoris reference genome sequence. J Biotechnol 2016;235:121–31. https://doi.org/10.1016/j.jbiotec.2016.04.023.

[21] 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. https://doi.org/10.1186/s12864-016-2876-y.

[22] Bretthauer RK, Castellino FJ. Glycosylation of Pichia pastoris-derived proteins. Biotechnol Appl Biochem 1999;30:193–200. https://doi.org/10.1111/J.1470-8744.1999.TB00770.X.

[23] Hamilton SR, Gerngross TU. Glycosylation engineering in yeast: the advent of fully humanized yeast. Curr Opin Biotechnol 2007;18:387–92. https://doi.org/10.1016/j.copbio.2007.09.001.

[24] Hamilton SR, Davidson RC, Sethuraman N, Nett JH, Jiang Y, Rios S, et al. Humanization of yeast to produce complex terminally sialylated glycoproteins. Science 2006;313:1441–3. https://doi.org/10.1126/science.1130256.

[25] Jacobs PP, Geysens S, Vervecken W, Contreras R, Callewaert N. Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology. Nat Protoc 2009;4:58–70. https://doi.org/10.1038/nprot.2008.213.

[26] Laukens B, De Wachter C, Callewaert N. Engineering the Pichia pastoris N-glycosylation pathway using the glycoswitch technology. Methods Mol Biol 2015;1321:103–22. https://doi.org/10.1007/978-1-4939-2760-9_8.

[27] Vervecken W, Callewaert N, Kaigorodov V, Geysens S, Contreras R. Modification of the N-glycosylation pathway to produce homogeneous, human-like glycans using GlycoSwitch plasmids. Methods Mol Biol 2007;389:119–38. https://doi.org/10.1385/1-59745-456-7:119.

[28] Holmes WJ, Darby RAJ, Wilks MDB, Smith R, Bill RM. Developing a scalable model of recombinant protein yield from Pichia pastoris: The influence of culture conditions, biomass and induction regime. Microb Cell Fact 2009;8:35. https://doi.org/10.1186/1475-2859-8-35.

[29] Tran AM, Nguyen TT, Nguyen CT, Huynh-Thi XM, Nguyen CT, Trinh MT, et al. Pichia pastoris versus Saccharomyces cerevisiae: a case study on the recombinant production of human granulocyte-macrophage colony-stimulating factor. BMC Res Notes 2017;10:148. https://doi.org/10.1186/s13104-017-2471-6.

[30] Cereghino JL, Cregg JM. Heterologous protein expression in the methylotrophic yeast Pichia pastoris . FEMS Microbiol Rev 2000;24:45–66. https://doi.org/10.1111/j.1574-6976.2000.tb00532.x.

[31] Wegner GH. Emerging applications of the methylotrophic yeasts. FEMS Microbiol Lett 1990;7:279–83. https://doi.org/10.1016/0378-1097(90)90467-5.

[32] Imura M, Nitta K, Iwakiri R, Matsuda F, Shimizu H, Fukusaki E. Comparison of metabolic profiles of yeasts based on the difference of the Crabtree positive and negative. J Biosci Bioeng 2019;129:52–8. https://doi.org/10.1016/j.jbiosc.2019.07.007.

[33] Ata Ö, Rebnegger C, Tatto NE, Valli M, Mairinger T, Hann S, et al. A single Gal4-like transcription factor activates the Crabtree effect in Komagataella phaffii. Nat Commun 2018;9:4911. https://doi.org/10.1038/s41467-018-07430-4.

[34] Burgard J, Gruber-Grünwald C, Altmann F, Zanghellini J, Valli M, Mattanovich D, et al. The secretome of Pichia pastoris in fed-batch cultivations is largely independent of the carbon source but changes quantitatively over cultivation time. Microb Biotechnol 2019;13:479–94. https://doi.org/10.1111/1751-7915.13499.

[35] Smeekens JM, Xiao H, Wu R. Global Analysis of Secreted Proteins and Glycoproteins in Saccharomyces cerevisiae. J Proteome Res 2017;16:1039–49. https://doi.org/10.1021/acs.jproteome.6b00953.

[36] Giardina BJ, Stanley BA, Chiang HL. Glucose induces rapid changes in the secretome of Saccharomyces cerevisiae. Proteome Sci 2014;12:9. https://doi.org/10.1186/1477-5956-12-9.

[37] Huang CJ, Damasceno LM, Anderson KA, Zhang S, Old LJ, Batt CA. A proteomic analysis of the Pichia pastoris secretome in methanol-induced cultures. Appl Microbiol Biotechnol 2011;90:235–47. https://doi.org/10.1007/s00253-011-3118-5.

[38] Mattanovich D, Graf A, Stadlmann J, Dragosits M, Redl A, Maurer M, et al. Genome, secretome and glucose transport highlight unique features of the protein production host Pichia pastoris. Microb Cell Fact 2009;8:29. https://doi.org/10.1186/1475-2859-8-29.

[39] Klein M, Swinnen S, Thevelein JM, Nevoigt E. Glycerol metabolism and transport in yeast and fungi: established knowledge and ambiguities. Environ Microbiol 2017;19:878–93. https://doi.org/10.1111/1462-2920.13617.

[40] Looser V, Bruhlmann B, Bumbak F, Stenger C, Costa M, Camattari A, et al. Cultivation strategies to enhance productivity of Pichia pastoris: A review. Biotechnol Adv 2015;33:1177–93. https://doi.org/10.1016/j.biotechadv.2015.05.008.

[41] Klein M, Islam Z ul, Knudsen PB, Carrillo M, Swinnen S, Workman M, et al. The expression of glycerol facilitators from various yeast species improves growth on glycerol of Saccharomyces cerevisiae. Metab Eng Commun 2016;3:252–7. https://doi.org/10.1016/j.meteno.2016.09.001.

[42] Klein M, Swinnen S, Thevelein JM, Nevoigt E. Glycerol metabolism and transport in yeast and fungi: established knowledge and ambiguities. Environ Microbiol 2017;19:878–93. https://doi.org/10.1111/1462-2920.13617.

[43] Vogl T, Glieder A. Regulation of Pichia pastoris promoters and its consequences

for protein production. N Biotechnol 2013;30:385–404. https://doi.org/10.1016/j.nbt.2012.11.010.

[44] Felber M, Pichler H, Ruth C. Strains and Molecular Tools for Recombinant Protein Production in Pichia pastoris. Yeast Metab. Eng., 2014, p. 87–111.

[45] Zavec D, Gasser B, Mattanovich D. Characterization of methanol utilization negative Pichia pastoris for secreted protein production: New cultivation strategies for current and future applications. Biotechnol Bioeng 2020;117:1394–405. https://doi.org/10.1002/bit.27303.

[46] Lin Cereghino GP, Lin Cereghino J, Jay Sunga A, Johnson MA, Lim M, Gleeson MAG, et al. New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of Pichia pastoris. Gene 2001;263:159–69. https://doi.org/10.1016/S0378-1119(00)00576-X.

[47] Gonçalves AM, Pedro AQ, Maia C, Sousa F, Queiroz JA, Passarinha LA. Pichia pastoris: A recombinant microfactory for antibodies and human membrane proteins. J Microbiol Biotechnol 2013;23:587–601. https://doi.org/10.4014/jmb.1210.10063.

[48] Juturu V, Wu JC. Heterologous Protein Expression in Pichia pastoris: Latest Research Progress and Applications. ChemBioChem 2018;19:7–21. https://doi.org/10.1002/cbic.201700460.

[49] Várnai A, Tang C, Bengtsson O, Atterton A, Mathiesen G, Eijsink VGH. Expression of endoglucanases in Pichia pastoris under control of the GAP promoter. Microb Cell Fact 2014;13:57. https://doi.org/10.1186/1475-2859-13-57.

[50] Gleeson MA, White CE, Meininger DP, Komives EA. Generation of protease-deficient strains and their use in heterologous protein expression. Methods Mol Biol 1998;103:81–94. https://doi.org/10.1385/0-89603-421-6:81.

[51] Lin-Cereghino J, Lin-Cereghino GP. Vectors and Strains for Expression. Methods Mol. Biol., vol. 389, Methods Mol Biol; 2007, p. 11–25. https://doi.org/10.1007/978-1-59745-456-8_2.

[52] 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:e39720. https://doi.org/10.1371/journal.pone.0039720.

[53] Prielhofer R, Barrero JJ, Steuer S, Gassler T, Zahrl R, Baumann K, et al. GoldenPiCS: A Golden Gate-derived modular cloning system for applied synthetic biology in the yeast Pichia pastoris. BMC Syst Biol 2017;11. https://doi.org/10.1186/s12918-017-0492-3.

[54] Vogl T, Glieder A. Regulation of Pichia pastoris promoters and its consequences for protein production. N Biotechnol 2013;30:385–404. https://doi.org/10.1016/j.nbt.2012.11.010.

[55] Türkanoğlu Özçelik A, Yılmaz S, Inan M. Pichia pastoris promoters. Methods Mol. Biol., vol. 1923, Humana Press Inc.; 2019, p. 97–112. https://doi.org/10.1007/978-1-4939-9024-5_3.

[56] Kurjan J, Herskowitz I. Structure of a yeast pheromone gene (MFα): A putative α-factor precursor contains four tandem copies of mature α-factor. Cell 1982;30:933–43. https://doi.org/10.1016/0092-8674(82)90298-7.

[57] Weber E, Engler C, Gruetzner R, Werner S, Marillonnet S. A Modular Cloning System for Standardized Assembly of Multigene Constructs. PLoS One 2011;6:16765. https://doi.org/10.1371/journal.pone.0016765.

[58] Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods 2009;6:343–5. https://doi.org/10.1038/nmeth.1318.

[59] Prielhofer R, Barrero JJ, Steuer S, Gassler T, Zahrl R, Baumann K, et al. GoldenPiCS: a Golden Gate-derived modular cloning system for applied synthetic biology in the yeast Pichia pastoris. BMC Syst Biol 2017;11:123. https://doi.org/10.1186/s12918-017-0492-3.

[60] Vogl T, Sturmberger L, Kickenweiz T, Wasmayer R, Schmid C, Hatzl A-M, et al. A Toolbox of Diverse Promoters Related to Methanol Utilization: Functionally

Verified Parts for Heterologous Pathway Expression in Pichia pastoris. ACS Synth Biol 2016;5:45. https://doi.org/10.1021/acssynbio.5b00199.

[61] Gassler T, Heistinger L, Mattanovich D, Gasser B, Prielhofer R. CRISPR/Cas9-mediated homology-directed genome editing in Pichia pastoris. Methods Mol. Biol., vol. 1923, Humana Press Inc.; 2019, p. 211–25. https://doi.org/10.1007/978-1-4939-9024-5_9.

[62] Liu Q, Shi X, Song L, Liu H, Zhou X, Wang Q, et al. CRISPR-Cas9-mediated genomic multiloci integration in Pichia pastoris. Microb Cell Fact 2019;18. https://doi.org/10.1186/s12934-019-1194-x.

[63] Weninger A, Jasmin |, Fischer E, Raschmanová H, Kniely C, Thomas Vogl |, et al. Expanding the CRISPR/Cas9 toolkit for Pichia pastoris with efficient donor integration and alternative resistance markers; Expanding the CRISPR/Cas9 toolkit for Pichia pastoris with efficient donor integration and alternative resistance markers. J Cell Biochem 2017;119:3183–98. https://doi.org/10.1002/jcb.26474.

[64] Ciofalo V, Barton N, Kreps J, Coats I, Shanahan D. Safety evaluation of a lipase enzyme preparation, expressed in Pichia pastoris, intended for use in the degumming of edible vegetable oil. Regul Toxicol Pharmacol 2006;45:1–8. https://doi.org/10.1016/j.yrtph.2006.02.001.

[65] Thompson CA. FDA approves kallikrein inhibitor to treat hereditary angioedema. Am J Heal Pharm 2010;67:93–93. https://doi.org/10.2146/news100005.

[66] Mattanovich D, Branduardi P, Dato L, Gasser B, Sauer M, Porro D. Recombinant protein production in yeasts. Methods Mol Biol 2012;824:329–58. https://doi.org/10.1007/978-1-61779-433-9_17.

[67] Agarraberes FA, Dice JF. Protein translocation across membranes. Biochim Biophys Acta - Biomembr 2001;1513:1–24. https://doi.org/10.1016/S0304-4157(01)00005-3.

[68] Hou J, Tyo KEJ, Liu Z, Petranovic D, Nielsen J. Metabolic engineering of recombinant protein secretion by Saccharomyces cerevisiae. FEMS Yeast Res

2012;12:491–510. https://doi.org/10.1111/j.1567-1364.2012.00810.x.

[69] Nyathi Y, Wilkinson BM, Pool MR. Co-translational targeting and translocation of proteins to the endoplasmic reticulum. BBA - Mol Cell Res 2013;1833:2392–402. https://doi.org/10.1016/j.bbamcr.2013.02.021.

[70] Agarraberes FA, Dice JF. Protein translocation across membranes. Biochim Biophys Acta - Biomembr 2001;1513:1–24. https://doi.org/10.1016/S0304-4157(01)00005-3.

[71] Shao, Sichen and Hegde R. Membrane Protein Insertion at the Endoplasmic Reticulum. Annu Rev Cell Dev Biol 2011;27:25–56. https://doi.org/https://doi.org/10.1146/annurev-cellbio-092910-154125.

[72] Voorhees RM, Hegde RS. Toward a structural understanding of co-translational protein translocation. Curr Opin Cell Biol 2016;41:91–9. https://doi.org/10.1016/j.ceb.2016.04.009.

[73] Friedman JR, Voeltz GK. The ER in 3D: A multifunctional dynamic membrane network. Trends Cell Biol 2011;21:709–17. https://doi.org/10.1016/j.tcb.2011.07.004.

[74] Austriaco, O. P. N. Endoplasmic reticulum involvement in yeast cell death. Front Oncol 2012;2:87. https://doi.org/10.3389/fonc.2012.00087.

[75] Stevenson J, Huang EY, Olzmann JA. Endoplasmic Reticulum–Associated Degradation and Lipid Homeostasis. Annu Rev Nutr 2016;36:511–42. https://doi.org/10.1146/annurev-nutr-071715-051030.

[76] Braakman I, Bulleid NJ. Protein Folding and Modification in the Mammalian Endoplasmic Reticulum. Annu Rev Biochem 2011;80:71–99. https://doi.org/10.1146/annurev-biochem-062209-093836.

[77] Tannous A, Brambilla Pisoni G, Hebert DN, Molinari M. N-linked sugar-regulated protein folding and quality control in the ER. Semin Cell Dev Biol 2015;41:79–89. https://doi.org/10.1016/j.semcdb.2014.12.001.

[78] de Ruijter JC, Koskela E V., Frey AD. Enhancing antibody folding and secretion by

tailoring the Saccharomyces cerevisiae endoplasmic reticulum. Microb Cell Fact 2016;15:87. https://doi.org/10.1186/s12934-016-0488-5.

[79] Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS, Walter P. Functional and Genomic Analyses Reveal an Essential Coordination between the Unfolded Protein Response and ER-Associated Degradation. Cell 2000;101:249–58. https://doi.org/10.1016/s0092-8674(00)80835-1.

[80] Delic M, Rebnegger C, Wanka F, Puxbaum V, Haberhauer-Troyer C, Hann S, et al. Oxidative protein folding and unfolded protein response elicit differing redox regulation in endoplasmic reticulum and cytosol of yeast. Free Radic Biol Med 2012;52:2000–12. https://doi.org/10.1016/j.freeradbiomed.2012.02.048.

[81] Bertolotti A, Zhang Y, Hendershot LM, Harding HP, Ron D. Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. Nat Cell Biol 2000;2:326–32. https://doi.org/10.1038/35014014.

[82] Gardner BM, Walter P. Unfolded Proteins are Ire1-Activating Ligands that Directly Induce the Unfolded Protein Response. Science 2011;333:1891–4. https://doi.org/10.1126/science.1209126.

[83] Young CL, Robinson AS. Protein Folding and Secretion: Mechanistic Insights Advancing Recombinant Protein Production in S. cerevisiae. Curr Opin Biotechnol 2014;30:168–77. https://doi.org/10.1038/jid.2014.371.

[84] Zahrl RJ, Mattanovich D, Gasser B. The impact of ERAD on recombinant protein secretion in Pichia pastoris (syn Komagataella spp.). Microbiology 2018;164:453–63. https://doi.org/10.1099/mic.0.000630.

[85] Baldridge RD, Rapoport TA. Autoubiquitination of the Hrd1 Ligase Triggers Protein Retrotranslocation in ERAD. Cell 2016;166:394–407. https://doi.org/10.1016/j.cell.2016.05.048.

[86] Luen Tang B, Wang Y, Ong YS, Hong W. COPII and exit from the endoplasmic reticulum. Biochim Biophys Acta 2005;1744:293–303. https://doi.org/10.1016/j.bbamcr.2005.02.007.

[87] Dancourt J, Barlowe C. Protein Sorting Receptors in the Early Secretory Pathway. Annu Rev Biochem 2010;79:777–802. https://doi.org/10.1146/annurev-biochem-061608-091319.

[88] Otte S, Barlowe C. Sorting signals can direct receptor-mediated export of soluble proteins into COPII vesicles. Nat Cell Biol 2004;6:1189–94. https://doi.org/10.1038/ncb1195.

[89] Yin Y, Garcia MR, Novak AJ, Saunders AM, Ank RS, Nam AS, et al. Surf4 (Erv29p) binds amino-terminal tripeptide motifs of soluble cargo proteins with different affinities, enabling prioritization of their exit from the endoplasmic reticulum. PLoS Biol 2018;16:e2005140. https://doi.org/10.1371/journal.pbio.2005140.

[90] Besada-Lombana PB, Da Silva NA. Engineering the early secretory pathway for increased protein secretion in Saccharomyces cerevisiae. Metab Eng 2019;55:142–51. https://doi.org/10.1016/j.ymben.2019.06.010.

[91] Papanikou E, Glick BS. Golgi compartmentation and identity. Curr Opin Cell Biol 2014;29:74–81. https://doi.org/10.1016/j.ceb.2014.04.010.

[92] Pfeffer SR. How the Golgi works: A cisternal progenitor model. Proc Natl Acad Sci U S A 2010;107:19614–8. https://doi.org/10.1073/pnas.1011016107.

[93] Casler JC, Papanikou E, Barrero JJ, Glick BS. Maturation-driven transport and AP-1-dependent recycling of a secretory cargo in the Golgi. J Cell Biol 2019;218:1582–601. https://doi.org/10.1083/jcb.201807195.

[94] Day KJ, Casler JC, Glick BS. Budding Yeast Has a Minimal Endomembrane System. Dev Cell 2018;44:56–72. https://doi.org/10.1016/j.devcel.2017.12.014.

[95] Fitzgerald I, Glick BS. Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting. Microb Cell Fact 2014;13:125. https://doi.org/10.1186/s12934-014-0125-0.

[96] Fuller RS, Sterne RE, Thorner J. Enzymes required for yeast prohormone processing. Annu Rev Physiol 1988;50:345–62.

https://doi.org/10.1146/annurev.physiol.50.1.345.

[97] Puxbaum V, Mattanovich D, Gasser B. Quo vadis? The challenges of recombinant protein folding and secretion in Pichia pastoris. Appl Microbiol Biotechnol 2015;99:2925–38. https://doi.org/10.1007/s00253-015-6470-z.

[98] Gasser B, Maurer M, Gach J, Kunert R, Mattanovich D. Engineering of Pichia pastoris for improved production of antibody fragments. Biotechnol Bioeng 2006;94:353–61. https://doi.org/10.1002/bit.20851.

[99] Gasser B, Sauer M, Maurer M, Stadlmayr G, Mattanovich D. Transcriptomics-Based Identification of Novel Factors Enhancing Heterologous Protein Secretion in Yeasts. Appl Environ Microbiol 2007;73:6499–507. https://doi.org/10.1128/AEM.01196-07.

[100] Damasceno LM, Anderson KA, Ritter G, Cregg JM, Old LJ, Batt CA. Cooverexpression of chaperones for enhanced secretion of a single-chain antibody fragment in Pichia pastoris. Appl Microbiol Biotechnol 2007;74:381–9. https://doi.org/10.1007/s00253-006-0652-7.

[101] Liu X, Wu D, Wu J, Chen J. Optimization of the production of Aspergillus niger a-glucosidase expressed in Pichia pastoris. World J Microbiol Biotechnol 2013;29:533–40. https://doi.org/10.1007/s11274-012-1207-y.

[102] Zhang W, Zhao HL, Xue C, Xiong XH, Yao XQ, Li XY, et al. Enhanced secretion of heterologous proteins in Pichia pastoris following overexpression of Saccharomyces cerevisiae chaperone proteins. Biotechnol Prog 2006;22:1090–5. https://doi.org/10.1021/bp060019r.

[103] Samuel P, Prasanna Vadhana AK, Kamatchi R, Antony A, Meenakshisundaram S. Effect of molecular chaperones on the expression of Candida antarctica lipase B in Pichia pastoris. Microbiol Res 2013;168:615–20. https://doi.org/10.1016/j.micres.2013.06.007.

[104] Gasser B, Maurer M, Rautio J, Sauer M, Bhattacharyya A, Saloheimo M, et al. Monitoring of transcriptional regulation in Pichia pastoris under protein production conditions. BMC Genomics 2007;8. https://doi.org/10.1186/1471-

2164-8-179.

[105] Guerfal M, Ryckaert S, Jacobs PP, Ameloot P, Van Craenenbroeck K, Derycke R, et al. The HAC1 gene from Pichia pastoris: Characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins. Microb Cell Fact 2010;9:49. https://doi.org/10.1186/1475-2859-9-49.

[106] Lin X-Q, Liang S-L, Han S-Y, Zheng S-P, Ye Y-R, Lin Y. Quantitative iTRAQ LC-MS/MS proteomics reveals the cellular response to heterologous protein overexpression and the regulation of HAC1 in Pichia pastoris. J Proteomics 2013;91:58–72. https://doi.org/10.1016/j.jprot.2013.06.031.

[107] De Ruijter JC, Koskela E V, Frey AD. Enhancing antibody folding and secretion by tailoring the Saccharomyces cerevisiae endoplasmic reticulum. Microb Cell Fact 2016;15:87. https://doi.org/10.1186/s12934-016-0488-5.

[108] Yang S, Kuang Y, Li H, Liu Y, Hui X, Li P, et al. Enhanced Production of Recombinant Secretory Proteins in Pichia pastoris by Optimizing Kex2 P1' site. PLoS One 2013;8:e75347. https://doi.org/10.1371/journal.pone.0075347.

[109] Rakestraw JA, Sazinsky SL, Piatesi A, Antipov E, Wittrup KD. Directed evolution of a secretory leader for the improved expression of heterologous proteins and full-length antibodies in Saccharomyces cerevisiae. Biotechnol Bioeng 2009;103:1192–201. https://doi.org/10.1002/bit.22338.

[110] Brake AJ, Merryweather JP, Coit DG, Heberlein UA, Masiarz FR, Mullenbach GT, et al. Alpha-factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 1984;81:4642–6. https://doi.org/10.1073/pnas.81.15.4642.

[111] Julius D, Blair L, Brake A, Sprague G, Thorner J. Yeast α factor is processed from a larger precursor polypeptide: The essential role of a membrane-bound dipeptidyl aminopeptidase. Cell 1983;32:839–52. https://doi.org/10.1016/0092-8674(83)90070-3.

[112] Vedvick T, Buckholz RG, Engel M, Urcan M, Kinney J, Provow S, et al. High-level secretion of biologically active aprotinin from the yeast Pichia pastoris. J Ind

Microbiol 1991;7:197–201. https://doi.org/10.1007/BF01575883.

[113] Ng DTW, Brown JD, Walter P. Signal sequences specify the targeting route to the endoplasmic reticulum membrane. J Cell Biol 1996;134:269–78. https://doi.org/10.1083/jcb.134.2.269.

[114] Yim C, Jung S jun, Kim JEH, Jung Y, Jeong SD, Kim H. Profiling of signal sequence characteristics and requirement of different translocation components. Biochim Biophys Acta - Mol Cell Res 2018;1865:1640–8. https://doi.org/10.1016/j.bbamcr.2018.08.018.

[115] Fuller RS, Sterne RE, Thorner J. Enzymes required for yeast prohormone processing. Annu Rev Physiol 1988;50:345–62. https://doi.org/10.1146/annurev.physiol.50.1.345.

[116] Chahal S, Wei P, Moua P, Pil S, Park J, Kwon J, et al. Structural characterization of the α-mating factor prepro-peptide for secretion of recombinant proteins in Pichia pastoris. Gene 2017;598:50–62. https://doi.org/10.1016/j.gene.2016.10.040.

[117] Lin-Cereghino GP, Stark CM, Kim D, Chang J, Shaheen N, Poerwanto H, et al. The effect of α-mating factor secretion signal mutations on recombinant protein expression in Pichia pastoris. Gene 2013;519:311–7. https://doi.org/10.1016/j.gene.2013.01.062.

[118] Mateljak I, Tron T, Alcalde M. Brief report Evolved a-factor prepro-leaders for directed laccase evolution in Saccharomyces cerevisiae. Microb Biotechnol 2017;10:1830–6. https://doi.org/10.1111/1751-7915.12838.

[119] Ahn Min-Jung Jang Kok Siong Ang Hongweon Lee Eui-Sung Choi Dong-Yup Lee J, Ahn Á M-J Jang Á Lee Á E-S Choi JH, Jang M, Lee H, Choi E, Ahn Á Lee Á E-S Choi JH, et al. Codon optimization of Saccharomyces cerevisiae mating factor alpha prepro-leader to improve recombinant protein production in Pichia pastoris. Biotechnol Lett 2016;38:2137–43. https://doi.org/10.1007/s10529-016-2203-3.

[120] Kjeldsen T. Yeast secretory expression of insulin precursors. vol. 54. Springer-Verlag; 2000.

[121] Kajiwara S, Yamada R, Ogino H. Secretory Overexpression of Bacillus thermocatenulatus Lipase in Saccharomyces cerevisiae Using Combinatorial Library Strategy. Biotechnol J 2018;13:e1700409. https://doi.org/10.1002/biot.201700409.

[122] Liang S, Li C, Ye Y, Lin Y. Endogenous signal peptides efficiently mediate the secretion of recombinant proteins in Pichia pastoris. Biotechnol Lett 2013;35:97–105. https://doi.org/10.1007/s10529-012-1055-8.

[123] Aw R, Mckay PF, Shattock RJ, Polizzi KM. A systematic analysis of the expression of the anti-HIV VRC01 antibody in Pichia pastoris through signal peptide optimization. Protein Expr Purif 2018;149:43–50. https://doi.org/10.1016/j.pep.2018.03.013.

[124] Massahi A, Çalık P. In-silico determination of Pichia pastoris signal peptides for extracellular recombinant protein production. J Theor Biol 2015;364:179–88. https://doi.org/10.1016/j.jtbi.2014.08.048.

[125] García-Ortega X, Cámara E, Ferrer P, Albiol J, Luis Montesinos-Seguí J, 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? N Biotechnol 2019;53:24–34. https://doi.org/10.1016/j.nbt.2019.06.002.

[126] 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. Microb Cell Fact 2006;5:17. https://doi.org/10.1186/1475-2859-5-17.

[127] Dragosits M, Stadlmann J, Albiol J, Baumann K, Maurer M, Gasser B, et al. The Effect of Temperature on the Proteome of Recombinant Pichia pastoris. J Proteome Res 2009;8. https://doi.org/10.1021/pr8007623.

[128] Jahic M, Wallberg F, Bollok M, Garcia P, Enfors S-O. Temperature limited fed-batch technique for control of proteolysis in Pichia pastoris bioreactor cultures. Microb Cell Fact 2003;2:6.

[129] Cregg JM, Vedvick TS, Raschke WC. Recent Advances in the Expression of Foreign Genes in Pichia pastoris. Bio/Technology 1993;11:905–10. https://doi.org/10.1038/nbt0893-905.

[130] Sinha J, Plantz BA, Inan M, Meagher MM. Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeastPichia pastoris: Case study with recombinant ovine interferon-Tau. Biotechnol Bioeng 2005;89:102–12. https://doi.org/10.1002/bit.20318.

[131] Zapadka KL, Becher FJ, Gomes dos Santos AL, Jackson SE. Factors affecting the physical stability (aggregation) of peptide therapeutics. Interface Focus 2017;7. https://doi.org/10.1098/rsfs.2017.0030.

[132] Lee CY, Lee SJ, Jung KH, Katoh S, Lee EK. High dissolved oxygen tension enhances heterologous protein expression by recombinant Pichia pastoris. Process Biochem 2003;38:1147–54. https://doi.org/10.1016/S0032-9592(02)00280-7.

[133] Ponte X, Luis Montesinos-Seguí J, Valero F. Bioprocess efficiency in Rhizopus oryzae lipase production by Pichia pastoris under the control of PAOX1 is oxygen tension dependent. Process Biochem 2016;51:1954–63. https://doi.org/10.1016/j.procbio.2016.08.030.

[134] Vandermies M, Fickers P. Bioreactor-scale strategies for the production of recombinant protein in the yeast yarrowia lipolytica. Microorganisms 2019;7:2. https://doi.org/10.3390/microorganisms7020040.

[135] Valero F. Bioprocess Engineering of Pichia pastoris, an Exciting Host Eukaryotic Cell Expression System. Protein Eng - Technol Appl 2013. https://doi.org/10.5772/56407.

[136] Invitrogen, Thermofisher. Pichia fermentation process guidelines, available online: http://tools.thermofisher.com/content/sfs/manuals/pichiaferm_prot.pdf (Accessed 29th March 2020) 2020.

[137] Zhang W, Inan M, Meagher MM. Rational design and optimization of fed-batch and continuous fermentations. Methods Mol Biol 2007;389:43–63. https://doi.org/10.1385/1-59745-456-7:43.

[138] Nieto-Taype MÁ, Garrigós J, Sánchez M, Valero F, Montesinos-Seguí JL, García -Ortega X. Rational-based selection of optimal operating strategies and gene dosage impact on recombinant protein production in Komagataella phafii (Pichia pastoris) 2019;13:315–27. https://doi.org/10.1111/1751-7915.13498.

[139] Garrigós-Martínez J, Nieto-Taype MA, Gasset-Franch A, Montesinos-Seguí JL, García-Ortega X, Valero F. Specific growth rate governs AOX1 gene expression, affecting the production kinetics of Pichia pastoris (Komagataella phaffii) P AOX1 -driven recombinant producer strains with different target gene dosage. Microb Cell Fact 2019;18:187. https://doi.org/10.1186/s12934-019-1240-8.

[140] López-Fernández J, Barrero JJ, Benaiges MD, Valero F. Truncated Prosequence of Rhizopus oryzae Lipase : Key Factor for Production Improvement and Biocatalyst Stability. Catalysts 2019;9:961. https://doi.org/10.3390/cata9110961.

[141] Barrigón JM, Montesinos JL, Valero F. Searching the best operational strategies for Rhizopus oryzae lipase production in Pichia pastoris Mut + phenotype: Methanol limited or methanol non-limited fed-batch cultures? Biochem Eng J 2013;75:47–54. https://doi.org/10.1016/j.bej.2013.03.018.

[142] Cos O, Ramon R, Montesinos L, Valero F. A Simple Model-Based Control for Pichia pastoris Allows a More Efficient Heterologous Protein Production Bioprocess. Biotechnol Bioeng 2006;95:145–54. https://doi.org/10.1002/bit.21005.

[143] Gassler T, Heistinger L, Mattanovich D, Gasser B, Prielhofer R. CRISPR/Cas9-mediated homology-directed genome editing in Pichia pastoris. Methods Mol Biol 2019;1923:211–25. https://doi.org/10.1007/978-1-4939-9024-5_9.

[144] De J, Robert M, Garcia-Ortega X, Luis Montesinos-Seguí J, Maria D, Freire G, et al. Continuous operation, a realistic alternative to fed-batch fermentation for the production of recombinant lipase B from Candida antarctica under the constitutive promoter PGK in Pichia pastoris. Biochem Eng J 2019;147:39–47. https://doi.org/10.1016/j.bej.2019.03.027.

[145] Cámara E, Monforte S, Albiol J, Ferrer P. Deregulation of methanol metabolism reverts transcriptional limitations of recombinant Pichia pastoris (Komagataella

spp) with multiple expression cassettes under control of the AOX1 promoter. Biotechnol Bioeng 2019;116:1710–20. https://doi.org/10.1002/bit.26947.

[146] Strack RL, Hein B, Bhattacharyya D, Hell SW, Keenan RJ, Glick BS. A rapidly maturing far-red derivative of DsRed-Express2 for whole-cell labeling. Biochemistry 2009;48:8279–81. https://doi.org/10.1021/bi900870u.

[147] Strack RL, Keenan RJ, Glick BS. Noncytotoxic DsRed derivatives for whole-cell labeling. Methods Mol Biol 2011;699:355–70. https://doi.org/10.1007/978-1-61737-950-5_17.

[148] Barrero JJ, Papanikou E, Casler JC, Day KJ, Glick BS. An improved reversibly dimerizing mutant of the FK506-binding protein FKBP. Cell Logist 2016;6:8. https://doi.org/10.1080/21592799.2016.1204848.

[149] Surribas A, Resina D, Ferrer P, Valero F. Rivoflavin may interfere with on-line monitoring of secreted green fluorescence protein fusion proteins in Pichia pastoris. Microb Cell Fact 2007;6:15. https://doi.org/10.1186/1475-2859-6-15.

[150] Hasan F, Shah AA, Hameed A. Methods for detection and characterization of lipases: A comprehensive review. Biotechnol Adv 2009;27:782–98. https://doi.org/10.1016/j.biotechadv.2009.06.001.

[151] Guillén M, Benaiges MD, Valero F. Immobilization and stability of a Rhizopus oryzae lipase expressed in Pichia pastoris: Comparison between native and recombinant variants. Biotechnol Prog 2011;27:1232–41. https://doi.org/10.1002/btpr.654.

[152] Resina D, Maurer M, Cos O, Arnau C, Carnicer M, Marx H, et al. Engineering of bottlenecks in Rhizopus oryzae lipase production in Pichia pastoris using the nitrogen source-regulated FLD1 promoter. N Biotechnol 2009;25:396–403. https://doi.org/10.1016/j.nbt.2009.01.008.

[153] Resina D, Bollók M, Khatri NK, Valero F, Neubauer P, Ferrer P. Transcriptional response of P. pastoris in fed-batch cultivations to Rhizopus oryzae lipase production reveals UPR induction. Microb Cell Fact 2007;6:21. https://doi.org/10.1186/1475-2859-6-21.

[154] Rúa ML, Schmidt-Dannert C, Wahl S, Sprauer A, Schmid RD. Thermoalkalophilic lipase of Bacillus thermocatenulatus. J Biotechnol 1997;56:89–102. https://doi.org/10.1016/S0168-1656(97)00079-5.

[155] Rúa ML, Atomi H, Schmidt-Dannert C, Schmid RD. High-level expression of the thermoalkalophilic lipase from Bacillus thermocatenulatus in Escherichia coli. Appl Microbiol Biotechnol 1998;49:405–10. https://doi.org/10.1007/s002530051190.

[156] Rúa ML, Schmidt-Dannert C, Wahl S, Sprauer A, Schmid RD. Thermoalkalophilic lipase of Bacillus thermocatenulatus: Large-scale production, purification and properties: Aggregation behaviour and its effect on activity. J Biotechnol 1997;56:89–102. https://doi.org/10.1016/S0168-1656(97)00079-5.

[157] Quyen DT, Schmidt-Dannert C, Schmid RD. High-level expression of a lipase from Bacillus thermocatenulatus BTL2 in Pichia pastoris and some properties of the recombinant lipase. Protein Expr Purif 2003;28:102–10. https://doi.org/10.1016/S1046-5928(02)00679-4.

2

Objectives

The main objective of this work has been the development and validation of a new secretion signal sequence capable of overcoming the limitations shown by the conventional secretion signals commonly used in the methylotrophic yeast *Pichia pastoris*. In addition, strains producing model proteins with the novel secretion signal have been characterized at a transcriptional level in fed-batch bioreactor cultures to further identify different limitations present in the secretion pathway when producing recombinant proteins with this new secretion signal.

In order to achieve this goal, different specific objectives were defined:

- Construction and design of new secretion signals with potential improvements in comparison with the conventional system.
- Find a fast and relievable system to characterize new secretion signals and the effect this signal is making to the cells.
- Find a robust system to quantify extracellular and intracellular (if possible) production levels for different model proteins.
- Strain characterization using cultures at different scale-level to study the effect of the new secretion signal in recombinant protein production and cell physiology.
- Transcriptional study of the new strains aiming at finding physiological insights derived from the use of the novel secretion signal.
- Interpretation of the transcriptional data sets as a basis for further strain optimization to boost protein production.

3

Overview of the study

In the last decades, heterologous protein production has been of great interest to produce several proteins with industrial and pharmaceutical purposes. Several cellular platforms have been created to obtain and maximize the amount of produced protein. Over all the expression organisms used as hosts for recombinant protein production, the yeast *Pichia pastoris* is an emerging cell platform with great interest. Its popularity has arisen due to several benefits including the capacity to grow at high cell densities at a bioreactor scale, the low cost of the culture media, and the availability of several inducible promoters that enable the decoupling of growth and protein production. In addition, this cell platform has also the capacity to secrete recombinant proteins out of the cells through the secretory pathway, which facilitates posterior protein purification. However, the capacity from the secretory pathway can be overwhelmed under the production of heterologous proteins yielding in an accumulation of intracellular protein in different compartments from the secretory pathway.

This work focuses on pinpointing those bottlenecks and limitations present at the secretory pathway with a special attention to a bottleneck that had been identified but never properly characterized before this study, particularly in *P. pastoris*, i.e. the translocation step. The limitation present at the translocation step was easily solved by designing and constructing a secretion signal that efficiently translocated model proteins into the secretory pathway. In addition, once the bottleneck in the translocation step was solved, other bottlenecks have been identified downstream the secretion pathway and new strategies to overcome these limitations have been proposed.

Chapter 4 focuses on a knowledge-based selection and characterization of a new secretion signal capable of overcoming the limitations from the current secretion signal, the α-mating factor from *Saccharomyces cerevisiae* (α-MF). For this purpose, the far-red fluorescent model protein E2-Crimson was used, as it can be tracked down along the secretion pathway and allows the identification of compartments where it can be accumulated. The designed hybrid secretion signal was the combination of the pre-Ost1 signal sequence and the pro region from α-MF with a Ser42 mutation. This secretion signal allowed the efficient translocation through the co-translational translocation of E2-Crimson and significantly reduced the aggregation of E2-Crimson at the ER lumen,

resulting in a boost of protein secretion in a flask-scale. Furthermore, the secretion signal variants obtained for E2-Crimson were also tested with a lipase called BTL2 and similar results were obtained. Overall, the preliminary evaluation of the new secretion signal in small-scale (shake-flask) experiments yielded very promising data. This prompted us to validate in Chapter 5 this novel tool at a bioreactor scale, under bioprocess-like conditions, to confirm its potential.

This study has been published as:

Juan J. Barrero, Jason C. Casler, Francisco Valero, Pau Ferrer and Benjamin S. Glick. (2018) An improved secretion signal enhances the secretion of model proteins from *Pichia pastoris.* Microbial Cell Factories. 17 (161).

Contributions:

I contributed to the study design, performed the bulk of the experimental work, data analysis, and interpretation, and drafted the manuscript. Jason C. Casler performed the immunoblot assay. Francisco Valero contributed to the design and characterization of the activity assay for BTL2, and revised the manuscript. Pau Ferrer and Benjamin S. Glick contributed to the study design, data analysis, and interpretation, and revised the manuscript.

In **chapter 5**, the performance of the improved secretion signal was studied at bioreactor scale with the model proteins from chapter 4 as well as an additional lipase called ROL, which is structurally different than the previous model proteins. As in chapter 4, the α-MF secretion signal was also used to compare the results seen with the improved secretion signal. All the strains were studied in bioreactor cultures operated in batch and fed-batch. Notably, the producing strains with the improved secretion signal showed a relative increase in terms of final protein concentration similar to what had been observed at shake-flask scale, as well as improved production yields. More importantly, cells with the new secretion signal showed an increased cell performance, which lead to an increased growth rate, allowing for increased overall productivities in batch cultures, as cells reached the end of the fermentation faster. Although the levels of protein production were increased with the improved secretion signal, new bottlenecks might appear downstream the translocation step which might decrease the full potential of the new secretion signal. In chapter 6, a transcriptional analysis was performed to identify new potential bottlenecks appearing after the translocation step.

This study has been submitted as:

Juan J. Barrero, Alejandro Pagazartaundua, Benjamin S. Glick, Francisco Valero and Pau Ferrer. (2020) Bioreactor-scale cell performance and protein production can be substantially increased by using a secretion signal that drives co-translational translocation in *Pichia pastoris*. (Manuscript under review in New Biotechnology)

Contributions:

I contributed to the study design, performed the bulk of the experimental work, data analysis, and interpretation, and drafted the manuscript. Alejandro Pagazartaundua was a master student working under my supervision and assisted in the E2-Crimson fed-batches as well as the cytometer experiments during his internship. Francisco Valero, Pau Ferrer and Benjamin S. Glick contributed to the study design, data analysis, and interpretation, and revised the manuscript.

Finally, in **chapter 6,** we performed a transcriptional analysis to study new potential limitations appearing downstream the secretion pathway once the initial translocation bottleneck is no longer present. This study was carried out in a fed-batch process using the strains expressing E2-Crimson with either the improved secretion signal or the α-MF. The results show a clear accumulation of E2-Crimson at the ER lumen, particularly after a few hours from the start of induction. This accumulation lead to an overexpression of genes related with the Unfolded Protein Response (UPR) and, consequently, genes related with the ER-Associated Degradation (ERAD) were overexpressed as well, which probably triggered the intracellular degradation of the accumulated protein. With this study, some cell engineering strategies have been proposed that, in combination with the improved secretion signal, could maximize production yields of recombinant proteins.

Juan J. Barrero, Francisco Valero and Pau Ferrer. (2021) A transcriptional analysis reveals potential targets to increase protein secretion in a fed-batch cultivation for the yeast *Pichia pastoris*. (This study will be part of a future publication together with additional experimental work to be carried out)

Contributions:

I contributed to the study design, performed all the experimental work, data analysis, and interpretation. Francisco Valero and Pau Ferrer contributed to the study design, data analysis, and interpretation.

4

An improved secretion signal enhances the secretion of model proteins from *Pichia pastoris*

Keywords: Translocation, Secretion, *Pichia pastoris*, Alpha-factor, Ost1, Heterologous protein production, Aggregation

Table of Contents

1. Summary

Proteins can be secreted from a host organism with the aid of N-terminal secretion signals. For *Pichia pastoris* (*Komagataella* spp.), the most commonly used secretion signal is the N-terminal portion of pre-pro-α-factor from *Saccharomyces cerevisiae*. However, this secretion signal promotes posttranslational translocation into the endoplasmic reticulum (ER), so proteins that can fold in the cytosol may be inefficiently translocated and thus poorly secreted. In addition, if a protein self-associates, the α-factor pro region can potentially cause aggregation, thereby hampering export from the ER. This study addresses both limitations of the pre-pro-α-factor secretion signal.

We engineered a hybrid secretion signal consisting of the *S. cerevisiae* Ost1 signal sequence, which promotes cotranslational translocation into the ER, followed by the alpha-factor pro region. Secretion and intracellular localization were assessed using as a model protein the tetrameric red fluorescent protein E2-Crimson. When paired with the α-factor pro region, the Ost1 signal sequence yielded much more efficient secretion than the α-factor signal sequence. Moreover, an allelic variant of the α-factor pro region reduced aggregation of the E2-Crimson construct in the ER. The resulting improved secretion signal enhanced secretion of E2-Crimson up to 20-fold compared to the levels obtained with the original α-factor secretion signal. Similar findings were obtained with the lipase BTL2, which exhibited 10-fold enhanced secretion with the improved secretion signal.

The improved secretion signal confers dramatic benefits for the secretion of certain proteins from *P. pastoris*. These benefits are likely to be most evident for proteins that can fold in the cytosol and for oligomeric proteins.

2. Introduction

For heterologous protein production in *P. pastoris*, the most common secretion signal is that of the *S. cerevisiae* α-factor mating pheromone [1]. This secretion signal consists of two parts: a 19-amino acid N-terminal signal sequence that directs translocation into the endoplasmic reticulum (ER), followed by a 66-amino acid pro region that mediates receptor-dependent packaging into ER-derived COPII transport vesicles (Figure S1) [2,3]. The α-factor signal sequence is removed by a signal peptidase in the ER lumen, and the pro region is cleaved by the Kex2 processing protease in the Golgi [4,5]. This bipartite secretion signal has proven to be effective for secreting multiple heterologous proteins in *P. pastoris*, but the level of secretion varies widely, prompting efforts to improve secretion efficiency [6–11]. Several of those attempts focused on modifying the α-factor secretion signal [12–15].

A limitation of the α-factor secretion signal is that the signal sequence portion directs posttranslational translocation across the ER membrane [16,17]. As a result, if the α-factor secretion signal is fused to a protein that can fold in the yeast cytosol, the protein may be unable to traverse the ER translocon and enter the secretory pathway. We encountered this problem when monitoring secretion in *S. cerevisiae* or *P. pastoris* using a monomeric superfolder GFP (msGFP) [18]. The solution was to replace the α-factor signal sequence with the Ost1 signal sequence, which directs cotranslational translocation across the ER membrane, thereby ensuring that msGFP folds only after reaching the ER lumen [19,20].

Here, we have extended this analysis by using the tetrameric far-red fluorescent protein E2-Crimson as a model for heterologous protein secretion in *P. pastoris*. E2-Crimson folds and oligomerizes efficiently, and it acquires fluorescence rapidly [21]. The fluorescence signal provides a convenient way to visualize potential roadblocks in the secretory pathway using fluorescence microscopy. There are advantages to using E2-Crimson instead of further investigating secretion of msGFP. E2-Crimson is oligomeric, so the data are complementary to those obtained with the monomeric msGFP. Moreover, E2-Crimson fluoresces at red wavelengths, so we can track secretion without

interference from the green fluorescence of riboflavin, a yeast culture medium component that is produced at high levels by methanol-grown *P. pastoris* cells [22,23].

Our results indicate that E2-Crimson can become trapped along the secretory pathway in two ways. First, when the α-factor signal sequence is used, E2-Crimson fails to cross the ER membrane, presumably because the protein folds prior to posttranslational translocation. As with msGFP, this problem can be overcome by using the Ost1 signal sequence. Second, when fused to a commonly used variant of the α-factor pro region, E2-Crimson aggregates in the ER lumen, presumably because the pro region has a self-association tendency that is amplified by the oligomeric nature of E2-Crimson. This problem can be overcome with an allelic variant in which a single amino acid difference in the pro region suppresses aggregation. Combining the two modifications yielded an improved secretion signal that drives highly efficient secretion of E2-Crimson.

An important question is whether these improvements extend beyond the model fluorescent proteins. As a case study, we chose the BTL2 lipase from *Bacillus thermocatenulatus* [24,25]. Lipases are of major industrial value [26], and BTL2 is promising because it is thermostable as well as catalytically active at high pH and in the presence of organic solvents [27,28]. *P. pastoris* has been engineered to secrete BTL2 [29]. A recent study of BTL2 secretion from *S. cerevisiae* showed that the choice of secretion signal was particularly important [30], hinting that BTL2 might be prone to folding prior to translocation. In support of this idea, use of the improved secretion signal in *P. pastoris* strongly enhances BTL2 secretion. This finding suggests that the improved secretion signal will be broadly useful.

3. Materials and methods

3.1. Strains and plasmids

P. pastoris strains were derivatives of X33 (Thermo Fisher Invitrogen). All strains were selected and grown in rich medium (YPD) supplemented with either Zeocin (100 µg/mL), hygromycin (250 µg/mL), or G418 (500 µg/mL) depending on the integrated plasmid. Buffered minimal glycerol (BMG) and buffered minimal methanol (BMM) media recipes were taken from the instruction manual for Thermo Fisher Invitrogen's *Pichia* Expression Kit.

Plasmids were created and modified by standard methods including site-directed mutagenesis [31] and In-Fusion cloning (TaKaRa/Clontech). Primers were purchased from IDT. The gene encoding BTL2 was codon-optimized for *P. pastoris* by GenScript. Expression of E2-Crimson and BTL2 were driven by the inducible *AOX1* promoter, while expression of msGFP-HDEL and Htb2-iGFP were driven by the *KAR2* and *GAP* promoters, respectively. Genetic engineering procedures were designed and recorded using SnapGene software (GSL Biotech). The supplementary information contains a compressed folder of SnapGene files for the plasmids used in this study (Additional File 1), and those files can be opened with the free SnapGene Viewer (www.snapgene.com/products/snapgene_viewer). Key plasmids will be deposited with Addgene.

Plasmids were linearized and then transformed by electroporation [32] using 100 ng of linear DNA, an amount that limited the number of copies integrated. Single-copy integration of E2-Crimson and BTL2 constructs at the *AOX1* locus were verified using primers 5'-GAAATAGACGCAGATCGGGAAC-3' and 5'-GAAGGTAGACCCATGGGTTGTTG-3'. The pre-Kar2-msGFP-HDEL construct was integrated at the *HIS4* locus, and single copy integration was verified using primers 5'-GCTCTAGCCAGTTTGCTGTCCAAAC-3' and 5'-GGATGTTAGATGCCGGTTAGATC-3'. The Htb2-iGFP construct was integrated at the *GAP* locus, and single-copy integration was verified using primers 5'-GATGACAATGGACCAAATTGTTGCAAGG-3' and 5'-CCGTTAATACCGACAGTGATAGCC-3'. Additionally, for strains with BTL2 constructs, droplet digital PCR [33] was performed to ensure single-copy integration using primers 5'-GGGTATGAACGCTTTTTCTGCTGTTG-3'

and 5'-GATCAACGTTACAAGTACCCATATCATTCC-3' for the BTL2 gene, or 5'-CCTGAGGCTTTGTTCCACCCATCT-3' and 5'-GGAACATAGTAGTACCACCGGACATAACGA-3' for the actin gene as a control.

3.2. Assaying secretion of E2-Crimson

From each strain, eight transformed colonies were streaked on YPD plates containing the appropriate antibiotic and then re-streaked twice on new plates to avoid mixed cell populations. After identifying single-copy integrants for each strain, a pre-screening of six clones to assess the level of E2-Crimson secretion (Figure S2) was used to identify two representative clones. Further analysis was performed in parallel with these two clones, which were stored frozen at -80°C. Clones were retrieved from the frozen stocks to make saturated YPD precultures that were kept for up to 2-3 weeks at 4°C.

For a given strain, each of the two clones was analyzed in triplicate as follows. The E2-Crimson secretion assay was initiated by inoculating a 1:1000 dilution of a preculture into 5 mL of YPD in a 15-mL culture tube. After a day of incubation at 30°C with shaking at 220 rpm in an Infors HT incubator, an aliquot of the culture was diluted to an optical density at 600 nm (OD_{600}) of 0.2 in 5 mL of BMG in a 15-mL culture tube. This tube was incubated under the same conditions as before. The following day, the culture was centrifuged at 3000 rpm (2000xg) for 5 min and resuspended in 25 mL BMM to attain an OD_{600} of 1.0. 25 mL of this culture was placed in a 250-mL baffled flask (Corning), and during this induction phase, the cells were incubated at 25°C with shaking at 150 rpm to reduce loss of methanol. Additional flasks containing water were present in the shaker to generate a humid atmosphere and minimize evaporation. After one day of induction, an additional dose of 125 µL methanol was added (yielding a final concentration of 0.5%), and the incubation was continued for another day. After 48 h of induction, 1.5 mL of the culture was centrifuged at 6000 rpm (2000xg) in a microcentrifuge for 2 min to separate the cell pellet from the supernatant. The cell pellet was washed with phosphate buffered saline (PBS) and then resuspended in 1.5 mL PBS.

300 µL each of the pellet and supernatant fractions were transferred in triplicate to a 96-well plate (Costar) to measure E2-Crimson fluorescence using a Synergy Neo

Microplate Reader (BioTek). The excitation and emission wavelengths were 611 nm and 646 nm, respectively, and the gain was 150. The results were normalized by dividing by the final OD_{600} value for the culture.

3.3. Assaying secretion of BTL2

The basic procedures described above for E2-Crimson were adapted to obtain strains expressing BTL2 and to monitor BTL2 secretion. To quantify the secretion of BTL2, lipolytic activity was measured in duplicate using a lipase colorimetric assay (Roche Diagnostics). Briefly, 0.5 mL of a suitably diluted supernatant from each strain was mixed with 0.5 mL Tris-HCl buffer (200 mM, pH 7.25), placed in a thermostatically controlled cuvette, and incubated at 50°C for 5 min. Then 0.3 mL of substrate (1,2-O-dilauryl-rac-glycero-3-glutaric-(methylresorufin)-ester) was mixed with the pre-warmed sample and monitored at 580 nm for 7 min in a Specord 200 Plus Spectrophotometer (Analytic Jena). Lipolytic activity was measured using the values between minutes 3 and 5. The absorbance increase per second was used to determine the protein activity, with one unit of lipolytic activity defined as the amount of lipase needed to hydrolyze 1 µmol of ester bond per minute.

3.4. Immunoblot for E2-Crimson

Strains expressing E2-Crimson were grown and induced as described above. A 25-mL culture was centrifuged at 4°C at 4700 rpm (5000xg) for 5 min, then the supernatant was carefully removed and placed on ice. 111 µl of 100% w/v trichloroacetic acid (TCA) was added per mL of supernatant. The sample was vortexed briefly and left on ice for 20 min. Then the samples were then centrifuged at maximum speed in a microcentrifuge for 15 min. The supernatant was discarded, and the precipitate was washed with 100% ethanol, resuspended in 50 µL SDS-PAGE sample buffer, boiled for 10 min, vortexed, and centrifuged at maximum speed in a microcentrifuge for 5 min to remove insoluble material. A 20-µL aliquot was loaded on a 4-20% Mini-PROTEAN TGX Precast Protein Gel (Bio-Rad).

Meanwhile, the cell pellet was washed twice with 5 mL H_2O and then resuspended in 5 mL H_2O. A 1-mL aliquot was transferred to a snap-cap tube. The cells were centrifuged at 5000 rpm (1400xg) in a microcentrifuge for 5 min. The supernatant was discarded, and the pellet was resuspended in 250 µL 20% w/v TCA. Then 250 µL of 0.5-mm glass beads were added and the sample was vortexed at maximum speed for three 1-min pulses, with 1 min on ice between pulses. 800 µL of 5% (w/v) TCA was added and the sample was briefly mixed. Then 800 µL of the liquid was transferred to a fresh snap-cap tube and left on ice for 15 min. Finally, the sample was centrifuged and processed in the same as the sample from the supernatant containing the secreted proteins, except that the sample from the cell pellet was resuspended in 100 µL SDS-PAGE sample buffer.

Proteins were transferred from the SDS-PAGE gel to a PVDF membrane using the Trans-Blot Turbo System (Bio-Rad). The membrane was then blocked for 1 h with shaking at room temperature in TBST + 5% milk, where TBST is TBS (50 mM Tris-HCl at pH 7.6, 150 mM NaCl) plus 0.05% Tween 20. The blocked membrane was incubated with shaking overnight at 4°C in TBST + 5% milk containing a 1:500 dilution of Living Colors DsRed Monoclonal Antibody (Clontech/TaKaRa). Then the membrane was washed three times for 5 min each in TBST, and incubated with TBST + 5% milk containing a 1:1000 dilution of goat anti-mouse antibody conjugated to Alexa Fluor 647 (Thermo Fisher) for 1 h. After three washes in TBST, the membrane was washed once more in TBS prior to imaging with a LI-COR Odyssey CLx imaging system.

3.5. Fluorescence microscopy

Images for Figures 3 and S4 were captured after the 48-h induction period. Prior to imaging, cultures were spun briefly in a microcentrifuge and then resuspended in PBS to avoid fluorescence background from secreted E2-Crimson.

For Figures 4 and 7, the strains were grown in BMG and then transferred to a 5-mL culture tube containing BMM for induction at a starting OD_{600} of 0.2. The following day, the cultures were processed and imaged as in Figure 3.

Images for Figures 3, S4, and S5 were captured as Z-stacks using an LSM 880 confocal microscope (Zeiss) equipped with a 1.4-NA/100x oil objective. Images for Figure 7 were

captured as Z-stacks using an SP5 confocal microscope (Leica) equipped with a 1.4-NA/63x oil objective. The Z-stacks were average projected, and the brightness and contrast were adjusted evenly in all images. A Gaussian blur filter was used to smooth the red and green signals. Image processing was performed using ImageJ (https://imagej.nih.gov/ij/).

4. Results

The Ost1 signal sequence and a variant of the α-factor pro region synergistically promote efficient secretion of E2-Crimson

It was previously shown that the Ost1 signal sequence is more effective than the α-factor signal sequence at promoting secretion of msGFP [18]. Our goal was to test whether those findings could be extended to an oligomeric model protein. For this purpose, we generated a pre-pro-α-factor-E2-Crimson construct, which contains the α-factor signal sequence and pro region, and compared it to a pre-Ost1-pro-α-factor-E2-Crimson construct, which contains the Ost1 signal sequence and the α-factor pro region (Figure 1).

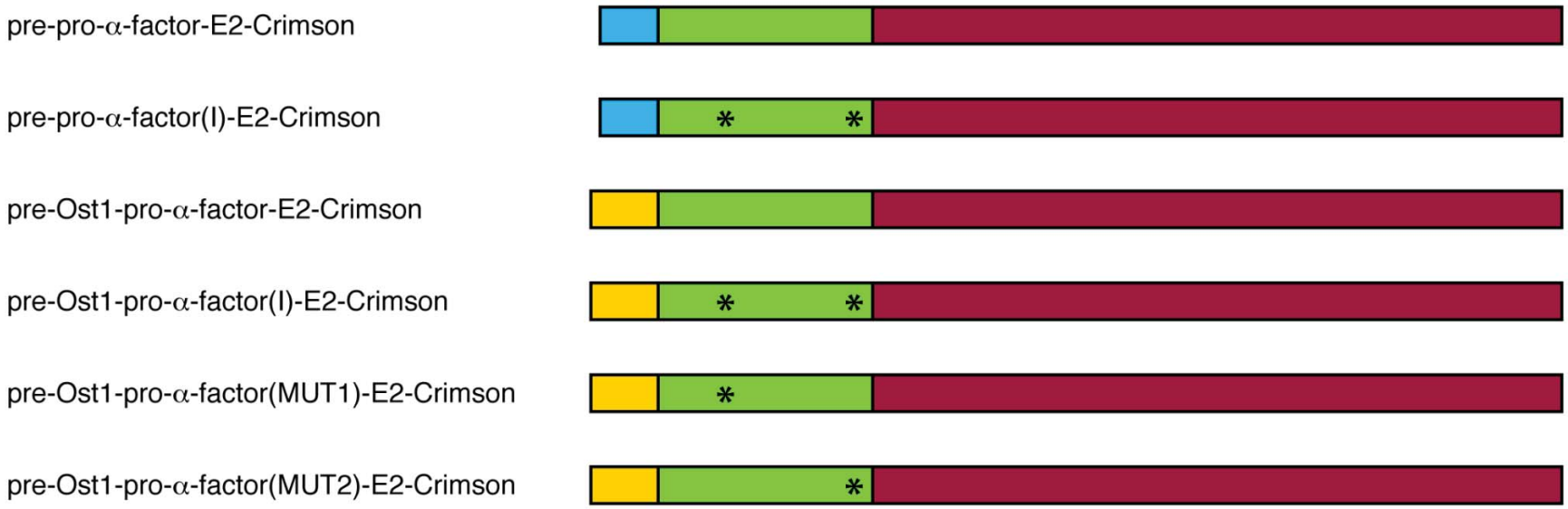

Figure 1. Constructs used in this study. Blue is the α-factor signal sequence, green is the α-factor pro region, yellow is the Ost1 signal sequence, and red is E2-Crimson. The wild-type α-factor pro region variant designated here as pro-α-factor contains Leu42. The pro-α-factor(I) or "Invitrogen" variant contains Ser42 as well as the Asp83-to-Glu mutation (see Figure S1), as represented by the asterisks. The (MUT1) and (MUT2) variants of the α-factor pro region contain the individual Ser42 and Glu83 mutations, respectively, relative to the wild-type Leu42 variant.

The other variable we tested was the sequence of the α-factor pro region. Our earlier work employed a pro region variant that contains Leu at position 42 (where the numbering is based on the pre-pro-α-factor precursor) (Figure S1). This Leu42 variant is commonly used for both biotechnology and basic science applications [2,34]. By contrast, the originally described allele of the α-factor gene contains Ser at position 42 [35]. This Ser42 variant is present in Invitrogen's widely used pPICZα family of plasmids, which also contain a trio of point mutations that create an XhoI restriction site while changing Asp83 to Glu. The Leu42 variant of the pro region is referred to here simply as pro-α-factor, while the "Invitrogen" Ser42 variant with the XhoI site is referred to as pro-α-factor(I). The constructs with the α-factor and Ost1 signal sequences were modified to include pro-α-factor(I), yielding a total of four constructs that represented all combinations of the signal sequences and pro regions (Figure 1).

Expression of these constructs was driven by the *AOX1* promoter [36]. After 48 h of methanol induction, the levels of intracellular and extracellular E2-Crimson fluorescence were measured using a fluorimeter. For each construct, six single-copy integrant clones were tested to confirm that the results were reasonably consistent (Figure S2), and two representative clones were used for further analysis. The pre-pro-α-factor-E2-Crimson reference construct contained the α-factor signal sequence followed by pro-α-factor. Figure 2A shows that in the context of the α-factor signal sequence, pro-α-factor(I) increased secretion ~3-fold. In a parallel test, when paired with pro-α-factor, the Ost1 signal sequence increased secretion ~12-fold. When the Ost1 signal sequence was paired with pro-α-factor(I), secretion was increased up to 20-fold (Figure 2A). The amount of intracellular fluorescence (Figure 2B) was inversely correlated with the amount of extracellular fluorescence (Figure 2A). Total fluorescence recovery was highest with the pre-Ost1-pro-α-factor(I) construct (Figure 2C), possibly because the other constructs led to degradation of protein molecules that failed to be secreted. These results indicate that an improved secretion signal consisting of the Ost1 signal sequence and pro-α-factor(I) is remarkably effective at promoting secretion of E2-Crimson.

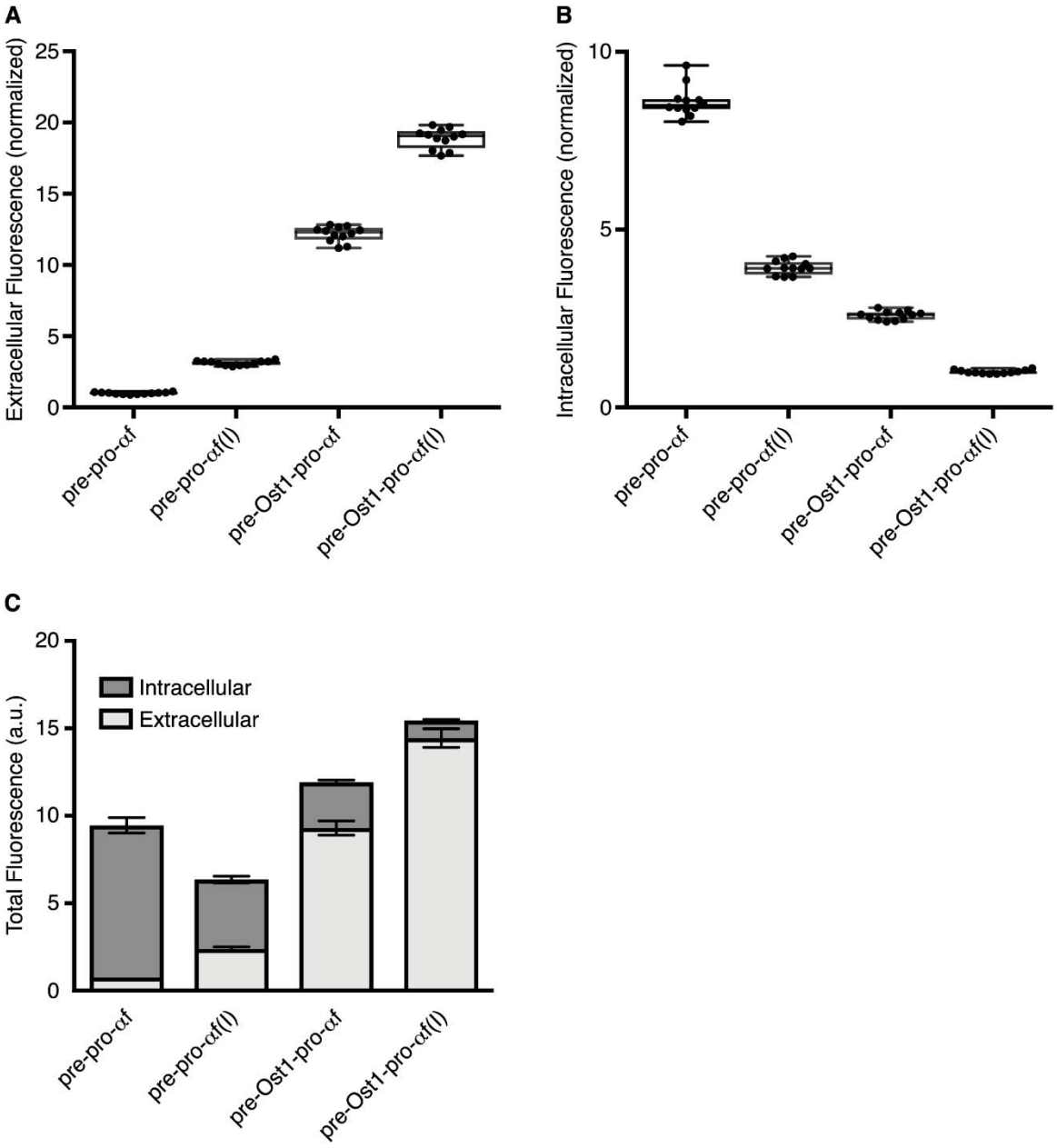

Figure 2. Extracellular and intracellular fluorescence signals with the different secretion signals. Fluorescence signals for extracellular and intracellular E2-Crimson were measured by fluorimetry after 48 h of methanol induction using different secretion signals. pre-pro-αf, α-factor signal sequence followed by pro-α-factor; pre-pro-αf(I), α-factor signal sequence followed by pro-α-factor(I); pre-Ost1-pro-αf, Ost1 signal sequence followed by pro-α-factor; pre-Ost1-pro-αf(I), Ost1 signal sequence followed by pro-α-factor(I). (**A**) Box plot showing E2-Crimson fluorescence in the culture medium for the different secretion signals. Each fluorescence signal was divided by the OD_{600} at the end of the incubation. Then the signals were normalized by setting the signal for pre-pro-αf to 1. (**B**) Same as (A), except that intracellular fluorescence signals were normalized by setting the signal for pre-Ost1-pro-αf(I) to 1. (**C**) Total extracellular and intracellular signals are plotted for the different secretion signals. Bars show standard deviations. a.u., arbitrary units.

The fluorescence measurements were verified qualitatively by immunoblotting (Figure S3). A protein that migrated at the position expected for mature E2-Crimson (26 kDa) was seen in the medium, and this gel band was most intense with the pre-Ost1-pro-α-factor(I) construct. Thus, the improved secretion signal appears to be proteolytically processed by *P. pastoris* cells in the same manner as the α-factor secretion signal. Compared to secreted E2-Crimson, cell-associated E2-Crimson migrated more slowly. This gel band was most intense with the constructs containing the α-factor signal sequence. As described below, most of the cell-associated E2-Crimson was probably molecules that failed to cross the ER membrane completely, so the higher apparent molecular weight reflects the presence of one or both parts of the secretion signal.

Intracellular E2-Crimson constructs become trapped during or after translocation into the ER

The constructs that showed substantial intracellular accumulation were presumably becoming trapped at early stages in the secretory pathway. To characterize those bottlenecks, we used fluorescence microscopy to visualize the location of the accumulated E2-Crimson. Figure 3 shows the cellular fluorescence patterns, at two brightness levels, for cells expressing the four constructs described above.

Based on the previous work with msGFP [18], we anticipated that the α-factor signal sequence would drive posttranslational translocation, and would therefore lead to accumulation of translocation intermediates in which the folded E2-Crimson domain remained on the cytosolic side of the ER membrane. Our fluorescence images fit this prediction. The two constructs with the α-factor signal sequence yielded fluorescent rings typical of the ER (Figure 3A,B), which consists of the nuclear envelope plus peripheral ER elements. Those rings were indeed the nuclear envelope as confirmed by labeling the nuclear DNA with histone H2B (Htb2) fused to GFP (Figure 4A). For the constructs containing the α-factor signal sequence, their strong ER labeling (Figure S4) combined with their relatively weak secretion suggests that they were trapped in transit across the ER membrane, with the signal sequences penetrating into the ER lumen while the folded E2-Crimson domains remained on the cytosolic side of the ER membrane.

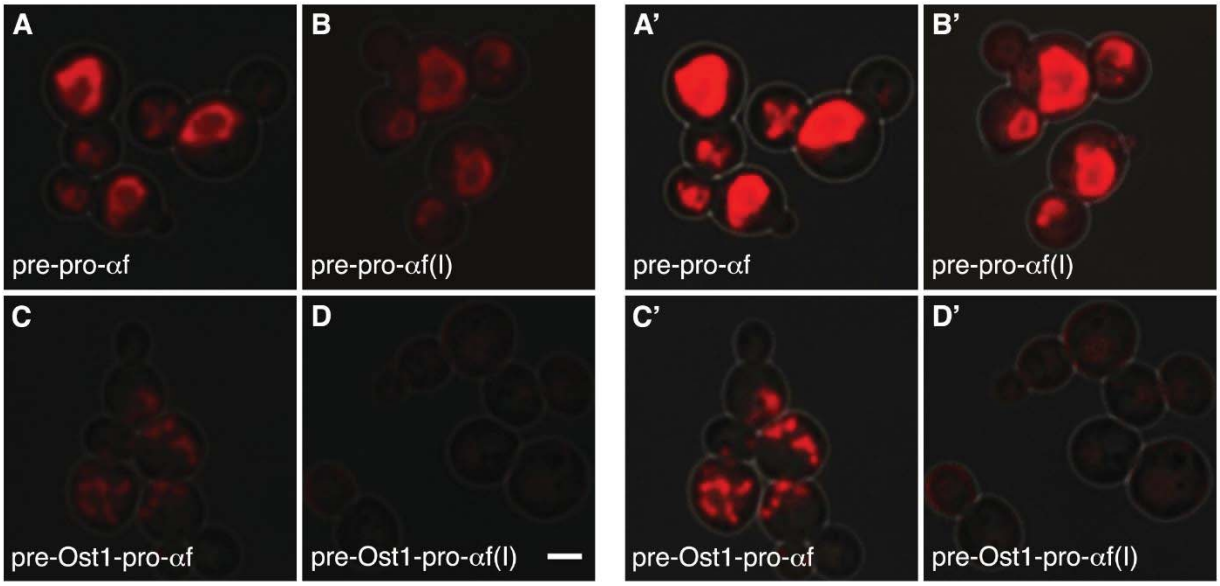

Figure 3. Images of intracellular fluorescence with the different secretion signals. The abbreviations are as in Figure 2. Projected confocal Z-stacks of E2-Crimson fluorescence were merged with differential interference contrast images of the cells. The rings in panels A and B represent the nuclear envelope, and the spots in panel C represent aggregates in the ER lumen. Panels A' through D' are the same images as panels A through D but adjusted to a higher brightness level. Scale bar, 2 μm.

If this interpretation is correct, the E2-Crimson constructs containing the α-factor signal sequence might be expected to "clog" the translocons in the ER [37]. To test this idea, we expressed an additional construct in which the Kar2 signal sequence was fused to GFP-HDEL. This fusion protein normally labels the ER lumen [38], and we saw the expected rings of ER-localized green fluorescence prior to methanol induction (Figure S5). However, in methanol-induced cells expressing the E2-Crimson constructs with the α-factor signal sequence, much of the green fluorescence was cytosolic (Figure 4B). Moreover, cells expressing the E2-Crimson constructs containing the α-factor signal sequence were often unusually large (Figure S4), consistent with a toxic effect of those constructs. The combined results support the idea that the α-factor signal sequence generates an intermediate that becomes trapped in the translocons during passage across the ER membrane.

The Ost1 signal sequence drives cotranslational translocation, and should therefore enable E2-Crimson to reach the ER lumen. With the pre-Ost1-pro-α-factor-E2-Crimson construct, there was no labeling of the nuclear envelope. Instead, punctate structures were observed in the cells (Figure 3C). Those structures were apparently aggregates located in the ER lumen because they also labeled with GFP-HDEL (Figure 4B). In this strain the GFP-HDEL was present in the ER rather than the cytosol, indicating that the translocons were not clogged (Figure 4B). The implication is that the Ost1 signal sequence overcomes the problem of translocating E2-Crimson into the ER but does not prevent subsequent aggregation in the ER lumen.

When pro-α-factor was replaced with the "Invitrogen" variant in the pre-Ost1-pro-α-factor(I)-E2-Crimson construct, no aggregates were seen in the ER lumen. Indeed, very little intracellular red fluorescence was seen (Figure 3D), consistent with the results of the fluorescence assays (Figure 2). Our interpretation is that pro-α-factor can lead to aggregation in the ER lumen, and that this effect is avoided by using pro-α-factor(I) instead.

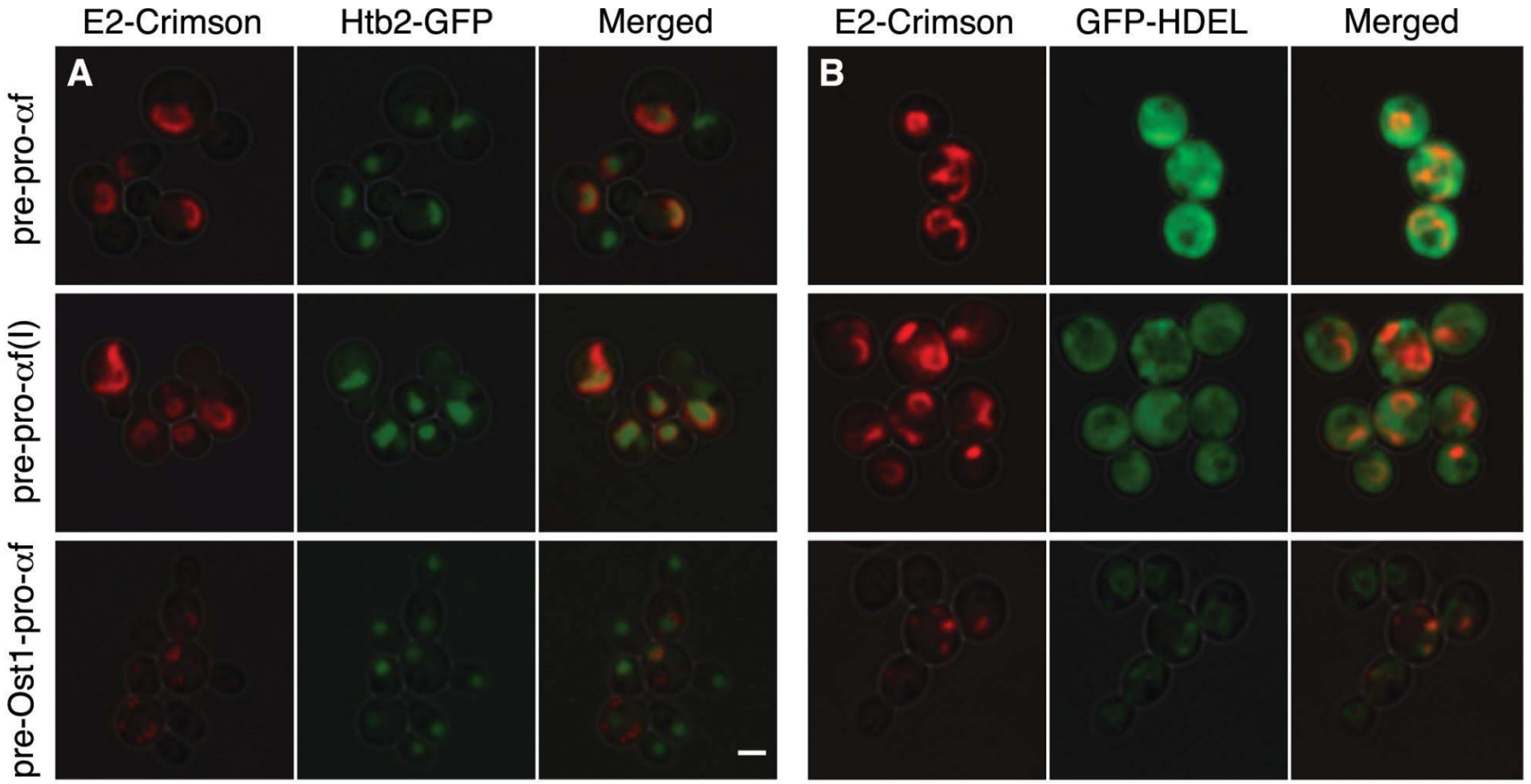

Figure 4. Confirmation that intracellular E2-Crimson constructs were associated with the ER. The abbreviations for the constructs are as in Figure 2. (**A**) Htb2-GFP represents histone 2B tagged with GFP to label the nucleus. For the constructs with the α-factor signal sequence, much of the intracellular red fluorescence was in the nuclear envelope. For the construct with the Ost1 signal sequence, punctate aggregates were visible. (**B**) GFP-HDEL represents ER-targeted GFP with a C-terminal HDEL tetrapeptide for ER retention. For the constructs with the α-factor signal sequence, most of the GFP-HDEL remained in the cytosol. For the construct with the Ost1 signal sequence, GFP-HDEL exhibited a typical ER pattern, and was present in the same locations as the E2-Crimson aggregates. Scale bar, 2 μm.

The superior behavior of pro-α-factor(I) is due to Ser42

We tested whether the enhanced secretion obtained with pro-α-factor(I) was due to one or both of the amino acid differences relative to pro-α-factor. For this purpose, pro-α-factor was modified by introducing either a point mutation that changed Leu42 to Ser, or a point mutation that changed Asp83 to Glu (Figure S1). These variants were designated pro-α-factor(MUT1) and pro-α-factor(MUT2), respectively. The MUT1 and MUT2 variants were tested in the context of the Ost1 signal sequence.

As shown in Figure 5, measurements of extracellular fluorescence gave an unambiguous answer: the MUT1 change (Leu42 to Ser) was necessary and sufficient for enhancing E2-Crimson secretion. Moreover, intracellular aggregates were seen in cells expressing pre-Ost1-pro-α-factor(MUT2)-E2-Crimson but not in cells expressing pre-Ost1-pro-α-factor(MUT1)-E2-Crimson (data not shown). To gain insight into the potential mechanism of the MUT1 mutation, we analyzed the α-factor pro region using a program called AGGRESCAN, which estimates the aggregation propensity of a polypeptide sequence [39]. A stretch of amino acids containing Leu42 was predicted to be aggregation-prone, and the predicted aggregation propensity was substantially reduced by mutating Leu42 to Ser (Figure S6). We conclude that mutation of Leu42 to the less hydrophobic Ser is crucial for suppressing aggregation of E2-Crimson constructs in the ER.

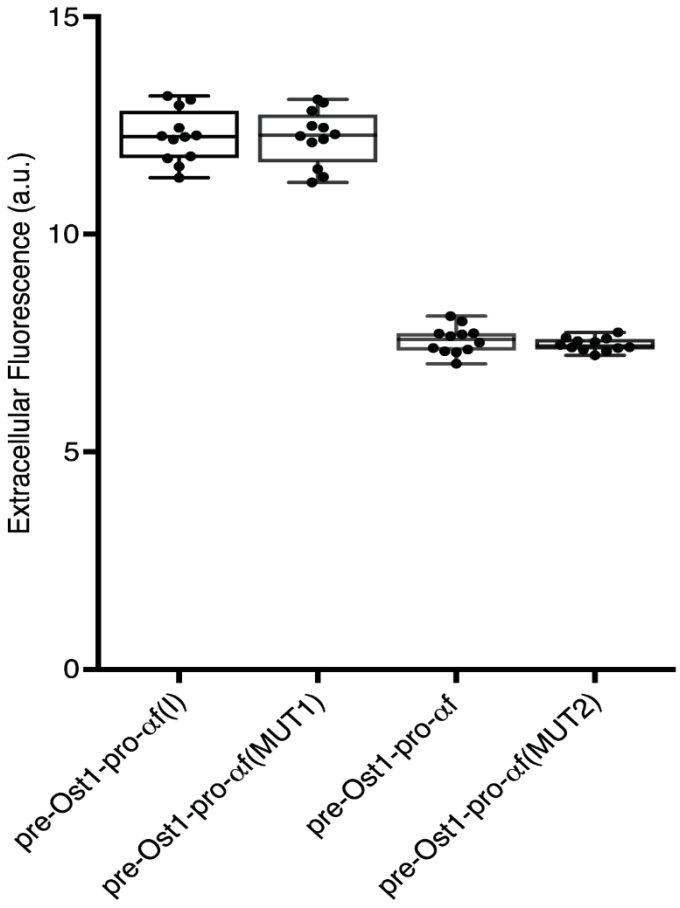

Figure 5. Separate analysis of the two differences that distinguish pro-α-factor(I) from pro-α-factor. The experiment was performed as in Figure 2A, except that the signals were not normalized and the Ost1 signal sequence was used together with either pro-α-factor, or pro-α-factor(I), or the MUT1 or MUT2 variant of the α-factor pro region.

Secretion of the BTL2 lipase is enhanced by the improved secretion signal

An obvious question is whether the enhanced secretion that we have documented for fluorescent proteins will also be seen for proteins of industrial interest. As a test case, we chose the BTL2 lipase for the reasons outlined in the Introduction. Figure 6 shows the effects of different secretion signals on the secretion of BTL2 from *P. pastoris*, as determined by measuring lipolytic activity in the medium. The results are similar to those obtained with E2-Crimson. For constructs with the Ost1 signal sequence, lipolytic activity in the medium was substantially higher than for constructs with the α-factor signal sequence. For constructs with either the α-factor or Ost1 signal sequence, pro-α-factor(I) produced better results than pro-α-factor. These two effects were additive, and

the improved signal sequence yielded about 10-fold more lipolytic activity in the medium than the original pre-pro-α-factor signal sequence (Figure 6A).

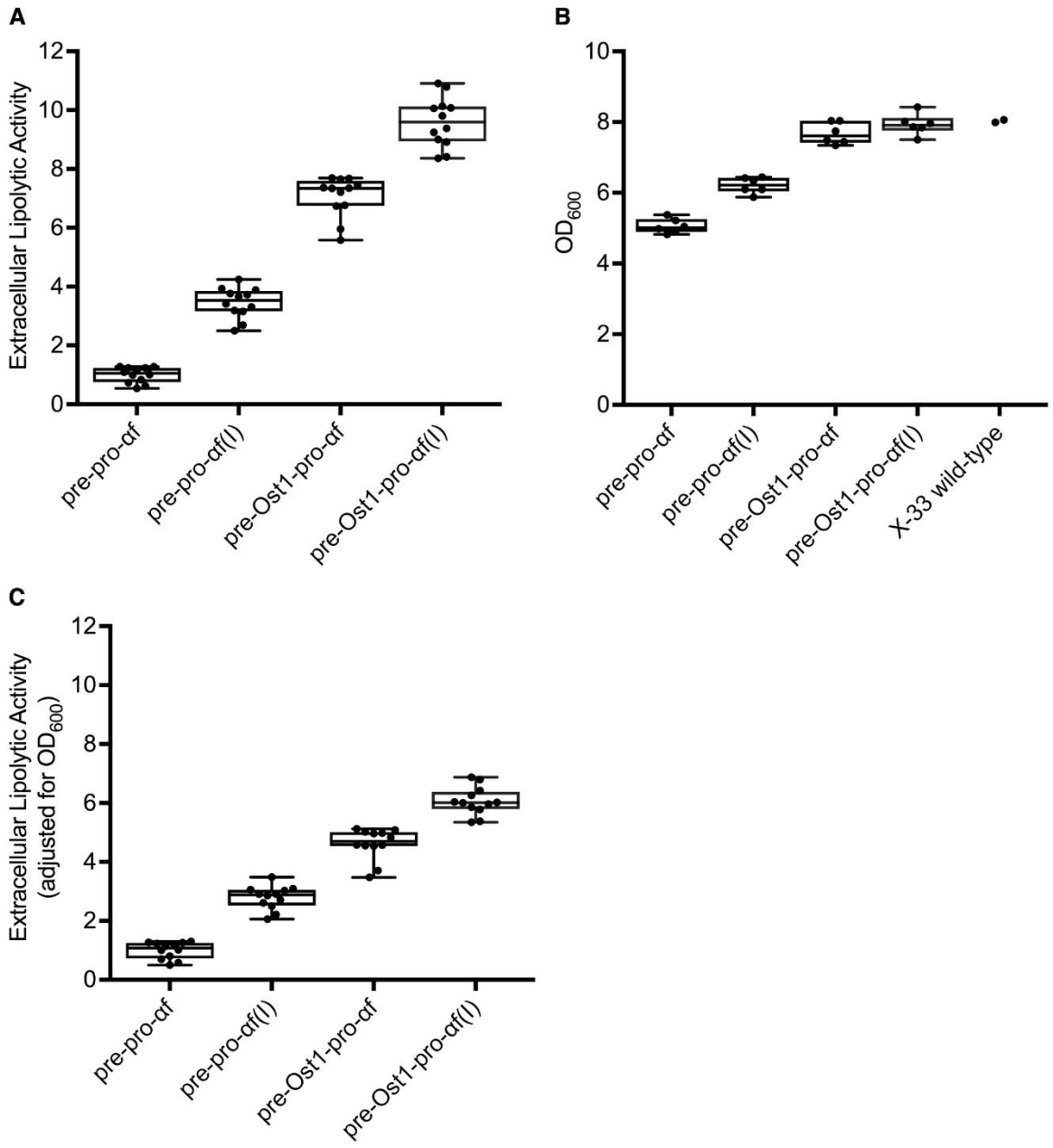

Figure 6 Extracellular BTL2 activity with the different secretion signals. This experiment was performed as in Figure 2A, except that the secreted protein was BTL2. The abbreviations for the constructs are as in Figure 2. (**A**) Box plot showing BTL2 lipolytic activity in the culture medium for the different secretion signals. The signals were normalized by setting the signal for pre-pro-αf to 1. (**B**) Box plot showing the final OD_{600} of the cultures at the end of the screening period. Two cultures of the parental X33 wild-type strain were processed in parallel as a control. (**C**) Same as (A), except that each lipolytic activity value was divided by the OD_{600} at the end of the incubation. Then the signals were normalized by setting the signal for pre-pro-αf to 1. Bars show standard deviations.

One difference compared to the results with E2-Crimson was that the BTL2 constructs containing the α-factor signal sequence inhibited cell growth. At the end of the screening period, the OD_{600} of the cultures was lower for those constructs than for the ones containing the Ost1 signal sequence (Figure 6B). The differences in cell density can account for some of the differences in the amount of secreted BTL2. However, when the lipolytic activity in the medium was normalized to the OD_{600} of the cultures, the beneficial effects of the Ost1 signal sequence and pro-α-factor(I) were still evident (Figure 6C).

Our interpretation is that like E2-Crimson, BTL2 can fold prematurely in the cytosol and clog the translocon during posttranslational translocation. To test this idea, the strains expressing the BTL2 constructs were transformed with the construct expressing ER-targeted GFP-HDEL. For cells expressing BTL2 constructs with the α-factor signal sequence, most of the fluorescence was cytosolic, whereas for cells expressing BTL2 constructs with the Ost1 signal sequence, most of the fluorescence showed a typical ER pattern (Figure 7). This result supports the idea that efficient secretion of BTL2 requires cotranslational translocation.

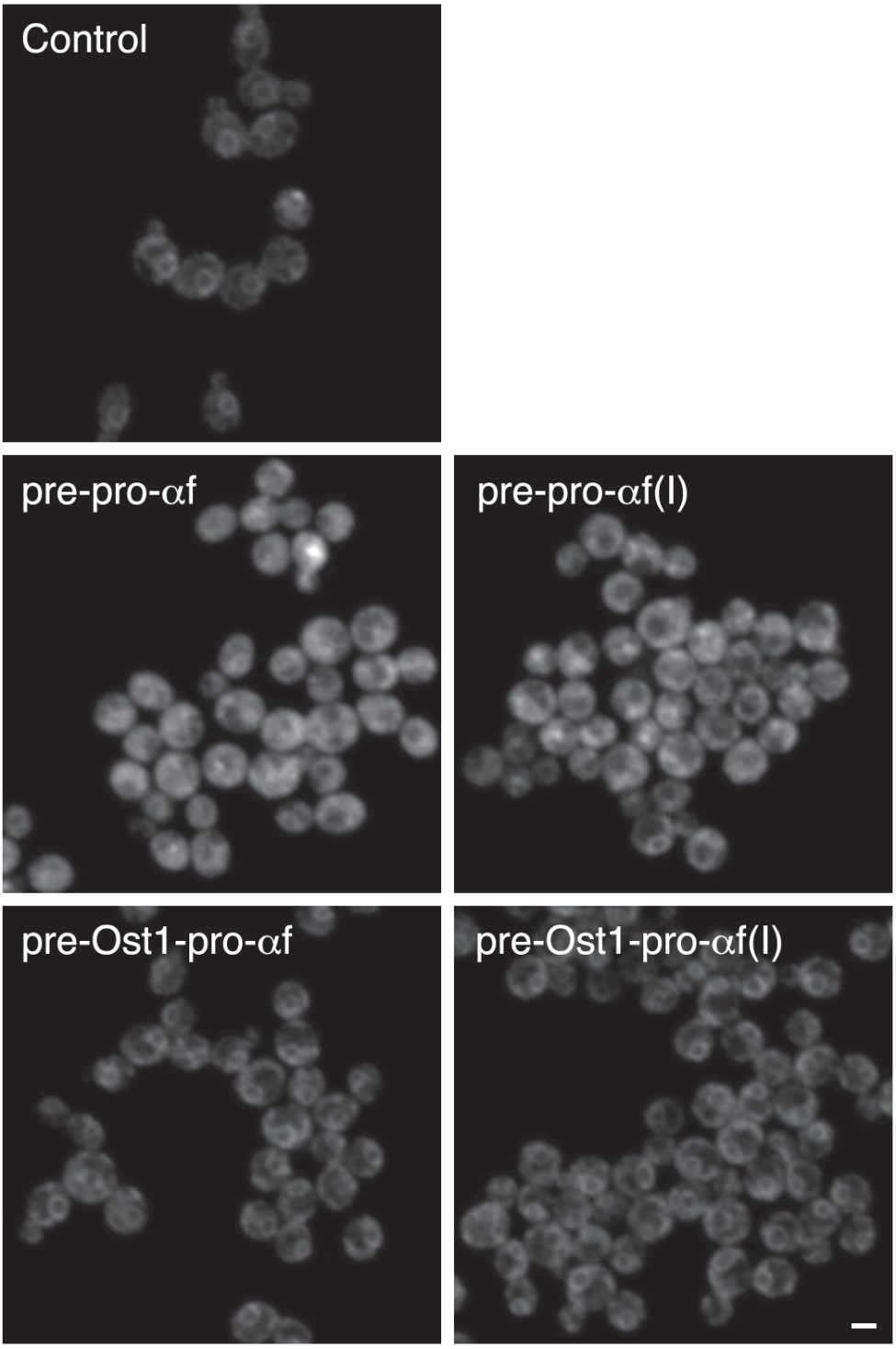

Figure 7 Clogging of the ER translocon by posttranslational translocation of BTL2. This experiment was performed as in Figure 4, except that the secreted protein was BTL2. The abbreviations for the constructs are as in Figure 2. Shown are fluorescence images of ER-targeted GFP-HDEL. In the "Control" sample, no BTL2 construct was expressed, and GFP-HDEL exhibited a typical ER pattern. For the constructs with the α-factor signal sequence, most of the GFP-HDEL remained in the cytosol. For the constructs with the Ost1 signal sequence, GFP-HDEL exhibited a typical ER pattern. Scale bar, 2 μm.

5. Discussion

We extended our previous cell biological approach [18] to devise an improved secretion signal for producing heterologous proteins in *P. pastoris*. The primary model protein was the fluorescent protein E2-Crimson, which was useful for three reasons. First, E2-Crimson can fold rapidly in the cytosol, so it serves as an example of a protein that requires efficient cotranslational translocation to enter the secretory pathway. Second, E2-Crimson is tetrameric, so it is likely to amplify any aggregation tendency of a secretion signal. Third, E2-Crimson emits red fluorescence, so it enables the detection of intracellular protein by fluorescence microscopy and of extracellular protein by fluorimetry. As described below, the analysis of E2-Crimson revealed ways to enhance the secretion signal.

In *S. cerevisiae*, the N-terminal signal sequence plays a major role in determining whether translocation is cotranslational or posttranslational [16,40,41]. Most soluble proteins that enter the *S. cerevisiae* secretory pathway, including pre-pro-α-factor, undergo posttranslational translocation [42]. However, the Ost1 signal sequence directs efficient cotranslational translocation in *S. cerevisiae* [18,19,43]. Although translocation into the ER has not been studied in *P. pastoris*, the mechanisms are probably similar to those observed in *S. cerevisiae*. Indeed, we reported that replacement of the α-factor signal sequence with the Ost1 signal sequence strongly increased secretion of msGFP in *P. pastoris* [18]. A similar beneficial effect of the Ost1 signal sequence is described here for secretion of E2-Crimson in *P. pastoris*.

After reaching the ER lumen, proteins can be exported by bulk flow, but this process is relatively slow [44]. Faster export is mediated by signal-dependent ER export receptors that concentrate secretory proteins in COPII vesicles [3,45]. For example, the α-factor pro region contains an ER export signal that is recognized by the transmembrane Erv29 receptor [3,46,47]. This receptor-driven export probably helps to explain why the α-factor secretion signal is often effective for heterologous protein production. Yet the presence of the α-factor pro region in a secretion signal also carries risk, because the pro region remains attached to the secretory protein until being removed by the Kex2 processing protease in the Golgi [4]. When a secretory protein is present at high levels

in the ER lumen, the α-factor pro region might cause aggregation, particularly if the secretory protein is oligomeric. Such aggregation in the ER was seen when the α-factor pro region was linked to E2-Crimson. By contrast, an allelic variant of the α-factor pro region caused no aggregation of E2-Crimson. We traced this effect to amino acid 42 in the pro region, where Leu42 promotes aggregation but Ser42 does not. The Ser42 variant is present in the widely used *P. pastoris* expression vectors supplied by Invitrogen, and it has likely benefited *P. pastoris* researchers who used those vectors. Thus, α-factor pro region variants that contain Ser42 can be used to drive rapid signal-mediated ER exit of a heterologous protein without the side effect of promoting ER aggregation. The combined effect of the Ost1 signal sequence and a Ser42 variant of the α-factor pro region was dramatic—compared to the Leu42 variant of the α-factor secretion signal, the improved secretion signal enhanced secretion of E2-Crimson approximately 20-fold.

To test whether the benefits of the improved secretion signal extend to secreted proteins of practical value, we tested the lipase BTL2. This protein was a good candidate for four reasons. First, *P. pastoris* had previously been engineered to secrete active BTL2 [29] . Second, even though BTL2 is naturally produced as an extracellular enzyme [25], it was active when expressed intracellularly in *E. coli* [27], indicating that the protein can fold in the cytosol. Third, the secretion of BTL2 from *S. cerevisiae* was recently found to be influenced by the signal sequence [30]. Fourth, although BTL2 is a monomeric protein, it has a tendency to aggregate at high concentrations [28,48]. These properties suggested that BLT2 would benefit from cotranslational translocation into the ER and from the presence of Ser42 in the α-factor pro region. Indeed, both the Ost1 signal sequence and a Ser42 variant of the α-factor pro region enhanced secretion of BTL2. Compared to the Leu42 variant of the α-factor secretion signal, the improved secretion signal yielded better growth of the BTL2-expressing cells, and enhanced secretion of BTL2 approximately 10-fold.

As part of this analysis, we developed a method to assess whether a heterologous protein fused to the α-factor secretion signal accumulates in the *P. pastoris* cytosol and clogs the ER translocon. A *P. pastoris* strain expresses an ER-targeted GFP-HDEL construct, which normally gives a distinctive fluorescence pattern. When the translocon

is clogged due to failed posttranslational translocation, GFP fluorescence shifts to the cytosol. Both E2-Crimson and BTL2 produced such a shift in GFP fluorescence. For the future, it will be interesting to test whether this simple assay reliably identifies heterologous proteins that will be secreted more efficiently with the improved secretion signal.

6. Conclusions

By combining the Ost1 signal sequence with a Ser42 variant of the α-factor pro region, we obtained an improved secretion signal for *P. pastoris*. With the E2-Crimson and BTL2 model proteins, this secretion signal boosted secretion up to 20-fold and 10-fold, respectively, relative to the Leu42 variant of the α-factor secretion signal. It will be interesting to test whether the improved secretion signal enhances the secretion from *P. pastoris* of other proteins, particularly proteins that can fold prior to translocation and proteins that oligomerize in the ER, and whether these enhancements are also seen at high cell densities in bioreactor-scale fermentations. If so, the improved secretion signal could replace the common variants of the α-factor secretion signal as the default standard for producing heterologous proteins in *P. pastoris***.**

List of Abbreviations

BMG, buffered minimal glycerol medium; BMM, buffered minimal methanol medium; ER, endoplasmic reticulum; msGFP, monomeric superfolder GFP; OD_{600}, optical density at 600 nm; PBS, phosphate buffered saline; YPD, rich yeast medium

7. References

[1] Brake AJ, Merryweather JP, Coit DG, Heberlein UA, Masiarz FR, Mullenbach GT, et al. Alpha-factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 1984;81:4642–6. https://doi.org/10.1073/pnas.81.15.4642.

[2] Otte S, Barlowe C. Sorting signals can direct receptor-mediated export of soluble proteins into COPII vesicles. Nat Cell Biol 2004;6:1189–94. https://doi.org/10.1038/ncb1195.

[3] Dancourt J, Barlowe C. Protein sorting receptors in the early secretory pathway. Annu Rev Biochem 2010;79:777–802. https://doi.org/10.1146/annurev-biochem-061608-091319.

[4] Fuller RS, Sterne RE, Thorner J. Enzymes required for yeast prohormone processing. Annu Rev Physiol 1988;50:345–62. https://doi.org/10.1146/annurev.physiol.50.1.345.

[5] Julius D, Blair L, Brake A, Sprague G, Thorner J. Yeast α factor is processed from a larger precursor polypeptide: The essential role of a membrane-bound dipeptidyl aminopeptidase. Cell 1983;32:839–52. https://doi.org/10.1016/0092-8674(83)90070-3.

[6] Puxbaum V, Mattanovich D, Gasser B. Quo vadis? The challenges of recombinant protein folding and secretion in Pichia pastoris. Appl Microbiol Biotechnol 2015;99:2925–38. https://doi.org/10.1007/s00253-015-6470-z.

[7] Looser V, Bruhlmann B, Bumbak F, Stenger C, Costa M, Camattari A, et al. Cultivation strategies to enhance productivity of Pichia pastoris: A review. Biotechnol Adv 2015;33:1177–93. https://doi.org/10.1016/j.biotechadv.2015.05.008.

[8] Gasser B, Prielhofer R, Marx H, Maurer M, Nocon J, Steiger M, et al. Pichia pastoris : protein production host and model organism for biomedical research. Future Microbiol 2013;8:191–208. https://doi.org/10.2217/fmb.12.133.

[9] Sturmberger L, Chappell T, Geier M, Krainer F, Day KJ, Vide U, et al. Refined Pichia pastoris reference genome sequence. J Biotechnol 2016;235:121–31. https://doi.org/10.1016/j.jbiotec.2016.04.023.

[10] Daly R, Hearn MTW. Expression of heterologous proteins in Pichia pastoris: A useful experimental tool in protein engineenring and production. J Mol Recognit 2005;18:119–38. https://doi.org/10.1002/jmr.687.

[11] 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;98:5301–17. https://doi.org/10.1007/s00253-014-5732-5.

[12] Lin-Cereghino GP, Stark CM, Kim D, Chang J, Shaheen N, Poerwanto H, et al. The effect of α-mating factor secretion signal mutations on recombinant protein expression in Pichia pastoris. Gene 2013;519:311–7. https://doi.org/10.1016/j.gene.2013.01.062.

[13] Joo HH, Xue L, Tsai JW, Park SPJ, Kwon J, Patel A, et al. Structural characterization of the α-mating factor prepro-peptide for secretion of recombinant proteins in Pichia pastoris. Gene 2017;598:50–62. https://doi.org/10.1016/j.gene.2016.10.040.

[14] Massahi A, Çalik P. In-silico determination of Pichia pastoris signal peptides for extracellular recombinant protein production. J Theor Biol 2015;364:179–88. https://doi.org/10.1016/j.jtbi.2014.08.048.

[15] Ahn J, Jang MJ, Ang KS, Lee H, Choi ES, Lee DY. Codon optimization of Saccharomyces cerevisiae mating factor alpha prepro-leader to improve recombinant protein production in Pichia pastoris. Biotechnol Lett 2016;38:2137–43. https://doi.org/10.1007/s10529-016-2203-3.

[16] Ng DTW, Brown JD, Walter P. Signal sequences specify the targeting route to the endoplasmic reticulum membrane. J Cell Biol 1996;134:269–78. https://doi.org/10.1083/jcb.134.2.269.

[17] Plath K, Mothes W, Wilkinson BM, Stirling CJ, Rapoport TA. Signal sequence

recognition in posttranslational protein transport across the yeast ER membrane. Cell 1998;94:795–807. https://doi.org/10.1016/S0092-8674(00)81738-9.

[18] Fitzgerald I, Glick BS. Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting. Microb Cell Fact 2014;13:125. https://doi.org/10.1186/s12934-014-0125-0.

[19] Willer M, Forte GMA, Stirling CJ. Sec61p is required for ERAD-L: Genetic dissection of the translocation and ERAD-L functions of Sec61P using novel derivatives of CPY. J Biol Chem 2008;283:33883–8. https://doi.org/10.1074/jbc.M803054200.

[20] Forte GMA, Pool MR, Stirling CJ. N-terminal acetylation inhibits protein targeting to the endoplasmic reticulum. PLoS Biol 2011;9. https://doi.org/10.1371/journal.pbio.1001073.

[21] Strack RL, Hein B, Bhattacharyya D, Hell SW, Keenan RJ, Glick BS. A rapidly maturing far-red derivative of DsRed-Express2 for whole-cell labeling. Biochemistry 2009;48:8279–81. https://doi.org/10.1021/bi900870u.

[22] Surribas A, Resina D, Ferrer P, Valero F. Rivoflavin may interfere with on-line monitoring of secreted green fluorescence protein fusion proteins in Pichia pastoris. Microb Cell Fact 2007;6:15. https://doi.org/10.1186/1475-2859-6-15.

[23] Marx H, Mattanovich D, Sauer M. Overexpression of the riboflavin biosynthetic pathway in Pichia pastoris. Microb Cell Fact 2008;7:23. https://doi.org/10.1186/1475-2859-7-23.

[24] Rúa ML, Schmidt-Dannert C, Wahl S, Sprauer A, Schmid RD. Thermoalkalophilic lipase of Bacillus thermocatenulatus. J Biotechnol 1997;56:89–102. https://doi.org/10.1016/S0168-1656(97)00079-5.

[25] Schmidt-Dannert C, Rúa ML, Atomi H, Schmid RD. Thermoalkalophilic lipase of Bacillus thermocatenulatus. I. Molecular cloning, nucleotide sequence, purification and some properties. Biochim Biophys Acta - Lipids Lipid Metab 1996;1301:105–14. https://doi.org/10.1016/0005-2760(96)00027-6.

[26] Hasan F, Shah AA, Hameed A. Methods for detection and characterization of lipases: A comprehensive review. Biotechnol Adv 2009;27:782–98. https://doi.org/10.1016/j.biotechadv.2009.06.001.

[27] Rúa ML, Atomi H, Schmidt-Dannert C, Schmid RD. High-level expression of the thermoalkalophilic lipase from Bacillus thermocatenulatus in Escherichia coli. Appl Microbiol Biotechnol 1998;49:405–10. https://doi.org/10.1007/s002530051190.

[28] Rúa ML, Schmidt-Dannert C, Wahl S, Sprauer A, Schmid RD. Thermoalkalophilic lipase of Bacillus thermocatenulatus: Large-scale production, purification and properties: Aggregation behaviour and its effect on activity. J Biotechnol 1997;56:89–102. https://doi.org/10.1016/S0168-1656(97)00079-5.

[29] Quyen DT, Schmidt-Dannert C, Schmid RD. High-level expression of a lipase from Bacillus thermocatenulatus BTL2 in Pichia pastoris and some properties of the recombinant lipase. Protein Expr Purif 2003;28:102–10. https://doi.org/10.1016/S1046-5928(02)00679-4.

[30] Kajiwara S, Yamada R, Ogino H. Secretory Overexpression of Bacillus thermocatenulatus Lipase in Saccharomyces cerevisiae Using Combinatorial Library Strategy. Biotechnol J 2018;13:e1700409. https://doi.org/10.1002/biot.201700409.

[31] Wenyan Wang BAM. Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange site-directed mutagenesis. Biotechniques 1999;26:2–4.

[32] Sears IB, O'Connor J, Rossanese OW, Glick BS. A versatile set of vectors for constitutive and regulated gene expression in Pichia pastoris. Yeast 1998;14:783–90. https://doi.org/10.1002/(SICI)1097-0061(19980615)14.

[33] Cámara E, Albiol J, Ferrer P. Droplet digital PCR-aided screening and characterization of Pichia pastoris multiple gene copy strains. Biotechnol Bioeng 2016;113:1542–51. https://doi.org/10.1002/bit.25916.

[34] Rakestraw JA, Sazinsky SL, Piatesi A, Antipov E, Wittrup KD. Directed evolution of

a secretory leader for the improved expression of heterologous proteins and full-length antibodies in Saccharomyces cerevisiae. Biotechnol Bioeng 2009;103:1192–201. https://doi.org/10.1002/bit.22338.

[35] Kurjan J, Herskowitz I. Structure of a yeast pheromone gene (MFα): A putative α-factor precursor contains four tandem copies of mature α-factor. Cell 1982;30:933–43. https://doi.org/10.1016/0092-8674(82)90298-7.

[36] 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;9:1316–23. https://doi.org/10.1128/mcb.9.3.1316.

[37] Ast T, Michaelis S, Schuldiner M, States U. The protease Ste24 clears clogged translocons. Cell 2017;164:103–14. https://doi.org/10.1016/j.cell.2015.11.053.The.

[38] Rossanese OW, Reinke CA, Bevis BJ, Hammond AT, Sears IB, O'Connor J, et al. A role for actin, Cdc1p, and Myo2p in the inheritance of late Golgi elements in Saccharomyces cerevisiae. J Cell Biol 2001;153:47–61. https://doi.org/10.1083/jcb.153.1.47.

[39] Conchillo-Solé O, de Groot NS, Avilés FX, Vendrell J, Daura X, Ventura S. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics 2007;8:65. https://doi.org/10.1186/1471-2105-8-65.

[40] Mandon EC, Trueman SF, Gilmore R. Protein translocation across the rough endoplasmic reticulum. Cold Spring Harb Perspect Biol 2013;5. https://doi.org/10.1101/cshperspect.a013342.

[41] Benham AM. Protein secretion and the endoplasmic reticulum. Cold Spring Harb Perspect Biol 2012;4. https://doi.org/10.1101/cshperspect.a012872.

[42] Ast T, Cohen G, Schuldiner M. A network of cytosolic factors targets SRP-independent proteins to the endoplasmic reticulum. Cell 2013;152:1134–45. https://doi.org/10.1016/j.cell.2013.02.003.

[43] Pechmann S, Chartron JW, Frydman J. Local slowdown of translation by nonoptimal codons promotes nascent-chain recognition by SRP in vivo. Nat Struct Mol Biol 2014;21:1100–5. https://doi.org/10.1038/nsmb.2919.

[44] Barlowe C, Helenius A. Cargo capture and bulk flow in the early secretory pathway. Annu Rev Cell Dev Biol 2016;32:197–222. https://doi.org/10.1146/annurev-cellbio-111315-125016.

[45] Barlowe CK, Miller EA. Secretory protein biogenesis and traffic in the early secretory pathway. Genetics 2013;193:383–410. https://doi.org/10.1534/genetics.112.142810.

[46] Foley DA, Sharpe HJ, Otte S. Membrane topology of the endoplasmic reticulum to Golgi transport factor Erv29p. Mol Membr Biol 2007;24:259–68. https://doi.org/10.1080/09687860601178518.

[47] Belden WJ, Barlowe C. Role of Erv29p in collecting soluble secretory proteins into ER-derived transport vesicles. Science 2001;294:1528–31. https://doi.org/10.1126/science.1065224.

[48] Schlieben NH, Niefind K, Schomburg D. Expression, purification, and aggregation studies of His-tagged thermoalkalophilic lipase from Bacillus thermocatenulatus. Protein Expr Purif 2004;34:103–10. https://doi.org/10.1016/j.pep.2003.11.014.

8. Appendix

Figure S1 The *S. cerevisiae* α-factor secretion signal. The 5′ portion of the *MFα1* gene encoding pre-pro-α-factor is depicted. This sequence encodes the Leu42 allele. Cleavage sites for signal peptidase and Kex2 are marked, and Leu42 and Asp83 are highlighted. This image was generated using SnapGene software.

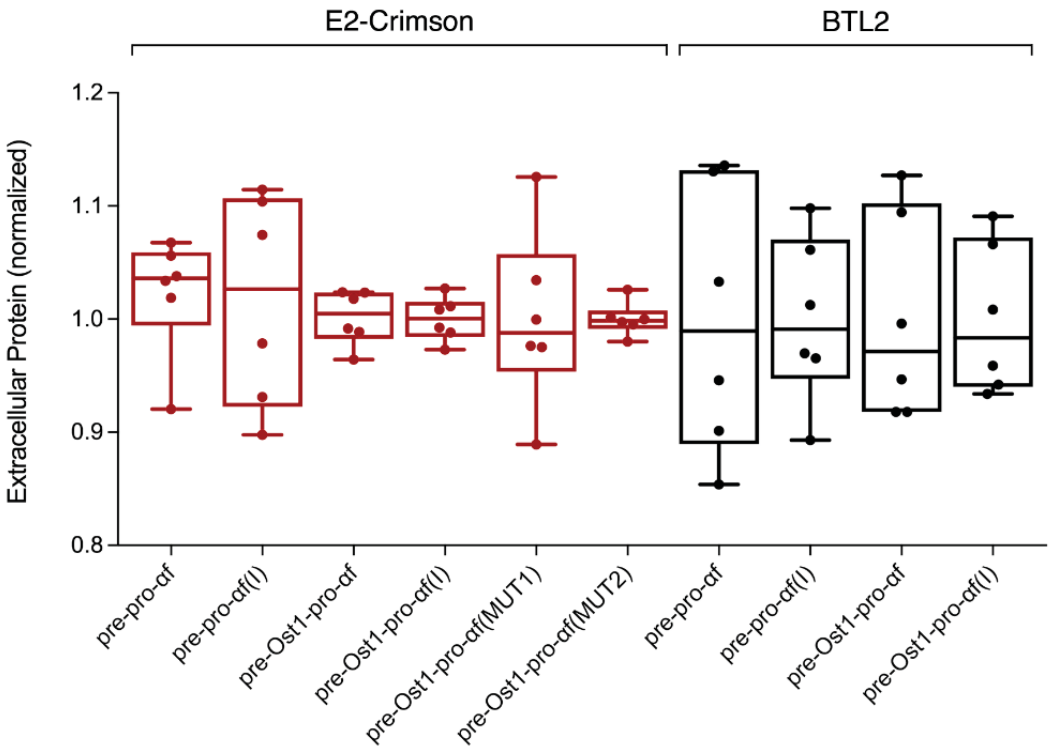

Figure S2 Low variation in secreted protein levels between clones. Each of the indicated constructs was transformed into *P. pastoris* cells, and six independent clones with confirmed single integrations were cultured and then analyzed as in Figure 2A to measure extracellular protein. For each construct, the average fluorescence level for the six clones was defined as 1.0. Two representative clones were chosen for further analysis of a given construct. The red boxes represent E2-Crimson constructs, and the black boxes represent BTL2 constructs.

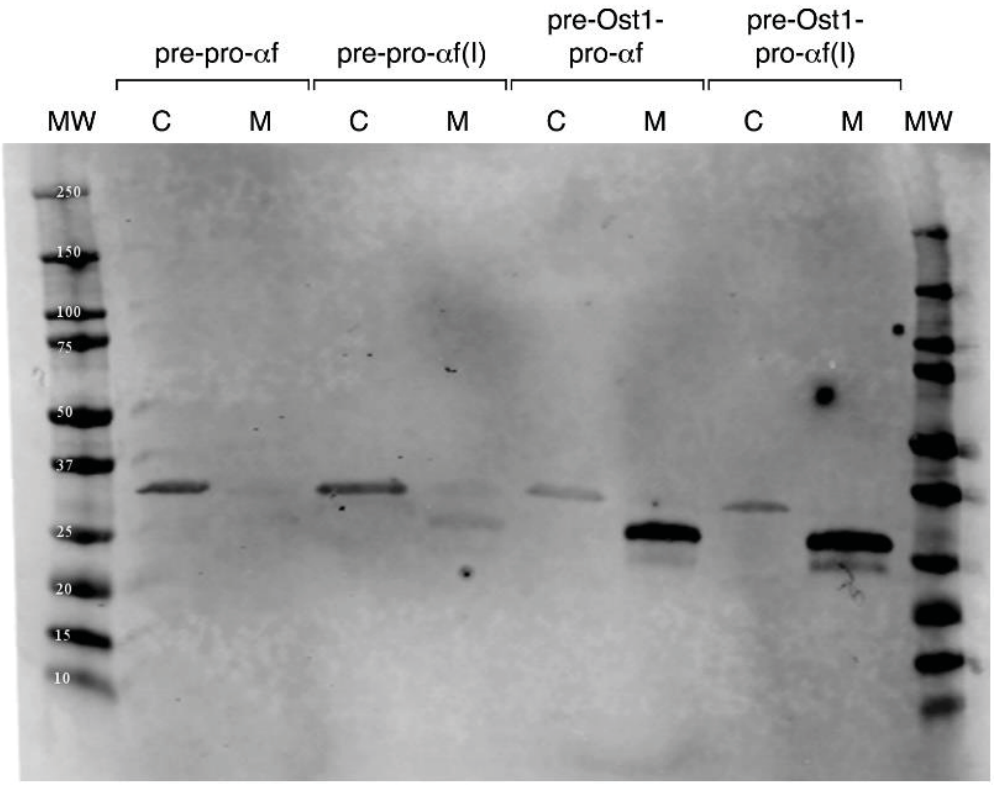

Figure S3 Immunoblot showing cell-associated and secreted E2-Crimson. Strains expressing E2-Crimson fused to the indicated secretion signals were grown and induced. Fractions containing cell-associated E2-Crimson ("C") or E2-Crimson in the extracellular medium ("M") were subjected to SDS-PAGE and immunoblotting. The molecular weight ("MW") marker was the Precision Plus Protein Dual Color Standards (Bio-Rad). Molecular weights of the protein standards are indicated.

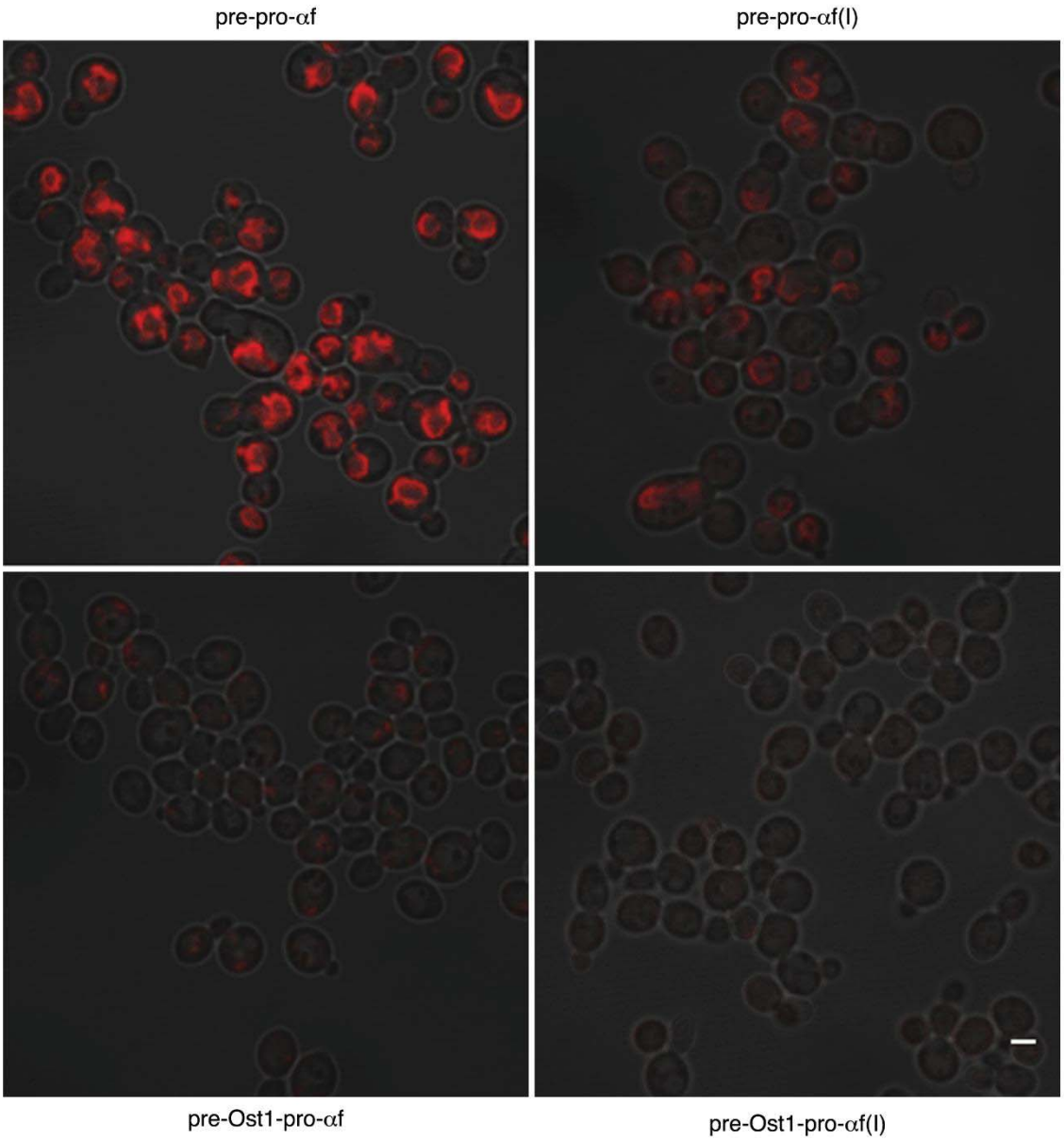

Figure S4 Large fields of cells expressing the various constructs. From the experiment of Figure 3, large fields were captured to illustrate that the fluorescence patterns were consistent among the cells in a population, and that cells expressing constructs with the α-factor signal sequence were often unusually large. Scale bar, 2 μm.

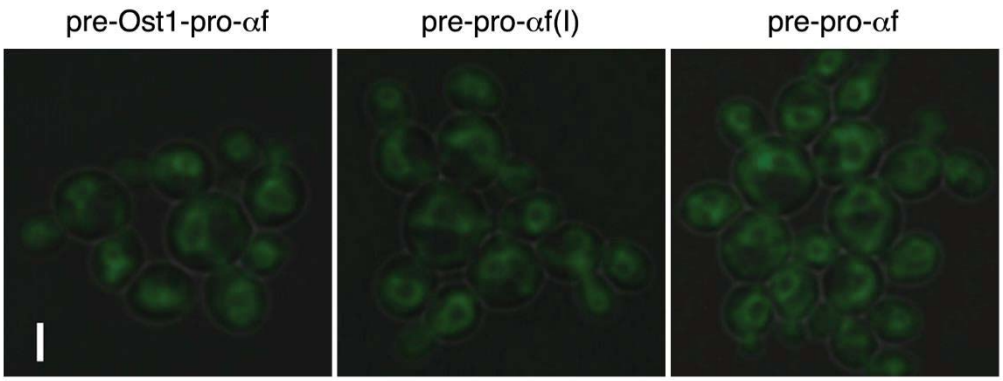

Figure S5 Distribution of ER-targeted GFP-HDEL before induced expression of E2-Crimson constructs. The same cultures shown in Figure 4B were examined by fluorescence microscopy before methanol-induced expression of the E2-Crimson constructs. All of the strains showed a typical ER pattern for GFP-HDEL. Scale bar, 2 μm.

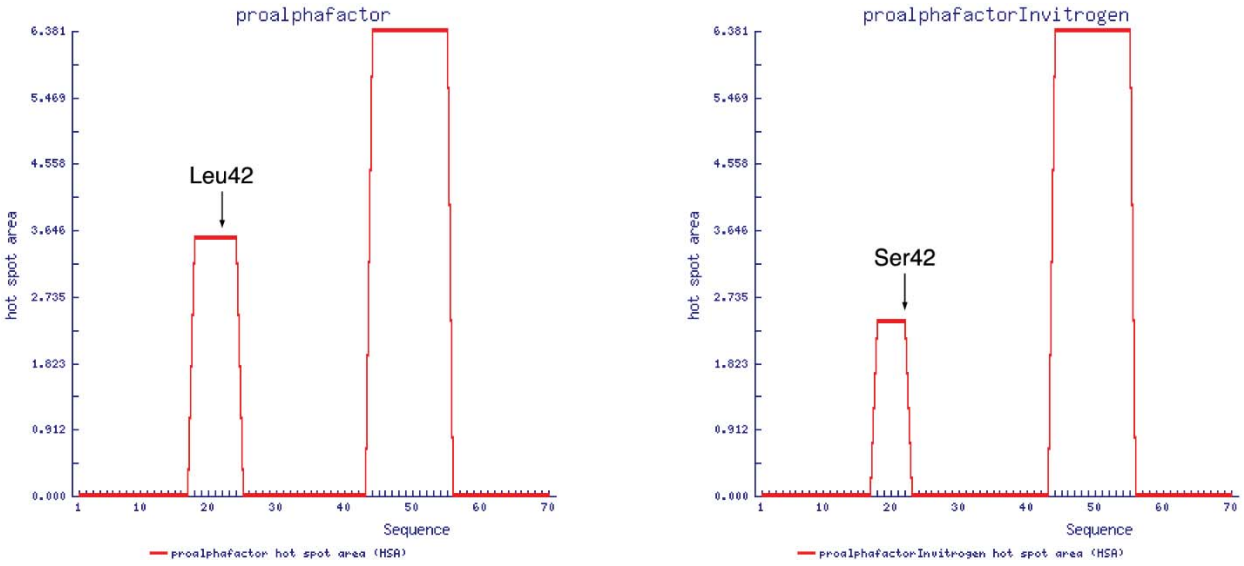

Figure S6 Theoretical prediction of aggregation propensities in the two allelic variants of the α-factor pro region. Left, the Leu42 variant of the α-factor pro region was analyzed using the online AGGRESCAN tool (http://bioinf.uab.es/aap/). The position of Leu42 in a predicted aggregation-prone region is marked. Right, the same analysis was performed for the Ser42 variant of the α-factor pro region.

5

Bioreactor-scale cell performance and protein production can be substantially increased by using a secretion signal that drives co-translational translocation in *Pichia pastoris*.

Key words: *Pichia pastoris*, heterologous protein production, signal sequence, translocation, Ost1, secretion, alpha-factor.

Table of Contents

1. Summary

The protein secretion pathway has been long recognized as a limiting process for recombinant protein secretion in yeast, in which many roadblocks have been pinpointed. In the previous chapter, we have described the identification of a bottleneck at the ER translocation level. Thereby, this limitation could be largely overcome by using an improved chimeric secretion signal to drive proteins through the co-translational translocation pathway.

Here, we have further tested at bioreactor scale the improved secretion signal consisting of the pre-Ost1 signal sequence, which drives proteins through co-translational translocation, followed by the pro region from the secretion signal of the *Saccharomyces cerevisiae* α-factor mating pheromone. For comparison, we tested the commonly used full-length α-factor secretion signal, which drives proteins through post-translational translocation. These two secretion signals were fused to three different model proteins: the tetrameric red fluorescent protein E2-Crimson, which can be used to visualize roadblocks in the secretory pathway; the mature lipase 2 from *Bacillus thermocatenulatus* (BTL2); and the mature *Rhizopus oryzae* lipase (ROL), which is known to impose a strong burden on *P. pastoris* cells.

All strains were tested in a batch cultivation to study the different growth parameters obtained. The strains carrying the improved secretion signal showed increased final production of the proteins of interest. Interestingly, these strains were able to grow at significantly higher maximum specific growth rates than their counterparts carrying the conventional secretion signal. These results were corroborated in a 5-L fed-batch cultivation, where the final product concentration and volumetric productivity were also shown to be improved.

2. Introduction

In order to achieve secretion in yeast, it is necessary to add at the N-terminus of the protein of interest a secretion signal. For *P. pastoris*, the most used secretion signal is derived from the α-factor mating pheromone (α-MF) of *S. cerevisiae* [1]. This secretion signal contains two parts: a 19-amino-acid signal sequence (the pre region) that directs translocation from the cytoplasm to the lumen of the endoplasmic reticulum (ER), followed by a 66-amino-acid pro region that is recognized by the transmembrane Erv29 receptor to mediate rapid ER export through packaging into COPII vesicles that are sent to the Golgi [2,3]. Additionally, at the end of the pro region, there is a signal that allows the secretion signal to be removed from the heterologous protein by the Golgi-localized peptidases Kex2 and Ste13 [4]. Although the α-MF secretion signal has been shown to drive efficient secretion of multiple heterologous proteins, variable levels of secretion have been reported between different proteins [5]. Therefore, several attempts to improve the secretion efficiency have been made [6], many of them focused on modifying the α-MF secretion signal [7–9].

In the first part of this study (chapter 4), we identified two major limitations of the α-MF secretion signal [10]. The first limitation is that the signal sequence (pre-α-MF) drives proteins through post-translational translocation. Therefore, if a protein intended for secretion folds prematurely, it could fail to enter the ER and could get stuck in the cytoplasm or even clog the translocons [11]. To avoid this problem, we replaced pre-α-MF with the *S. cerevisia*e Ost1 signal sequence (pre-Ost1), which has been reported to drive efficient co-translational translocation [10,12,13]. The second limitation is that the pro-region of the α-MF secretion signal is prone to aggregation in the ER. The solution was to replace the leucine at position 42 with serine. This mutation reduced the aggregation of proteins in the ER and increased the secretion efficiency. The resulting chimeric secretion signal consisted of the pre-Ost1 signal sequence followed by the Ser42 variant of the α-MF pro region [10].

Thereafter, we have studied the effect of the different translocation pathways by using either the pre-Ost1 or the pre-α-MF signal sequence, in each case fused to the Ser42 variant of the α-MF pro region. Our study has been carried out using three different model proteins produced under control of the methanol-inducible P_{AOX1}: E2-Crimson, a

far-red fluorescent protein that allows the visualization of intracellular protein trafficking, and two microbial lipases, BTL2 and ROL. These two lipases are structurally different and can give a broad indication about the effect of using pre-Ost1 on cellular physiology and protein production levels. Furthermore, this comparison was studied in both batch and fed-batch cultivations, i.e., under bioprocess-relevant conditions. Although the levels of secretion varied between proteins, all of the strains carrying the pre-Ost1 signal sequence showed an improvement in terms of final protein production compared with the strains carrying the pre-α-MF signal sequence, and this effect was preserved across cultivation scales from shake flasks to 5-L bioreactors.

3. Materials and methods

3.1. Strains and plasmids

P. pastoris strains used in this work were derived from the strain X33 (Thermo Fisher, Invitrogen). Plasmids were created by In-Fusion cloning (TaKaRa/Clontech), and primers required for the plasmid construction were acquired from IDT. Expression of the genes encoding the three model proteins, E2-Crimson, BTL2, and ROL was driven by the methanol-inducible promoter P_{AOX1}, while expression of the gene encoding msGFP-HDEL was driven by the constitutive *KAR2* promoter. The model proteins were fused to either the full-length α-MF secretion signal consisting of the pre-α-MF signal sequence followed by the α-MF pro region with the Ser42 mutation [10], or a chimeric secretion signal consisting of the pre-Ost1 signal sequence followed by the same α-MF pro region. msGFP-HDEL was fused to the pre-Kar2 signal sequence. Genetic engineering procedures were designed and recorded by the use of SnapGene software (Insightful Science). All the plasmids used in this work, except the ROL plasmids, were described previously [10] and are available at Addgene.

Plasmids were linearized with the PmeI restriction enzyme and then transformed by electroporation using 100 ng of linear DNA, an amount that favored integration of a single copy of the expression cassette into the genome. Afterwards, screening was performed to detect positive colonies. Strains were grown and selected in a Yeast Extract - Peptone - Dextrose (YPD) rich medium, supplemented with either the antibiotic

Zeocin (100 µg/mL) or G418 (500 µg/mL). For each strain, 8 independent clones were tested but only 3 were selected for further shake flask-scale characterization, and just one was finally selected for further cultivation experiments. Single-copy integration was confirmed by droplet digital PCR, after purification of genomic DNA, using primers that flanked the integration locus. E2-Crimson, BTL2, and ROL single-copy integration at the *AOX1* locus was confirmed using primers 5'-GAAATAGACGCAGATCGGGAAC-3' and 5'-GAAGGTAGACCCATGGGTTGTTG-3'. The pre-Kar2-msGFP-HDEL construct was integrated at the *HIS4* locus by linearizing the plasmid with SalI, and single-copy integration was verified using primers 5'-GCTCTAGCCAGTTTGCTGTCCAAAC-3' and 5'-GGATGTTAGATGCCGGTTAGATC-3'.

YPD cultures for each strain were used to prepare 15% glycerol stock cell suspensions with final optical density (OD) at wavelength (λ) 600 nm equal to 60. These cell stocks were stored at -80°C.

3.2. Cultivation media

YPG pre-culturing medium contained per liter: 10 g yeast extract (Merck), 20 g peptone (Merck), and 20 g glycerol (PanReac AppliChem). This medium was supplemented with 100 µg/mL of Zeocin.

The basal salt synthetic medium for methanol batch and fed-batch cultivations contained per liter of distilled water: H_3PO_4 (85%), 26.7 mL; $CaSO_4$, 0.93 g; K_2SO_4, 18.2 g; $MgSO_4 \cdot 7H_2O$, 14.9 g; KOH, 4.13 g; glycerol, 40 g; biotin solution (200 mg/L), 4 mL; trace salts solution, 10 mL; and antifoam agent (A6426, Sigma–Aldrich Co., St. Louis, MO, USA), 0.5 mL. The trace salts solution contained per liter: $CuSO_4 \cdot 5H_2O$, 6.0 g; NaI, 0.08 g; $MnSO_4 \cdot H_2O$, 3.0 g; $Na_2MoO_4 \cdot 2H_2O$, 0.2 g; H_3BO_3, 0.02 g; $CoCl_2$, 0.5 g; $ZnCl_2$, 20.0 g; $FeSO_4 \cdot 7H_2O$, 65.0 g; biotin, 0.3 g; and concentrated H_2SO_4, 5 mL. The biotin and trace salts solutions were sterilized separately by filtration with a 0.22 µm pore filter (4433, PALL Life Sciences) or a 0.2 µm pore filter (17764-ACK0, Sartorius Stedim), respectively.

3.3. Pre-culture

Duplicate Erlenmeyer flask cultures (150 mL working volume) were prepared as follows: 150 mL of YPG medium containing 100 µg/mL of Zeocin were inoculated with defrosted cryo-stocks at an OD_{600} of 0.2 and incubated overnight at 30°C and 160 rpm (Infors Multitron Shaker, 25 mm shaking diameter). The pre-culture was then centrifuged at 6000 rpm for 10 min, and the harvested cells were resuspended in sterile water in a final volume of 100 mL and directly inoculated in the reactor. Methanol batches were inoculated at an initial OD_{600} of 2, while fed-batches were inoculated at initial OD_{600} of 1.

3.4. Bioreactor cultivation set-up and operational conditions

3.4.1. Methanol batch cultures

Batch cell cultures were carried out in a 2-L bioreactor controlled by the ez-Control system (Applikon Biotechnology B.V., Delft, NL) under the following cultivation conditions: working volume 1 L, temperature 30°C, pH 5.5 controlled by adding NH_4OH 30% (v/v), 1 vvm air flow and dissolved oxygen set-point at 30% air saturation controlled in cascade with agitation (from a minimum of 500 to a maximum of 700 rpm). Before starting the cultivation, 10 g/L of methanol were added, and samples were taken at different time points until the methanol was fully consumed. In table 1, $q_{p\ MAX}$ was calculated using the $Y_{P/X}$ and multiplied by μ_{MAX}. All methanol batches were performed in duplicates.

2.4.2. Fed-batch cultures

Fed-batch cultures were carried out in a 5-L bioreactor controlled by a Braun Biostat B system (Braun Biotech, Melsungen, Germany) under the following cultivation conditions: initial volume 2 L, temperature 30°C, pH 5.5 controlled by adding NH_4OH 30% (v/v) during the glycerol batch phase and KOH (5 M) during the transition phase and methanol induction phase, 2 L/min air flow and dissolved oxygen set at 30% air saturation controlled in cascade with agitation (from 900 to 1200 rpm during the glycerol batch phase and 1200 rpm during the transition phase and methanol induction phase), and air oxygen enrichment when required to maintain the culture at 30% oxygen.

First, the glycerol batch phase started with 40 g/L of glycerol. Second, when the glycerol was completely consumed, as detected by a sudden increase in the dissolved oxygen concentration (DO), a 5 h transition phase began [14]. Finally, the methanol induction phase started, in which methanol acted as the sole carbon source and as inducer simultaneously. The strategy was a non-limited fed-batch culture (MNLFB).. A predictive-PI control strategy was supplied to maintain a constant methanol concentration of 3 g/L [15].

In the transition phase, successive solutions of 250 mL of glycerol (50% v/v), and 1 L of pure methanol complemented with 10 mL of trace salts solution and 4 mL of biotin solution, were added as carbon sources. Transition phase started with 300 µL/min glycerol addition for 2 h. Afterwards, the methanol addition proportion was gradually increased using 160 µL/min of glycerol and 100 µL/min of methanol for the first hour, 100 µL/min of glycerol and 100 µL/min of methanol for the second hour, and 65 µL/min of glycerol and 100 µL/min of methanol for the third hour. During the methanol induction phase, methanol was the only carbon source. Glycerol solution and methanol were added using automatic microburettes S1 (Crison Instruments S.A., Alella, Barcelona, Spain). Previously, the microburette system had been sterilized by solutions of HCl 1 M, ethanol 70%, NaOH 1 M, and sterile water [16]. After transition, the nitrogen source was changed from 30% NH_4OH to NH_4Cl (100 g of NH_4Cl diluted in 500 mL of sterile water), because of interference of the NH_4OH with the methanol sensor. NH_4Cl solution was also complemented with 10 mL of trace salts solution and 4 mL of biotin solution. The NH_4Cl solution flow rate was directly linked to methanol addition by previously calculated NH_4Cl/methanol requirements (0.12 g/g) [17]. Furthermore, the pH was controlled after this point by KOH (5 M) addition. All cultivations ended when the culture reached a final biomass concentration of around 70 g/L of Dry Cell Weight (DCW), always below the maximal working volume (4 L). All the state variables and specific rates were calculated as previously described [18]. E2-Crimson fed-batch cultivation was performed in duplicate to obtain biological replicates for the fluorimetric and cytometry analyses shown in Figures 5 and 6.

3.5. Methanol measurement

Methanol concentration was monitored on-line using an immersed sensor in the culture (Raven Biotech, Vancouver, BC, Canada) [19]. The sensing element that detected the methanol was the Figaro TGS-822 (Figaro USA Inc., Glenview, USA) [20]. Methanol concentration was corroborated and/or corrected in the culture by HPLC off-line analyses, as reported elsewhere [19]. Relative standard deviation was below 5%.

3.6. Biomass analysis

Biomass concentration was measured as both dry cell weight (DCW) per liter of culture broth and OD_{600}. Samples were first centrifuged at 6000 rpm for 2 min (Espresso Microcentrifugue, Thermo Electron Corporation), to collect the biomass in a pellet. Pellets were washed and centrifuged twice in a 1 g/L citric acid and 9 g/L NaCl solution, and filtered through previously weighed 0.45 µm pore glass fiber filters (APPF04700, Merck Millipore Ltd.) by applying vacuum. Finally, filters were dried at 105°C for 24 h. OD_{600} and DCW measurements were performed in triplicate. Relative standard deviation was about 5%. In methanol batch experiments, cell growth was monitored by OD_{600} and then converted to DCW with a correlation set as 1 unit of DCW per 0.48 times the OD_{600}.

3.7. Heterologous protein production

3.7.1. Lipase activity assay

Enzyme activity assays were used to quantify protein secretion in the case of the BTL2 and ROL lipases. Activity was measured in triplicate using a lipase colorimetric assay, as previously described in [10]. Briefly, this assay consisted of mixing 0.5 mL of the suitably diluted culture supernatant from each strain with 0.5 mL of Tris-HCl buffer (200 mM, pH 7.25 for ROL or 100 mM, pH 7.25 for BTL2). The mixture was placed in a thermostatically controlled cuvette, at 30°C in the case of ROL or 50°C in the case of BTL2, which was pre-incubated for 2 min to reach the correct temperature. Then, 0.3 mL of substrate (Lipase assay reagent solution from Roche Diagnostics), containing 1,2-O-dilauryl-rac-glycero-3-glutaric-(methylresorufin)-ester) was added to the mixture and the absorbance was

measured at 580 nm for 7 min in a Specord 200 Plus Spectrophotometer (Analytic Jena). Data used to determine the activity were taken from minutes 3 to 5. The increased absorbance per second was used to calculate the activity, taking into account that one activity unit is defined as the amount of enzyme needed to hydrolyze 1 µmol of ester bond per minute. The relative standard deviation of the method was below 5%.

3.7.2. E2-Crimson fluorescence assay

Both extracellular and intracellular fluorescence were analyzed. Extracellular fluorescence was measured from the supernatant obtained after centrifugation at 6000 rpm in a microcentrifuge for 2 min. Intracellular fluorescence was measured from the solid fraction after centrifugation under the same conditions. The pellet corresponding to an OD_{600} of 1 in 1 mL was first washed with phosphate-buffered saline (PBS) and then resuspended in 1 mL of PBS. Samples were stored at -20°C until the time of analysis.

100 µL each of the extracellular and intracellular fractions were transferred to a 1-cm quartz cuvette, and fluorescence was measured in duplicate in a Varian Eclipse Fluorescence Spectrophotometer (Agilent Technologies). The excitation and emission wavelengths were 570 nm and 636 nm, respectively. Fluorescence intensity was measured in the same arbitrary units for all of the samples. Extracellular and intracellular fluorescence were measured in triplicate and relative standard deviation was about 5%.

3.8. Fluorescence microscopy

Images for Figure 3B and Figure 5C were captured as Z-stacks using an SP5 confocal microscope (Leica) equipped with a 1.4-NA/63× oil objective. Images for Figure 1B and Figure 3C were captured as Z-stacks using an Olympus Fluoview 1000 equipped with a UPlansApo/60x oil objective. The Z-stacks were average projected, and the brightness and contrast were adjusted evenly in all images. A Gaussian blur filter was used to smooth the red and green signals. Images were processed with Image J software (https://imagej.nih.gov/ij/).

3.9 Flow cytometry analysis

Cell size and viability were measured as described by Cámara and co-workers [21,22]. Intracellular ROS formation was measured using the dihydroethidium (DHE) fluorochrome. Samples were first sonicated and diluted the same way as when measuring cell size and viability. After the dilution, 0.2 mL of DHE solution (12.5 µg/mL in filtered PBS) was added to 1 mL of cell suspension. Then, samples were incubated at 30°C for 30 min with agitation. Finally, samples were centrifuged at 13,000 rpm for 3 min and resuspended in 1 mL of filtered PBS. In addition, a positive and negative control were prepared, in which the positive control was the same strain containing 20% v/v ethanol and the negative control sample lacked DHE.

4. Results

The pre-Ost1 signal sequence improves protein secretion and increases the maximum growth rate

We have previously shown that the pre-Ost1 signal sequence can drive passenger proteins through co-translational translocation into the ER lumen [10,12]. In shake flask cultures, this effect increased the final secreted protein titer in comparison with the commonly used pre-α-MF signal sequence [10]. However, it was unclear whether this improvement seen in small cultures could be scaled to the fermenter level, i.e., to bioprocess-relevant conditions. The goal of this work was to study how a secretion signal that replaces pre-α-MF with pre-Ost1 affects fermenter cultures in terms of product yields and cellular physiology.

The model proteins were BTL2 and E2-Crimson, which were previously examined in small-scale cultures [10], as well as the mature form of a *Rhizopus oryzae* lipase (ROL), which is structurally different from the other model proteins and has been shown to trigger a strong stress response [23] and to cause a significant metabolic burden even at relatively low production levels [24]. These proteins were fused to either a conventional secretion signal containing the pre-α-MF signal sequence or an improved secretion signal containing the pre-Ost1 signal sequence. Eight independent clones were evaluated for each strain, and the most representative clone was selected for subsequent batch and fed-batch cultivations.

As shown in **Figure 1A**, all of the proteins were secreted more efficiently in shake flask cultures when the strains carried the improved secretion signal. However, the magnitudes of the effects were protein dependent. Extracellular protein production was 6-fold higher for E2-Crimson, almost 3-fold higher for BTL2, and just 30% higher for ROL. For each strain, co-expression of GFP-HDEL was carried out to assess whether the heterologous proteins failed to cross the ER membrane and therefore clogged the ER translocons [10]. As shown in **Figure 1B**, when the three model proteins were fused to the conventional secretion signal containing pre-α-MF, GFP-HDEL accumulated in the cytosol, indicating that the model proteins had clogged the ER translocons. By contrast, when E2-Crimson and BTL2 were fused to the improved secretion signal containing pre-

Ost1, GFP-HDEL gave a typical ER pattern similar to that seen in the control strain, indicating that these model proteins had not clogged the ER translocons and therefore had presumably traversed the ER membrane. When ROL was fused to the improved secretion signal containing pre-Ost1, GFP-HDEL exhibited a peculiar ER pattern that likely reflected some type of ER stress response, but once again there was very little GFP-HDEL fluorescence in the cytosol. These observations support the idea that the pre-Ost1 signal sequence can improve the ability of multiple passenger proteins to traverse the ER membrane, but that the downstream effect on secretion depends on other properties of the passenger protein.

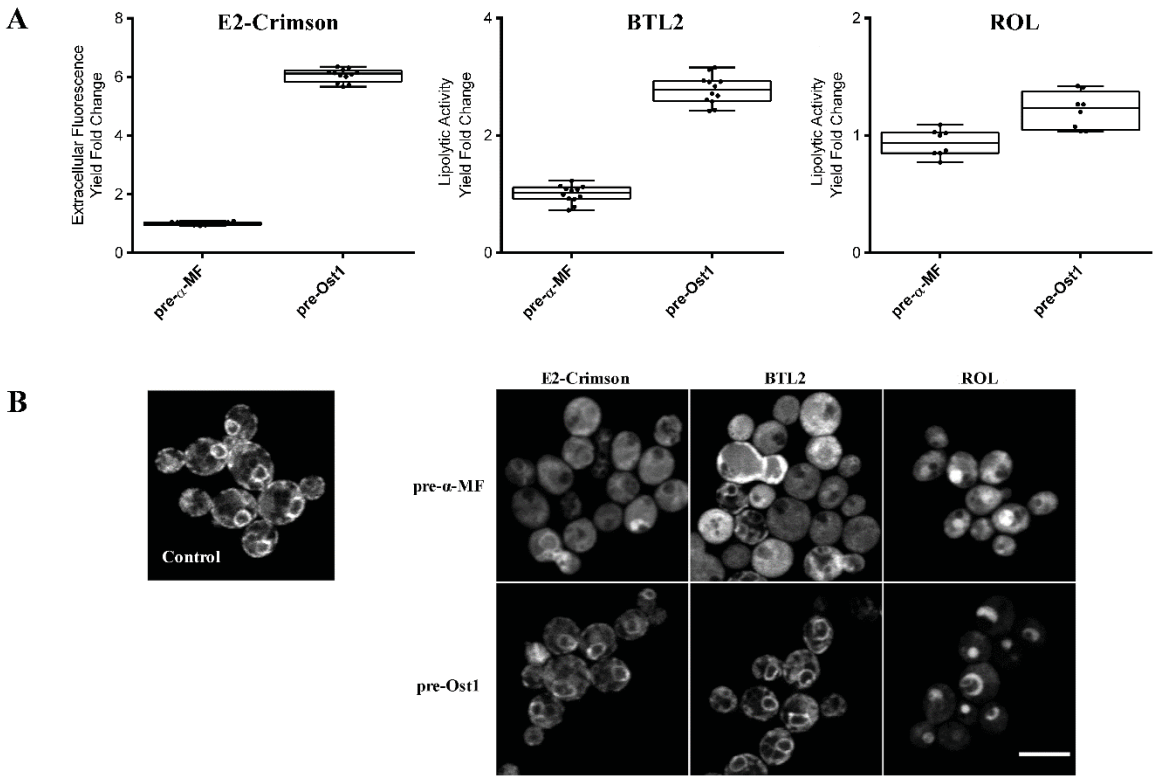

Figure 1. Effects of the different signal sequences in shake-flask cultures. (A) Box plot representing extracellular protein secretion ± standard deviation (SD) in shake-flask cultures for E2-Crimson, BTL2, and ROL, comparing secretion signals containing the pre-α-MF and pre-Ost1 signal sequences. Eight clones were evaluated for each strain. Box plots for E2-Crimson and BTL2 were obtained from [10]. The activity/fluorescence values were normalized by setting the value for the pre-α-MF reference strain to 1. **(B)** Fluorescence images of ER-targeted GFP-HDEL are shown. In the "Control" sample, no model protein was expressed. Scale bar, 5 µm.

To further study the benefits of using the pre-Ost1 signal sequence, batch cultivations with an initial methanol concentration of 10 g/L were performed for all of the strains, including the X33 wild-type control strain. Cell growth, methanol concentration, and extracellular protein concentration are represented in **Figure 2,** and related bioprocess parameters are listed in **Table 1**.

Heterologous protein production has often been related to a reduction in the specific growth rate (μ) and biomass yield [25,26]. In particular, this correlation has been previously observed for ROL [24]. Notably, such a negative effect was significantly diminished when the secretion signal contained the pre-Ost1 signal sequence instead of the pre-α-MF signal sequence **(Table 1)**. Moreover, use of the pre-Ost1 signal sequence notably increased both the protein titers and the maximum specific protein production rate (q_p). As would be expected, the methanol specific consumption rate (q_s) was also higher with use of the pre-Ost1 signal sequence in accordance with the increase of μ_{MAX} **(Table 1)**.

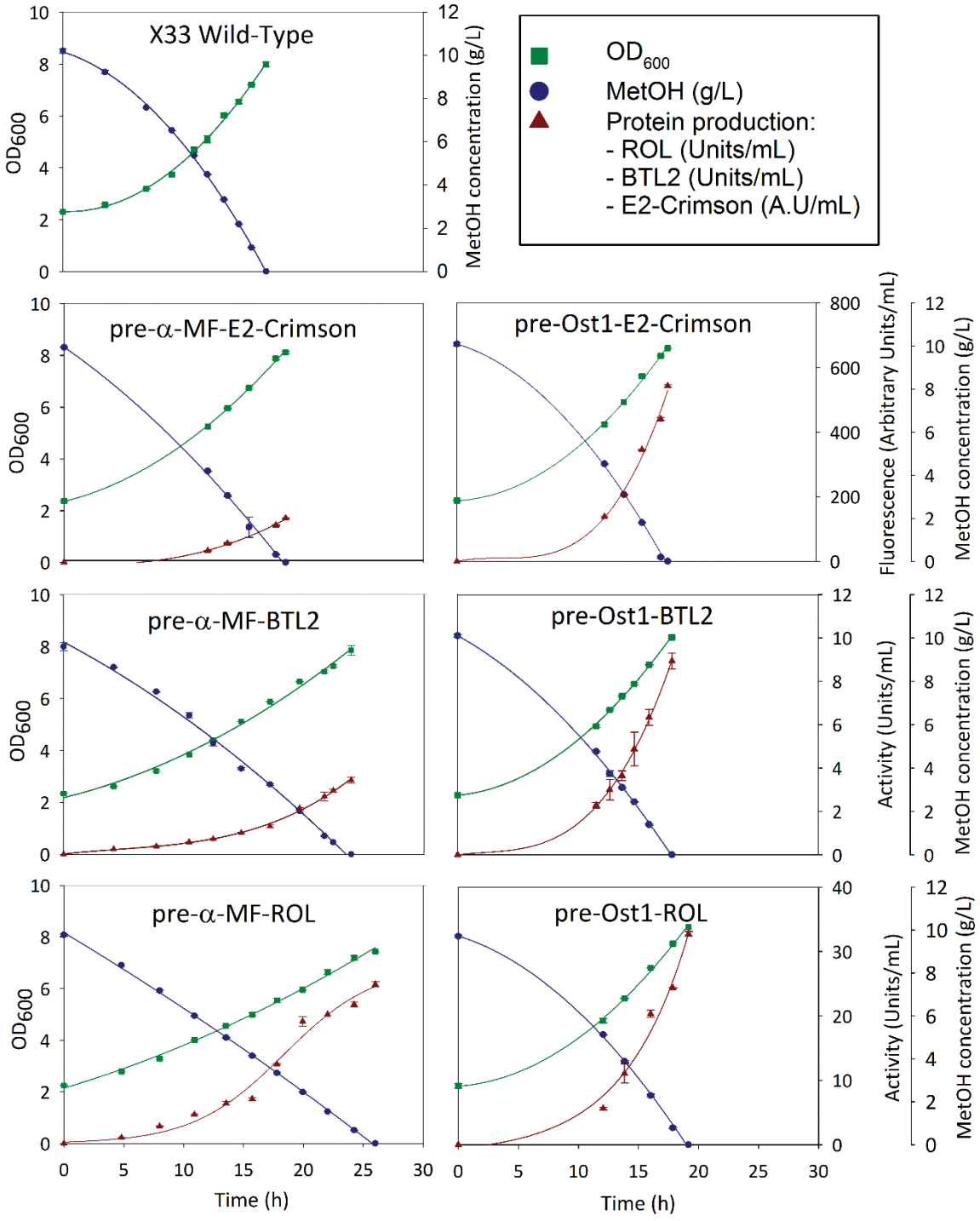

Figure 2. Results from methanol batch cultivations. Biomass (OD_{600}), concentration of methanol (MetOH), and protein production result s± SD are represented for each cultivation over time. Seven batches are shown, one for each model protein (E2-Crimson, BTL2, and ROL) and signal sequence (pre-α-MF and pre-Ost1) plus the X33 wild-type control strain.

	X33	E2-Crimson		BTL2		ROL	
	Wild-Type	pre-α-MF	pre-Ost1	pre-α-MF	pre-Ost1	pre-α-MF	pre-Ost1
Final protein *	-	879.65	3096.2	107.0	259.9	192.4	291.9
μ_{mean} (h^{-1})	0.081	0.050	0.050	0.047	0.072	0.026	0.040
q_s (gMetOH)/gDCW*h)	0.30	0.20	0.19	0.18	0.25	0.11	0.14
q_p (protein/gDCW*h)	-	740.0	2357	81.6	255.4	74.9	155.0
$Y_{(x/s)}$ (gDCW/gMetOH))	0.28	0.25	0.25	0.26	0.27	0.25	0.27
$Y_{(P/X)}$ (Total protein/gDCW)	-	13620	44962	1739	3487	2863	3873
Volumetric productivity (Total protein/(L*h))	-	19465	66250	2035	6047	2681	4906
Specific productivity (Total protein/(gDCW*h))	-	253	901	28.7	77.6	37.0	63.7

* Protein Units: E2-Crimson: (A.U./mL)
BTL2: (U.A./mL)
ROL: (U.A./mL)

Table 1. Methanol batch bioprocess parameters. Listed are values of maximum specific growth rate (μ_{MAX}), specific methanol (MetOH) consumption rate (q_s), specific protein production rate ($q_{p\ MAX}$), protein/biomass yield ($Y_{(p/x)}$), and biomass/methanol yield ($Y_{(x/s)}$) (±SD). Biomass results used in calculations were measured as dried cell weight (DCW). These parameters were calculated from cultivations performed in duplicate. E2-Crimson levels were measured in arbitrary units (A.U.) of fluorescence, and BTL2 and ROL levels were measured in units of lipase activity (U.A.).

E2-Crimson accumulates in the ER when using the pre-Ost1 signal sequence in a methanol batch cultivation

To track the possible accumulation of a model protein inside the cells during a methanol batch cultivation, E2-Crimson production by the pre-Ost1-E2-Crimson and pre-α-MF-E2-Crimson strains was monitored over time by measuring intracellular fluorescence levels (**Figure 3A**). In both strains, intracellular fluorescence increased during the cultivation period, but surprisingly, the amount of intracellular fluorescence was significantly higher when E2-Crimson was fused to the pre-Ost1 signal sequence. Insight into this phenomenon came from capturing cell images by fluorescence microscopy at the end of the batch cultivation (**Figure S1**). With the pre-α-MF-E2-Crimson strain, fluorescent rings were visible in some cells. This effect was previously observed for shake-flask cultures, and because GFP-HDEL accumulated in the cytosol in those cultures, we inferred that E2-Crimson became stuck during passage into the ER and clogged the translocons [10]. After a batch cultivation of the pre-α-MF-E2-Crimson strain, a similar cytosolic accumulation of GFP-HDEL was apparent, suggesting that E2-Crimson was once again becoming stuck during passage into the ER (**Figure 3B**). With the pre-Ost1-E2-Crimson strain, fluorescent rings were also visible after a batch cultivation (**Figure S1**). However, those E2-Crimson rings colocalized with GFP-HDEL in a normal ER pattern (**Figure 3B**), suggesting that the ER translocons were not clogged and that E2-Crimson was accumulating inside the ER. A likely interpretation is that under methanol batch conditions, the pre-Ost1 signal sequence continues to drive translocation across the ER membrane, but the resulting high level of E2-Crimson in the ER creates a new bottleneck at the level of ER export.

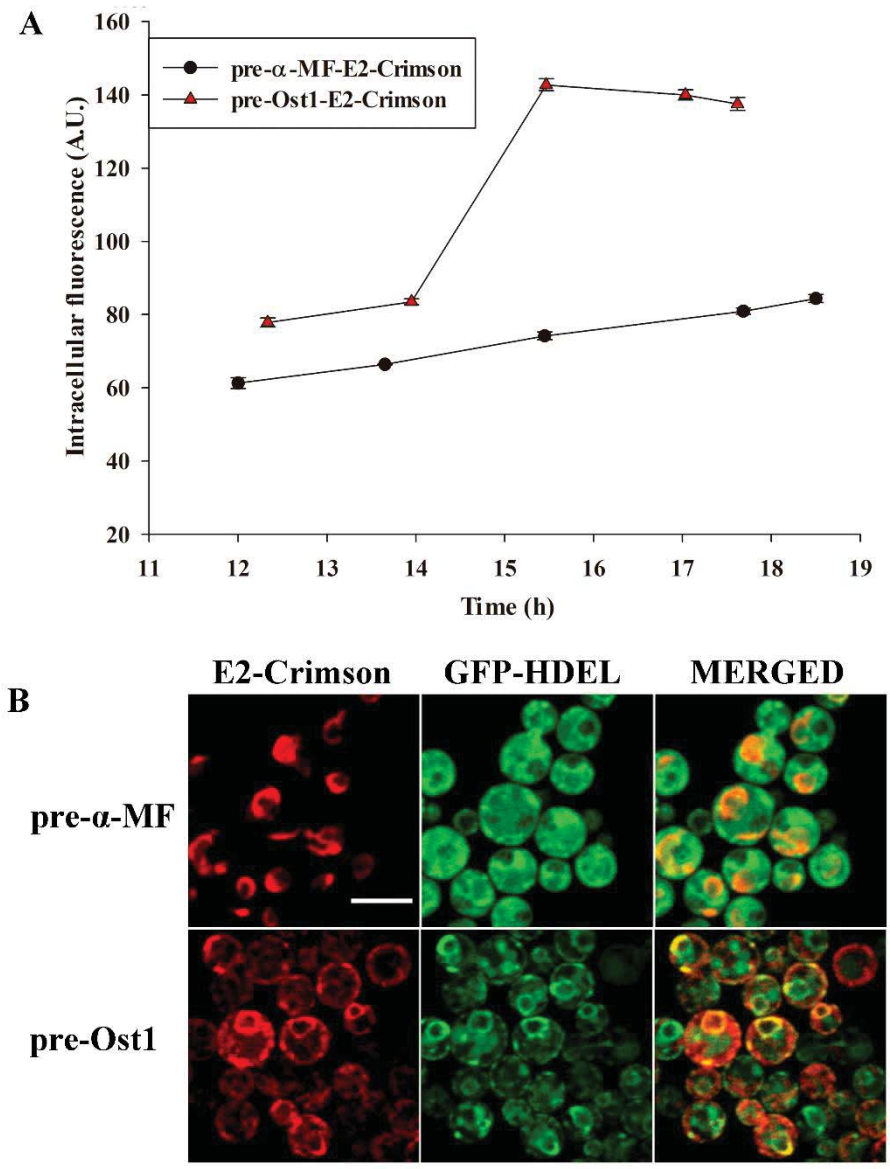

Figure 3. Intracellular E2-Crimson fluorescence in a methanol batch cultivation. (A) Evolution of intracellular E2-Crimson fluorescence (± SD) during batch cultivation with the different signal sequences. The data are plotted in arbitrary units (A.U.) and represent the amount of fluorescence per 1 unit of OD_{600}. **(B)** Images of intracellular fluorescence with the different signal sequences at the end of each batch cultivation containing the ER-targeted GFP-HDEL. Left: projected confocal Z-stacks of E2-Crimson fluorescence (red). Center: projected confocal Z-stacks of the GFP-HDEL. Right: The previous images merged. Scale bar, 5 µm.

Secretion of model proteins in fed-batch cultivations is enhanced by using the pre-Ost1 signal sequence

We next performed fed-batch cultivations to test how the cultures behaved under bioprocess-like conditions, i.e., at high cell densities. The methanol level was maintained at 3 g/L. Evolution of biomass and the concentrations of the secreted model proteins during the fed-batch cultivations are shown in **Figure 4**, and an X33 wild-type cultivation is shown in **Figure S2**. The bioprocess parameters are listed in **Table 2**. Each cultivation was run until the biomass concentration reached around 70 g/L of DCW.

Contrary to what was observed in E2-Crimson batch cultivations, the mean specific growth rate was maintained similarly when producing E2-Crimson with either the pre-Ost1 signal sequence or the pre-α-MF signal sequence (**Figure 4A**) at this methanol concentration set-point. However, when comparing secreted protein titers at this methanol concentration, extracellular fluorescence was 3.5-fold higher with the pre-Ost1-E2-Crimson strain quite similar to the value obtained in batch (3.8-fold higher). Moreover, volumetric productivity, specific productivity and $Y_{P/X}$ were, respectively, 3.4-fold, 3.6-fold, and 3.3-fold higher with the pre-Ost1-E2-Crimson strain compared to the pre-α-MF-E2-Crimson strain (**Table 2**).

For the BTL2 cultivations, the maximum targeted biomass concentration was reached in a shorter cultivation time with the pre-Ost1-BTL2 strain than with the pre-α-MF-BTL2 strain (**Figure 4B**), yielding a higher mean μ (**Table 2**). Regarding protein secretion, extracellular lipase activity was increased by 2.4-fold when using pre-Ost1 as the signal sequence, while volumetric productivity and specific productivity were 3.0-fold and 2.7-fold higher, respectively (**Table 2**).

For the ROL cultivations, the pre-Ost1-ROL strain gave a 1.5-fold increase in the final product titer compared to the pre-α-MF-ROL strain (**Figure 4C**). Because the growth rate of the pre-Ost1-ROL strain was significantly higher, the volumetric productivity and specific productivity were increased by almost 2-fold (**Table 2**).

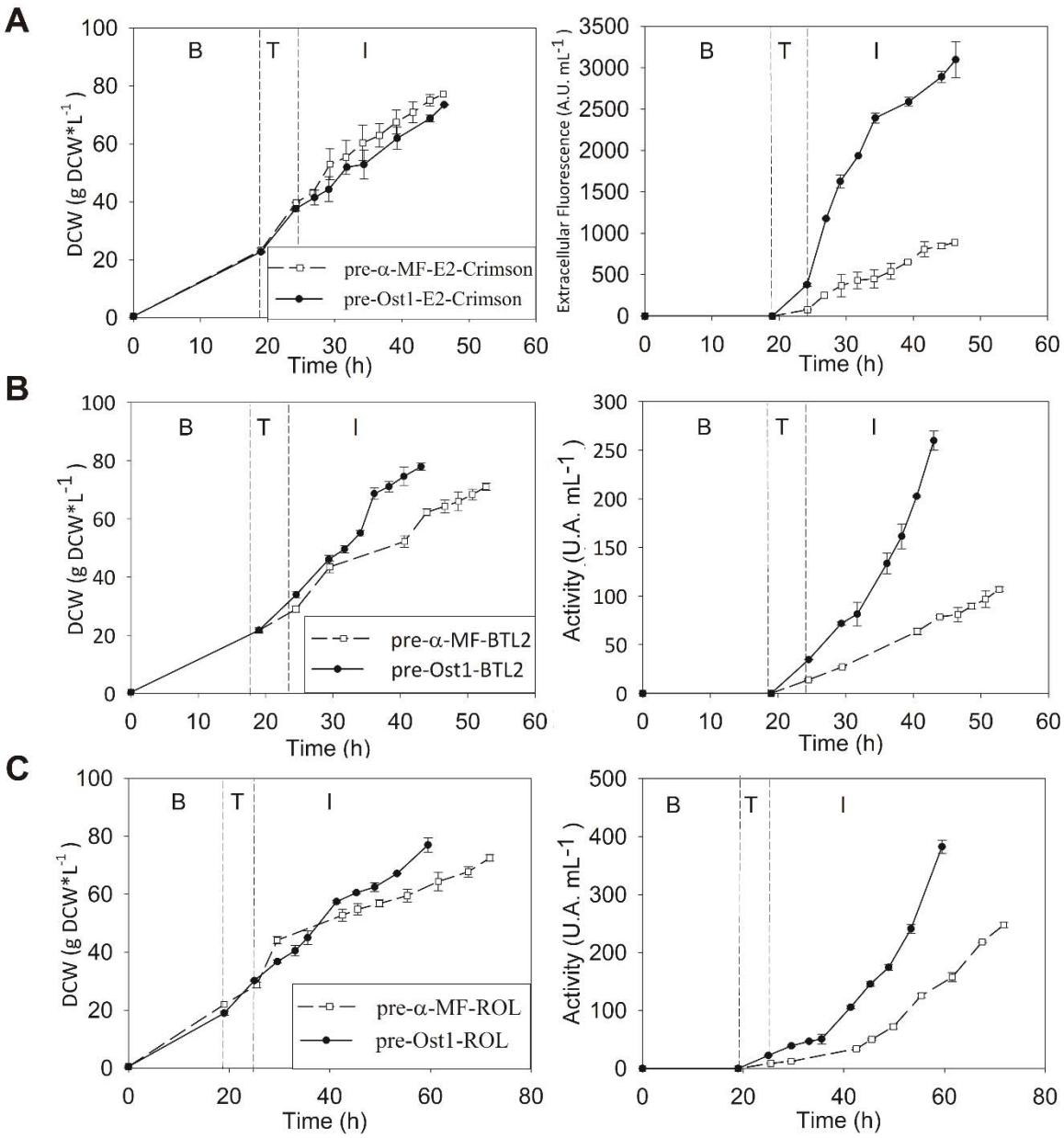

Figure 4. Results from fed-batch cultivations. Cell growth (left graphs) and fluorescence or lipolytic activity (right graphs) ± SD are plotted over time for the different signal sequences for **(A)** E2-Crimson, **(B)** BTL2, and **(C)** ROL. The letter symbols represent glycerol batch phase (B), transition phase (T), and methanol induction phase (I). A.U., arbitrary units; U.A., units of lipase activity. Error bars in E2-Crimson correspond to the differences between biological replicates and error bars in BTL2 and ROL represent the difference between technical triplicates.

	X33	E2-Crimson		BTL2		ROL	
	Wild-Type	**pre-α-MF**	**pre-Ost1**	**pre-α-MF**	**pre-Ost1**	**pre-α-MF**	**pre-Ost1**
Final protein *	-	879.65	3096.2	107.0	259.9	192.4	291.9
μ_{mean} (h^{-1})	0.081	0.050	0.050	0.047	0.072	0.026	0.040
q_s (gMetOH)/gDCW*h)	0.30	0.20	0.19	0.18	0.25	0.11	0.14
q_p (protein/gDCW*h)	-	740.0	2357	81.6	255.4	74.9	155.0
$Y_{(x/s)}$ (gDCW/gMetOH))	0.28	0.25	0.25	0.26	0.27	0.25	0.27
$Y_{(P/X)}$ (Total protein/gDCW)	-	13620	44962	1739	3487	2863	3873
Volumetric productivity (Total protein/(L*h))	-	19465	66250	2035	6047	2681	4906
Specific productivity (Total protein/(gDCW*h))	-	253	901	28.7	77.6	37.0	63.7

* Protein Units: E2-Crimson: (A.U./mL)
BTL2: (U.A./mL)
ROL:(U.A./mL)

Table 2. Fed-batch bioprocess parameters. Listed are values from wild-type (X33), E2-Crimson, BTL2, and ROL fed-batch cultivations for specific growth rate (μ), specific methanol (MetOH) consumption (q_s), specific protein production (q_p), protein/biomass yield ($Y_{(p/x)}$), biomass/methanol yield ($Y_{(x/s)}$), volumetric productivity, and specific productivity. Biomass results used in calculations were measured as dried cell weight (DCW). E2-Crimson levels were measured in arbitrary units (A.U.) of fluorescence, and BTL2 and ROL levels were measured in units of lipase activity (U.A.).

pre-Ost1 efficiently improves protein translocation but its accumulation at the ER level triggers ER stress

Building on the results with the methanol batch cultivations, the strains expressing E2-Crimson were examined during the fed-batch cultivations. In the transition phase, when addition of methanol started, intracellular fluorescence started to increase for both the pre-α-MF-E2-Crimson strain and the pre-Ost1-E2-Crimson strain. However, in the induction phase, intracellular fluorescence evolution was different with the two signal sequences (**Figure 5A**). In the pre-α-MF-E2-Crimson fed-batch cultivation, intracellular fluorescence increased during the first 12 h of methanol feeding, and then remained stable until the end of the cultivation. On the other hand, in the pre-Ost1-E2-Crimson cultivation, intracellular fluorescence initially accumulated much faster, but then decreased after 10 h of methanol feeding. These results agree with fluorescence microscopy images (**Figure 5B**), which revealed intracellular fluorescent rings similar to those seen after the methanol batch cultivations, and also confirmed that intracellular fluorescence was notably higher for the pre-α-MF-E2-Crimson strain at the end of the fed-batch cultivation.

We also measured average cell size, ROS accumulation, and cell viability during fed-batch cultivations of the pre-α-MF-E2-Crimson and pre-Ost1-E2-Crimson strains as well as the X33 reference strain (**Figure 6**). These parameters are good indicators of the physiological state of the cells. It has been reported that in methanol fed-batch cultivations of recombinant *P. pastoris*, metabolic stress increases while cell viability often decreases [27–29]. Intracellular protein accumulation is one of the main causes of metabolic stress, which can compromise cellular viability [29–31], and cell size and ROS levels are metabolic stress indicators [32,33]. In our analysis, cells size tended to increase during the cultivation period for all the strains (**Figure 6A**). However, such increases were smaller in the X33 wild-type strain than in the strains expressing E2-Crimson, and slightly smaller in the pre-Ost1-E2-Crimson strain than in the pre-α-MF-E2-Crimson strain. ROS levels also increased over time in all of the strains, and this effect was particularly increased in the strains expressing E2-Crimson (**Figure 6B**). In the pre-α-MF-E2-Crimson strain, the percentage of cells with high ROS levels gradually increased until reaching 33% of the population at the end of the fed-batch cultivation. Interestingly,

with the pre-Ost1-E2-Crimson strain, the percentage of cells with high ROS levels transiently reached higher levels than with the pre-α-MF-E2-Crimson strain after the transition phase and at the beginning of the induction phase, but started to decrease 15 h after, reaching significantly lower values than those observed for the pre-α-MF-E2-Crimson strain towards the end of the induction phase (**Figure 6B**). Finally, cell viability was significantly higher in the X33 reference strain than in E2-Crimson-producing strains, and slightly higher in the pre-Ost1-E2-Crimson strain than in the pre-α-MF-E2-Crimson strain (**Figure 6C**). Taken together, these results indicate that expression of E2-Crimson stresses the cells, and that the improved secretion signal may somewhat alleviate the stress during long fed-batch cultivations.

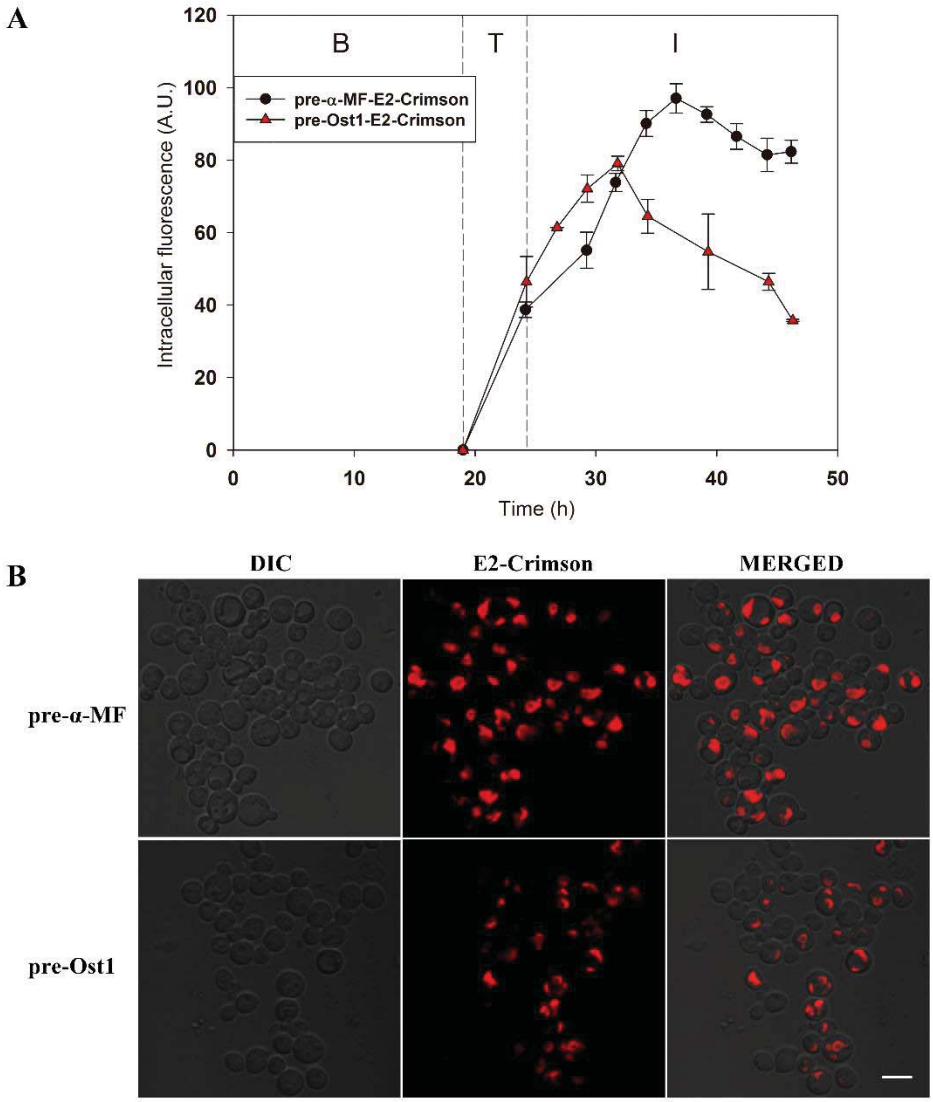

Figure 5. Intracellular E2-Crimson fluorescence in fed-batch cultivations. (A) Evolution of intracellular E2-Crimson fluorescence (± SD) during fed-batch cultivation with the different signal sequences. The letters represent glycerol batch phase (B), transition phase (T), and methanol induction phase (I). The data represent the amount of fluorescence per 1 unit of OD_{600}. Error bars represent the differences between biological replicates. **(B)** Images of intracellular fluorescence with the different signal sequences at the end of each fed-batch cultivation. Projected confocal Z-stacks of E2-Crimson fluorescence (red) were merged with differential interference contrast (DIC) images of the cells. Scale bar, 5 µm.

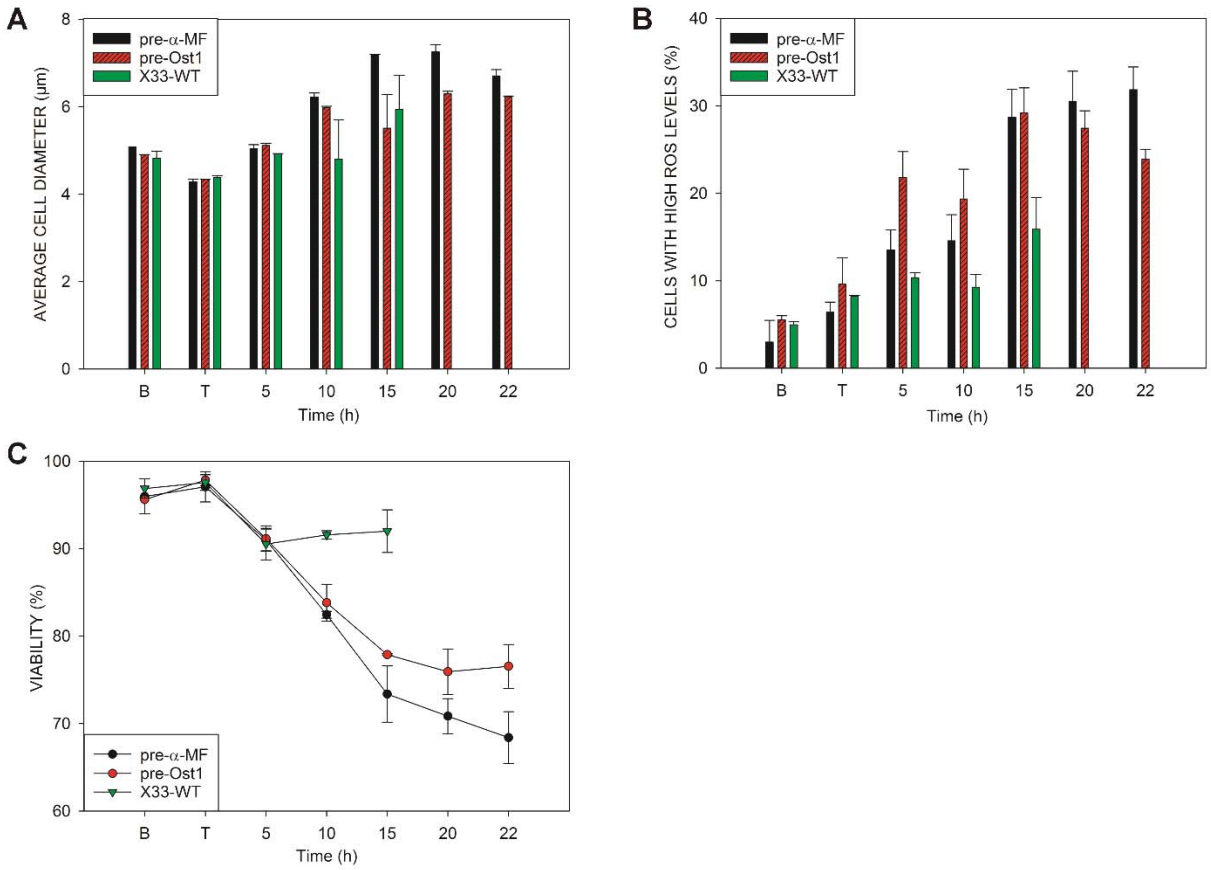

Figure 6. Effects of the signal sequences on cell size, ROS, and cell viability in fed-batch cultivations. Results are shown from fed-batch cultivations expressing E2-Crimson with either the pre-α-MF or the pre-Ost1 signal sequence (Figure 4A) as well as the X33 wild-type fed-batch cultivation (Figure S2). The letters and numbers represent glycerol batch phase (B), transition phase (T), and the time of methanol induction (numbers). **(A)** Average cell diameter (± SD). **(B)** Percentage of cells showing ROS levels (± SD). **(C)** Percentage of viable cells (± SD). Error bars represents de difference between biological replicates.

5. Discussion

In this study, we have extended the characterization of an improved chimeric secretion signal containing the pre-Ost1 signal sequence, which drives proteins across the ER membrane of *P. pastoris* by co-translational translocation [10]. The model proteins were E2-Crimson, BTL2, and ROL. These proteins were selected for the following reasons. E2-Crimson is a fluorescent protein that enables the visualization of intracellular trafficking and was the model protein of our previous work. BTL2 is a lipase of industrial interest that has been reported to fold rapidly, and its secretion can be dramatically affected by

the choice of secretion signal [34]. ROL is a lipase and a promising industrial biocatalyst [35,36] that we have used as a reporter protein in previous studies, and it was shown to trigger ER stress and cause a significant metabolic burden [23,24,37], making it an appealing choice to test whether the pre-Ost1 signal sequence can reduce cellular stress/metabolic burden in yeast. Each of these model proteins was fused to the pre-Ost1 signal sequence as well as the pre-α-MF signal sequence as a reference.

Each strain was tested in three different ways. First, a shake-flask culture was used to determine how the signal sequence affected the secretory pathway and the final levels of secretion. Second, a bioreactor batch culture was set up to study cell growth and product secretion kinetics under controlled conditions. Third, a fed-batch cultivation was carried out to see how the cells behaved at high densities under bioprocess-relevant conditions.

In the shake-flask cultures, all three model proteins showed significantly enhanced secretion with the pre-Ost1 signal sequence. However, the improvement levels varied obtaining the maximum improvement with E2-Crimson and the lowest with ROL. This result suggests that the importance of the ER translocation step depends on the protein being expressed.

Similar results were obtained in the bioreactor batch cultures. Importantly, growth rates were significantly superior in the strains carrying the pre-Ost1 secretion signal, leading to increased final productivities.

In the fed-batch cultivations, the pre-Ost1 signal sequence once again yielded increased protein productivities due to higher protein titers combined with faster growth rates that allowed earlier termination. However, the growth rates in the fed-batch cultivations were lower than in the bioreactor batch cultures, particularly for the strains expressing ROL. This reduced growth might be related to cellular stress due to extended time under inducing conditions, high cell densities, and high-level expression of proteins with cytotoxic properties.

One curious result was that in the E2-Crimson fed-batch cultivations, the strain carrying the pre-Ost1 signal sequence grew similarly to the strain carrying the pre-α-MF signal sequence. Under these conditions, E2-Crimson showed an ER accumulation that may

have produced a metabolic stress. Early in the cultivation, this ER accumulation was more pronounced with the pre-Ost1 signal sequence than with the pre-α-MF signal sequence, but after 10 h of induction, the ER accumulation seen with the pre-Ost1 signal sequence diminished and cellular ROS levels also decreased. A likely explanation is that accumulation of E2-Crimson inside the ER upon induction stressed the cells, which eventually responded by producing more chaperones and/or upregulating the ER degradation machinery. Notably, pre-Ost1-E2-Crimson cells seem to better cope with stress after such initial response. However, the cells were adapting to the conditions too late, resulting in a failure to increase the growth rate. This phenomenon bears further investigation.

We note that our test strains each contained a single copy of the gene to be expressed, whereas multiple copies of integrated genes are often used to maximize protein production [21,38,39]. Even if efficient protein translocation across the ER is achieved by using the pre-Ost1 signal sequence, very high expression levels could create a new bottleneck at the level of ER export. A potential strategy for addressing this issue is to overexpress the Erv29 protein, which serves as a receptor that recognizes an ER export signal in the pro region of the secretion signal [40–42]. Thus, further improvements may be possible due to continued rational engineering of the *P. pastoris* secretory pathway.

6. Conclusions

We show here that the pre-Ost1 signal sequence can improve heterologous protein secretion at a bioreactor scale by driving the proteins through co-translational translocation instead of post-translational translocation. Thus, pre-Ost1 has great potential to replace pre-α-MF at an industrial scale. Efficient co-translational translocation seems to reduce the metabolic burden for some model proteins, probably by ensuring that a larger fraction of the protein molecules correctly enter the secretory pathway. Further studies should provide insights into the physiological impact of this secretion strategy. To fully exploit the power of the improved secretion signal, additional cellular engineering may be needed to overcome bottlenecks that can appear

downstream of the translocation event, particularly under bioprocess-type (fed-batch) conditions.

7.References

[1] Brake AJ, Merryweather JP, Coit DG, Heberlein UA, Masiarz FR, Mullenbach GT, et al. α-Factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 1984;81:4642–6. https://doi.org/10.1073/pnas.81.15.4642.

[2] Otte S, Barlowe C. Sorting signals can direct receptor-mediated export of soluble proteins into COPII vesicles. Nat Cell Biol 2004;6:1189–94. https://doi.org/10.1038/ncb1195.

[3] Dancourt J, Barlowe C. Protein Sorting Receptors in the Early Secretory Pathway. Annu Rev Biochem 2010;79:777–802. https://doi.org/10.1146/annurev-biochem-061608-091319.

[4] Fuller RS, Sterne RE, Thorner J. Enzymes required for yeast prohormone processing. Annu Rev Physiol 1988;50:345–62. https://doi.org/10.1146/annurev.physiol.50.1.345.

[5] Looser V, Bruhlmann B, Bumbak F, Stenger C, Costa M, Camattari A, et al. Cultivation strategies to enhance productivity of Pichia pastoris: A review. Biotechnol Adv 2015;33:1177–93. https://doi.org/10.1016/j.biotechadv.2015.05.008.

[6] Puxbaum V, Mattanovich D, Gasser B. Quo vadis? The challenges of recombinant protein folding and secretion in Pichia pastoris. Appl Microbiol Biotechnol 2015;99:2925–38. https://doi.org/10.1007/s00253-015-6470-z.

[7] Lin-Cereghino GP, Stark CM, Kim D, Chang J, Shaheen N, Poerwanto H, et al. The effect of α-mating factor secretion signal mutations on recombinant protein expression in Pichia pastoris. Gene 2013;519:311–7. https://doi.org/10.1016/j.gene.2013.01.062.

[8] Ahn J, Jang MJ, Ang KS, Lee H, Choi ES, Lee DY. Codon optimization of Saccharomyces cerevisiae mating factor alpha prepro-leader to improve recombinant protein production in Pichia pastoris. Biotechnol Lett 2016;38:2137–43. https://doi.org/10.1007/s10529-016-2203-3.

[9] Joo HH, Xue L, Tsai JW, Park SPJ, Kwon J, Patel A, et al. Structural characterization of the α-mating factor prepro-peptide for secretion of recombinant proteins in Pichia pastoris. Gene 2017;598:50–62. https://doi.org/10.1016/j.gene.2016.10.040.

[10] Barrero JJ, Casler JC, Valero F, Ferrer P, Glick BS. An improved secretion signal enhances the secretion of model proteins from Pichia pastoris. Microb Cell Fact 2018;17:1–13. https://doi.org/10.1186/s12934-018-1009-5.

[11] Ast T, Michaelis S, Schuldiner M, States U. The protease Ste24 clears clogged translocons. Cell 2017;164:103–14. https://doi.org/10.1016/j.cell.2015.11.053.The.

[12] Fitzgerald I, Glick BS. Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting. Microb Cell Fact 2014;13:125. https://doi.org/10.1186/s12934-014-0125-0.

[13] Willer M, Forte GMA, Stirling CJ. Sec61p is required for ERAD-L: Genetic dissection of the translocation and ERAD-L functions of Sec61P using novel derivatives of CPY. J Biol Chem 2008;283:33883–8. https://doi.org/10.1074/jbc.M803054200.

[14] Cos O, Serrano A, Montesinos JL, Ferrer P, Cregg JM, Valero F. Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in Pichia pastoris fed-batch cultures. J Biotechnol 2005;117:321–35. https://doi.org/10.1016/j.jbiotec.2004.12.010.

[15] Cos O, Ramon R, Montesinos JL, Valero F. A simple model-based control for Pichia pastoris allows a more efficient heterologous protein production bioprocess. Biotechnol Bioeng 2006;95:145–54. https://doi.org/10.1002/bit.21005.

[16] Arnau C, Ramon R, Casas C, Valero F. Optimization of the heterologous production

of a Rhizopus oryzae lipase in Pichia pastoris system using mixed substrates on controlled fed-batch bioprocess. Enzyme Microb Technol 2010;46:494–500. https://doi.org/10.1016/j.enzmictec.2010.01.005.

[17] 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. Microb Cell Fact 2006;5:1–20. https://doi.org/10.1186/1475-2859-5-17.

[18] Barrigón JM, Montesinos JL, Valero F. Searching the best operational strategies for Rhizopus oryzae lipase production in Pichia pastoris Mut + phenotype: Methanol limited or methanol non-limited fed-batch cultures? Biochem Eng J 2013;75:47–54. https://doi.org/10.1016/j.bej.2013.03.018.

[19] Arnau C, Casas C, Valero F. The effect of glycerol mixed substrate on the heterologous production of a Rhizopus oryzae lipase in Pichia pastoris system. Biochem Eng J 2011;57:30–7. https://doi.org/10.1016/j.bej.2011.08.004.

[20] Valero F, Ramon R, Feliu JX, Berthet FX, Montesinos JL, Cos O. Improving the monitoring of methanol concentration during high cell density fermentation of Pichia pastoris. Biotechnol Lett 2004;26:1447–52. https://doi.org/10.1023/b:bile.0000045649.60508.c5.

[21] Cámara E, Albiol J, Ferrer P. Droplet digital PCR-aided screening and characterization of Pichia pastoris multiple gene copy strains. Biotechnol Bioeng 2016;113:1542–51. https://doi.org/10.1002/bit.25916.

[22] Cámara E, Monforte S, Albiol J, Ferrer P. Deregulation of methanol metabolism reverts transcriptional limitations of recombinant Pichia pastoris (Komagataella spp) with multiple expression cassettes under control of the AOX1 promoter. Biotechnol Bioeng 2019;116:1710–20. https://doi.org/10.1002/bit.26947.

[23] Resina D, Maurer M, Cos O, Arnau C, Carnicer M, Marx H, et al. Engineering of bottlenecks in Rhizopus oryzae lipase production in Pichia pastoris using the nitrogen source-regulated FLD1 promoter. N Biotechnol 2009;25:396–403. https://doi.org/10.1016/j.nbt.2009.01.008.

[24] Jordà J, Jouhten P, Cámara E, Maaheimo H, Albiol J, Ferrer P. Metabolic flux profiling of recombinant protein secreting Pichia pastoris growing on glucose:methanol mixtures. Microb Cell Fact 2012;11:57. https://doi.org/10.1186/1475-2859-11-57.

[25] Yu XW, Sun WH, Wang YZ, Xu Y. Identification of novel factors enhancing recombinant protein production in multi-copy Komagataella phaffii based on transcriptomic analysis of overexpression effects. Sci Rep 2017;7:16249. https://doi.org/10.1038/s41598-017-16577-x.

[26] Heyland J, Fu J, Blank LM, Schmid A. Carbon metabolism limits recombinant protein production in Pichia pastoris. Biotechnol Bioeng 2011;108:1942–53. https://doi.org/10.1002/bit.23114.

[27] Vanz AL, Lünsdorf H, Adnan A, Nimtz M, Gurramkonda C, Khanna N, et al. Physiological response of Pichia pastoris GS115 to methanol-induced high level production of the Hepatitis B surface antigen: Catabolic adaptation, stress responses, and autophagic processes. Microb Cell Fact 2012;11:103. https://doi.org/10.1186/1475-2859-11-103.

[28] Zepeda AB, Figueroa CA, Pessoa A, Farías JG. Free fatty acids reduce metabolic stress and favor a stable production of heterologous proteins in Pichia pastoris. Brazilian J Microbiol 2018;49:856–64. https://doi.org/10.1016/j.bjm.2018.03.008.

[29] Surribas A, Stahn R, Montesinos JL, Enfors SO, Valero F, Jahic M. Production of a Rhizopus oryzae lipase from Pichia pastoris using alternative operational strategies. J Biotechnol 2007;130:291–9. https://doi.org/10.1016/j.jbiotec.2007.04.009.

[30] Hohenblum H, Borth N, Mattanovich D. Assessing viability and cell-associated product of recombinant protein producing Pichia pastoris with flow cytometry. J Biotechnol 2003;102:281–90. https://doi.org/10.1016/S0168-1656(03)00049-X.

[31] Hohenblum H, Gasser B, Maurer M, Borth N, Mattanovich D. Effects of Gene Dosage, Promoters, and Substrates on Unfolded Protein Stress of Recombinant

Pichia pastoris. Biotechnol Bioeng 2004;85:367–75. https://doi.org/10.1002/bit.10904.

[32] Zhong Y, Yang L, Guo Y, Fang F, Wang D, Li R, et al. High-temperature cultivation of recombinant Pichia pastoris increases endoplasmic reticulum stress and decreases production of human interleukin-10. Microb Cell Fact 2014;13:1–10. https://doi.org/10.1186/s12934-014-0163-7.

[33] Ben Azoun S, Kallel H. Investigating the effect of carbon source on rabies virus glycoprotein production in Pichia pastoris by a transcriptomic approach. Microbiologyopen 2017;6:1–9. https://doi.org/10.1002/mbo3.489.

[34] Kajiwara S, Yamada R, Ogino H. Secretory Overexpression of Bacillus thermocatenulatus Lipase in Saccharomyces cerevisiae Using Combinatorial Library Strategy. Biotechnol J 2018;13:e1700409. https://doi.org/10.1002/biot.201700409.

[35] Valero F. Recent advances in pichia pastoris as host for Heterologous Expression System for Lipases: A review. Methods Mol Biol 2018;1835:205–16. https://doi.org/10.1007/978-1-4939-8672-9_11.

[36] Duarte SH, del Peso Hernández GL, Canet A, Benaiges MD, Maugeri F, Valero F. Enzymatic biodiesel synthesis from yeast oil using immobilized recombinant Rhizopus oryzae lipase. Bioresour Technol 2015;183:175–80. https://doi.org/10.1016/j.biortech.2015.01.133.

[37] López-Fernández J, Barrero JJ, Benaiges MD, Valero F. Truncated Prosequence of Rhizopus oryzae Lipase : Key Factor for Production Improvement and Biocatalyst Stability. Catalysts 2019;9:961. https://doi.org/10.3390/cata9110961.

[38] Nieto-Taype MÁ, Garrigós J, Sánchez M, Valero F, Montesinos-Seguí JL, García - Ortega X. Rational-based selection of optimal operating strategies and gene dosage impact on recombinant protein production in Komagataella phafii (Pichia pastoris) 2019;13:315–27. https://doi.org/10.1111/1751-7915.13498.

[39] Garrigós-Martínez J, Nieto-Taype MA, Gasset-Franch A, Montesinos-Seguí JL, García-Ortega X, Valero F. Specific growth rate governs AOX1 gene expression,

affecting the production kinetics of Pichia pastoris (Komagataella phaffii) P AOX1 -driven recombinant producer strains with different target gene dosage. Microb Cell Fact 2019;18:187. https://doi.org/10.1186/s12934-019-1240-8.

[40] Besada-Lombana PB, Da Silva NA. Engineering the early secretory pathway for increased protein secretion in Saccharomyces cerevisiae. Metab Eng 2019;55:142–51. https://doi.org/10.1016/j.ymben.2019.06.010.

[41] Yin Y, Garcia MR, Novak AJ, Saunders AM, Ank RS, Nam AS, et al. Surf4 (Erv29p) binds amino-terminal tripeptide motifs of soluble cargo proteins with different affinities, enabling prioritization of their exit from the endoplasmic reticulum. PLoS Biol 2018;16:e2005140. https://doi.org/10.1371/journal.pbio.2005140.

[42] Belden WJ, Barlowe C. Role of Erv29p in collecting soluble secretory proteins into ER-derived transport vesicles. Science 2001;294:1528–31. https://doi.org/10.1126/science.1065224.

8. Appendix

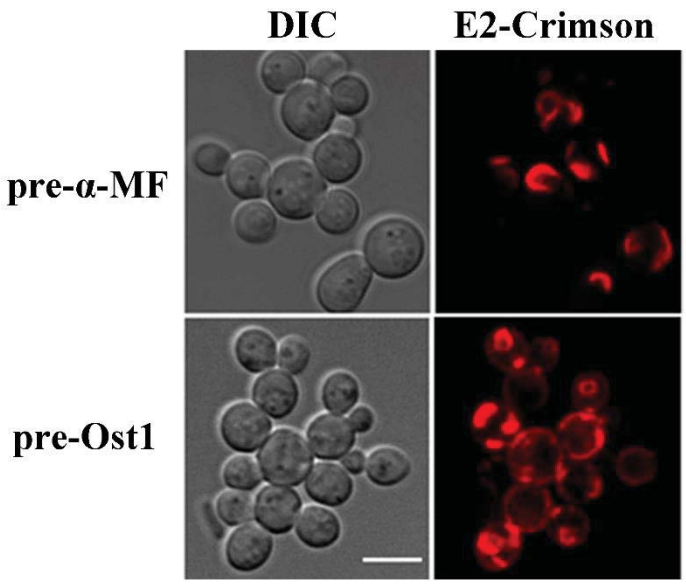

Figure S1. Images of intracellular fluorescence with the different signal sequences at the end of the E2-Crimson batch cultivation. Left: differential interference contrast (DIC) images of the cells. Right: projected confocal Z-stacks of E2-Crimson fluorescence (red). Scale bar, 5 µm.

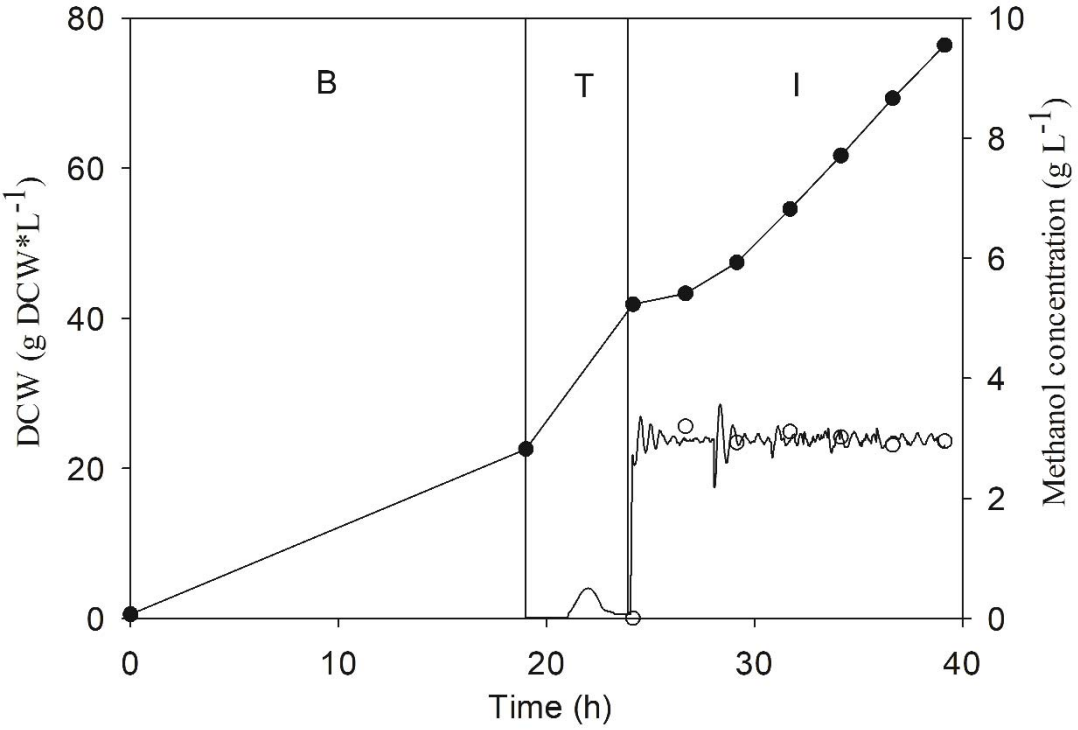

Figure S2. Fed-batch cultivation of the X33 wild-type strain. The cultivation was carried out at a constant methanol concentration (3 g L^{-1}). Time course of biomass is represented as black dots (●). Methanol concentration was measured off-line (○) and online (—). B, batch phase; T, transition phase; I, induction phase.

6

A transcriptional analysis of the yeast *Pichia pastoris* during a fed-batch cultivation reveals potential targets to increase protein secretion

Keywords: *Pichia pastoris*, protein secretion, ERAD, UPR, transcriptional analysis, real-time PCR, fed-batch.

Table of Contents

1. Summary

In this study, we have further characterized the performance of two different X33 derived strains containing the fluorescent reporter protein E2-Crimson fused to either the commonly used α-factor secretion signal (α-MF), which drives proteins through post-translational translocation, or the pre-Ost1 secretion signal sequence, which drives proteins through co-translational translocation, at a bioreactor scale. Both strains as well as the reference X33 strain were evaluated in 5-L fed-batch cultivations with the same experimental set up used in chapter 5. The cellular response to E2-crimson oversecretion along the fed-batch phase was analyzed using targeted transcriptomics with RT-PCR to study the impact of the secretion signal used for E2-Crimson on the expression levels of biomarker genes related to the secretory pathway and related cellular processes.

Our results indicate that genes related to the ER-Associated Degradation (ERAD) and the Unfolded Protein Response (UPR) are highly overexpressed over the induction phase, especially when E2-Crimson is tagged with pre-Ost1. This is one of the very few transcriptional studies carried out in *P. pastoris* cells growing in a fed-batch cultivation and the first study showing transcriptional differences when using different secretion signals rather than just an increment on protein production. Furthermore, this study can give us a better understanding on how this cellular system adapts to stress conditions and design new genetic engineering strategies to maximize production rates.

2. Introduction

As mentioned in the previous chapters, the first step in the protein secretion pathway is the translocation of the protein to be secreted into the Endoplasmic Reticulum (ER). Those proteins to be translocated are recognized by the cell machinery due to a secretion signal located at the N-terminal of the protein. As previously mentioned, the most common secretion signal to express recombinant proteins in *P. pastoris* is the alpha mating factor from *Saccharomyces cerevisiae* (α-MF) [1]. In the first part of this study, we demonstrated that the translocation step was greatly improved when the signal sequence (pre-region) from the α-MF was exchanged by the Ost1 signal sequence from *S. cerevisiae* [2,3]. The hybrid secretion signal allowed heterologous proteins to be translocated through the co-translational translocation, in which the translation and translocation occurs simultaneously. In the case of the pre region from the α-MF, the translocation and translation were happening separately [4]. These results have been shown in chapter 4 and were corroborated in a fermenter-scale in chapter 5.

Although the findings obtained in the previous chapters greatly improved the secretion efficiency, the translocation is just the first step from the secretion pathway and many other complications may occur. This led us to further study the secretion pathway to pinpoint other bottlenecks occurring downstream the pathway.

Once the recombinant protein has reached the ER lumen, heterologous proteins start to fold and obtain a tertiary structure with the aid of ER-resident chaperones, oxidoreductases and glycosylation enzymes [5,6]. The matured proteins are then recognized from the pro-region of the secretion signal by the protein Erv29 that sends protein to the ER Exit Sites (ERES), which are then sent to the Golgi and finally out of the cell [7,8]. While proteins are folding, there is a quality control checkpoint called ER-Associated Degradation (ERAD), which allows the ER to maintain a quality control and discard those misfolded proteins [9]. In the case of the ERAD present at the ER lumen (ERAD-L), the Hrd1 complex removes misfolded proteins by retrotranslocating them back to the cytosol, where they are ubiquitinated and degraded by the proteasome [10]. There are some reports suggesting that with heterologous proteins an excessive intracellular degradation may occur due to an overshooting ERAD [11,12]. This

phenomenon could be more detrimental in the context of protein production in which cells usually are forced to overproduce heterologous proteins leading to cellular stress.

Under ER- stress conditions, the Unfolded Protein Response (UPR) also plays a crucial role to decrease the stress derived from protein production. The transcription factor responsible for the UPR onset is Hac1 [13]. The mRNA from Hac1 contains an intron that prevents its correct translation, but under stress conditions there is a splicing of this intron mediated by Ire1, allowing the proper translation of Hac1 [14]. This transcription factor upregulates several genes including, foldases, chaperones and proteins from the ERAD [15]. Thus, it helps reducing the cellular stress by eliminating misfolded or aggregated proteins and the detoxification of the cell from elements such as reactive oxygen species (ROS) [16].

Here, we have performed a targeted transcriptional analysis of different marker genes involved in the secretion pathway to study the effect of the production of a protein called E2-Crimson with either the secretion signal pre-Ost1 or pre-α-MF, under the control of the methanol-inducible P_{AOX1}. Each strain was cultivated with a non-limited fed-batch culture strategy (MNLFB) utilizing a constant methanol concentration set point of 3 g/L [17]. Thereby, we aim to obtain a better understanding on how this cellular system adapts to stress conditions potentially caused by an increased translocation of E2-crimson and design new genetic engineering strategies to enhance secretion rates.

3. Materials and Methods

3.1. Strains and plasmids

P. pastoris strains used in this work were derived from the strain X33 (Thermo Fisher, Invitrogen). Plasmids were created by In-Fusion cloning (TaKaRa/Clontech), and primers required for the plasmid construction were acquired from IDT. Expression of the gene encoding E2-Crimson was driven by the methanol-inducible promoter P_{AOX1}. The model protein was fused to either the full-length α-MF secretion signal consisting of the pre-α-MF signal sequence followed by the α-MF pro region with the Ser42 mutation [2], or a chimeric secretion signal consisting of the pre-Ost1 signal sequence followed by the same α-MF pro region. Genetic engineering procedures were designed and recorded by the use of SnapGene software (Insightful Science). All the plasmids used in this work were described previously [2] and are available at Addgene.

Plasmids were linearized in the P_{AOX1} locus with the PmeI restriction enzyme and then transformed by electroporation using 100 ng of linear DNA, an amount that favored integration of a single copy of the expression cassette into the genome. Afterwards, screening was performed to detect positive colonies. Strains were grown and selected in a Yeast Extract - Peptone - Dextrose (YPD) rich medium, supplemented with the antibiotic Zeocin (100 µg/mL). Single-copy integration was confirmed by droplet digital PCR, after purification of genomic DNA, using primers that flanked the integration locus. E2-Crimson single-copy integration at the *AOX1* locus was confirmed using primers 5'-GAAATAGACGCAGATCGGGAAC-3' and 5'-GAAGGTAGACCCATGGGTTGTTG-3'.

YPD cultures for each strain were used to prepare 15% glycerol stock cell suspensions with final optical density (OD) at wavelength (λ) 600 nm equal to 60. These cell stocks were stored at -80°C.

3.2. Cultivation media

YPG pre-culturing medium contained per liter: 10 g yeast extract (Merck), 20 g peptone (Merck), and 20 g glycerol (PanReac AppliChem). This medium was supplemented with 100 µg/mL of Zeocin (Invivogen).

The basal salt synthetic medium for fed-batch cultivations contained per liter of distilled water: H_3PO_4 (85%), 26.7 mL; $CaSO_4$, 0.93 g; K_2SO_4, 18.2 g; $MgSO_4 \cdot 7H_2O$, 14.9 g; KOH, 4.13 g; glycerol, 40 g; biotin solution (200 mg/L), 4 mL; trace salts solution, 10 mL; and antifoam agent (A6426, Sigma–Aldrich Co., St. Louis, MO, USA), 0.5 mL. The trace salts solution contained per liter: $CuSO_4 \cdot 5H_2O$, 6.0 g; NaI, 0.08 g; $MnSO_4 \cdot H_2O$, 3.0 g; $Na_2MoO_4 \cdot 2H_2O$, 0.2 g; H_3BO_3, 0.02 g; $CoCl_2$, 0.5 g; $ZnCl_2$, 20.0 g; $FeSO_4 \cdot 7H_2O$, 65.0 g; biotin, 0.3 g; and concentrated H_2SO_4, 5 mL. The biotin and trace salts solutions were sterilized separately by filtration with a 0.22 µm pore filter (4433, PALL Life Sciences) and a 0.2 µm pore filter (17764-ACK0, Sartorius Stedim), respectively.

3.3. Pre-culture

Duplicate Erlenmeyer 1L flask cultures (150 mL working volume) were prepared as follows: 150 mL of YPG medium containing 100 µg/mL of Zeocin were inoculated with defrosted cryo-stocks at an OD_{600} of 0.2 and incubated overnight at 30°C and 160 rpm (Infors Multitron Shaker, 25 mm shaking diameter). The pre-culture was then centrifuged at 6000 rpm for 10 min, and the harvested cells were resuspended in sterile water in a final volume of 100 mL and directly inoculated in the reactor. Methanol batches were inoculated at an initial OD_{600} of 2, while fed-batches were inoculated at initial OD_{600} of 1.

3.4. Bioreactor cultivation set-up and operational conditions

Fed-batch cultures were carried out in a 5-L bioreactor Braun Biostat B (Braun Biotech, Melsungen, Germany) under the following cultivation conditions: initial volume 2 L, temperature 30°C, pH 5.5 controlled by adding 30% (v/v) NH_4OH during the glycerol batch phase and 5M KOH during the transition phase and methanol induction phase.

In the transition and induction phase, dissolved oxygen was set at 30% of air saturation and was controlled in cascade with agitation (from 900 to 1200 rpm during the glycerol batch phase and 1200 rpm during the transition phase and methanol induction phase), and air oxygen enrichment was added when required to maintain the culture at 30% oxygen always maintaining 2 L/min of enriched air flow rate.

First, the glycerol batch phase started with 40 g/L of glycerol as sole carbon source. Second, when the glycerol was completely consumed, as detected by a sudden increase in the dissolved oxygen concentration (DO), a 5 h transition phase began [18]. Finally, once transition phase finished the methanol induction phase started, in which methanol acted as the sole carbon source and as inducer simultaneously. The strategy was a non-limited fed-batch culture (MNLFB). A predictive-PI control strategy was supplied to maintain a constant methanol concentration of 3 g/L [17].

In the transition phase, successive solutions of 250 mL of 50% (v/v) glycerol, and 1 L of pure methanol complemented with 10 mL of trace salts solution and 4 mL of biotin solution, were added as carbon sources. Transition phase started with 300 µL/min glycerol addition for 2 h. Afterwards, the methanol addition proportion was gradually increased using 160 µL/min of glycerol and 100 µL/min of methanol for the first hour, 100 µL/min of glycerol and 100 µL/min of methanol for the second hour, and 65 µL/min of glycerol and 100 µL/min of methanol for the third hour. Glycerol solution and methanol were added using automatic microburettes S1 (Crison Instruments S.A., Alella, Barcelona, Spain). Previously, the microburette system had been sterilized by solutions of 1 M HCl, ethanol 70%, 1 M NaOH, and sterile water [19]. After transition, the nitrogen source was changed from 30% NH_4OH to NH_4Cl (100 g of NH_4Cl diluted in 500 mL of sterile water), because of interference of the NH_4OH with the methanol sensor. NH_4Cl solution was also complemented with 10 mL of trace salts solution and 4 mL of biotin solution. The NH_4Cl solution flow rate was directly linked to methanol addition by previously calculated NH_4Cl/methanol requirements (0.12 g/g) [20]. Furthermore, the pH was controlled after this point by 5 M KOH addition. All cultivations ended when the culture reached a final biomass concentration of around 70 g/L of Dry Cell Weight (DCW),). All the state variables and specific rates were calculated as previously described [21]. E2-Crimson fed-batch cultivation was performed in duplicate to obtain biological replicates for the fluorimetric, biomass and transcriptional analyses.

3.5. Methanol measurement

Methanol concentration was monitored on-line using an immersed sensor in the culture (Raven Biotech, Vancouver, BC, Canada) [22]. The sensing element that detected the methanol was the Figaro TGS-822 (Figaro USA Inc., Glenview, USA) [23]. Methanol

concentration was corroborated in the culture by HPLC off-line analyses, as reported elsewhere [22]. Relative standard deviation was below 5%.

3.6. Biomass analysis

Biomass concentration was measured as both dry cell weight (DCW) per liter of culture broth. For the DCW analysis, samples were first centrifuged at 6000 rpm for 2 min (Espresso Microcentrifugue, Thermo Electron Corporation), to collect the biomass in a pellet. Pellets were washed and centrifuged twice in a 1 g/L citric acid and 9 g/L NaCl solution, and filtered through previously weighed 0.45 µm pore glass fiber filters (APPF04700, Merck Millipore Ltd.) by applying vacuum. Finally, filters were dried at 105°C for 24 h. DCW measurements were performed in triplicate and relative standard deviation was about 5%.

3.7. E2-Crimson fluorescence assay

Both extracellular and intracellular fluorescence were analyzed. Extracellular fluorescence was measured from the supernatant obtained after centrifugation at 6000 rpm in a microcentrifuge for 2 min. Intracellular fluorescence was measured from the solid fraction after centrifugation under the same conditions. The pellet corresponding to an OD_{600} of 1 in 1 mL was first washed with phosphate-buffered saline (PBS) and then resuspended in 1 mL of PBS. Samples were stored at -20°C until the time of analysis.

100 µL each of the extracellular and intracellular fractions were transferred to a 1-cm quartz cuvette, and fluorescence was measured in duplicate in a Varian Eclipse Fluorescence Spectrophotometer (Agilent Technologies). The excitation and emission wavelengths were 570 nm and 636 nm, respectively. Fluorescence intensity was measured in the same arbitrary units for all of the samples. Extracellular and intracellular fluorescence were measured in triplicate and relative standard deviation was about 5%.

3.8. Total RNA extraction

Fed-batch samples for RNA isolation were collected according to bibliography [24]. Briefly, pellets from 1 mL culture broth samples were resuspended in 1 mL of TRIzol™ reagent (Waltham, Massachusetts, USA) and lysed with glass beads (425–600 µm, Sigma-Aldrich, St. Louis, MO, USA) for mechanical disruption. Cell lysis was attempted by alternating cycles of 30-s of vortexing and cooling. Disrupted cell lysis was then centrifuged for 5 minutes at maximum speed at 4°C. The liquid fraction was then transferred in a new tube and 200 µL of chloroform was added and mixed by inversion. The mixture was incubated for 3 minutes and centrifuged again for 15 minutes at maximum speed and at 4°C. The mixture separates into a lower red phenol-chloroform, an interphase, and a colorless upper aqueous phase. 500 µL of the colorless part was placed in a new tube and mixed with the same volume of ethanol 70%. The final mixture was placed in a column from the kit SV TOTAL RNA extraction (Promega) and further RNA extraction steps were carried out according to manufacturer's instructions. RNA integrity was checked by agarose electrophoresis and is shown in Figure A1. RNA concentration was determined using a Nanodrop instrument from Thermo Scientific™ (Waltham, MA, US).

Samples for transcriptional analysis were taken at the end of the batch, at the end of the transition phase, and every 5 hours until the end of the fed-batch, where an additional sample was taken (after 22 hours induction).

3.9. Synthesis of cDNA and determination of transcriptional levels

cDNA was synthetized with the iScript™ cDNA Synthesis kit (Bio-Rad, Hercules, CA, USA), following the manufacturer's instructions.

A set of primers were designed for specific target cDNA. The selected genes are listed in **Table 2** and the primers used to amplify part of the genes are in the Annex in **Table A1**. These genes include E2-Crimson as well as some genes involved in different parts from the secretion pathway.

For qPCR, reactions were done with SYBR™ Select Master Mix (Thermo Scientific™ Waltham, MA, US). Additionally, and as suggested by the manufacturer, to assure the maximum accuracy the reaction mixes were made by EpMotion® 5075 robot (Eppendorf, Germany). The amplification program was run on a QuantStudio 12 K Flex Real-Timer from Thermo Scientific™ (Waltham, MA, US), following the manufacturer's instructions. The annealing extension temperature was set at 59°C. Relative transcript levels were determined by using *ACT1* as a housekeeping gene and the percentage of Gene Expression shown in figures are based on the expression of *ACT1*. All the samples were performed in technical triplicates and biological duplicates.

4. Results

In a fed-batch, protein production rate is gradually reduced along the induction phase when using pre-Ost1 signal sequence

In the previous chapters, we showed that the pre-Ost1 signal sequence outperforms the conventional mating alpha factor signal sequence (pre-α-MF) in terms of final protein secretion titers and productivities. In addition, the production of proteins with the aid of pre-Ost1 also seems to decrease cellular stress derived from the expression of genes encoding heterologous proteins, supporting an increased growth rate of producer strains for most of the tested model proteins. However, we noted that in the case of E2-Crimson, there was no positive impact on growth performance when using the pre-Ost1 signal sequence in a fed-batch culture (Figure 1A). Yet, final protein secretion of E2-Crimson was increased more than 3-fold with pre-Ost1 (Figure 1B and Table 1). Therefore, we assumed that other limitations were arising downstream the secretion pathway, i.e. the translocation was not the only bottleneck along the E2-Crimson secretion process.

Our studies tracking down intracellular E2-Crimson fluorescence along the induction phase of fed-batch cultivations described in the previous chapter pointed out that it was getting retained at the ER lumen after being translocated, resulting in an intracellular accumulation of this protein (Figure 1C). However, after 10-15 hours of induction, the intracellular fluorescence started decreasing progressively until the end of the

fermentation, where the intracellular fluorescence was half of the maximum amount of fluorescence observed (Figure 1C).

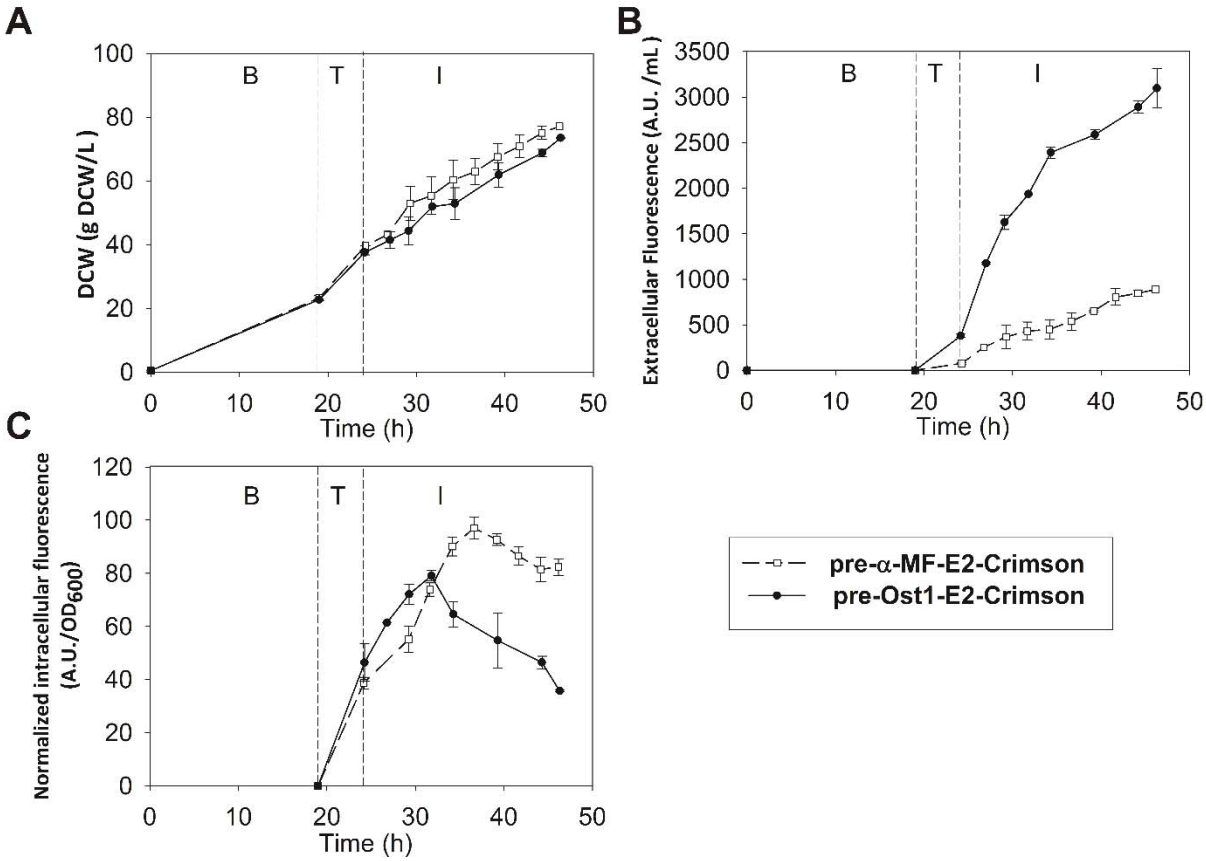

Figure 1. Results from fed-batch cultivations. Cell growth (**A**) and fluorescence (**B**) are plotted over time for the different signal sequences for E2-Crimson. Evolution of intracellular E2-Crimson fluorescence (**C**) during fed-batch cultivation with the different signal sequences. In figure 1C, the data represent the amount of fluorescence per 1 unit of OD_{600}. The letter symbols represent glycerol batch phase (B), transition phase (T), and methanol induction phase (I). A.U., arbitrary units. Error bars correspond to SD values from two biological replicates, each analyzed in triplicates.

Table 1. Fed-batch bioprocess parameters. Listed are values from wild-type (X33) and E2-Crimson fed-batch cultivations for specific growth rate (μ), specific methanol (MetOH) consumption (q_s), specific protein production (q_p), protein/biomass yield ($Y_{(p/x)}$), biomass/methanol yield ($Y_{(x/s)}$), volumetric productivity, and specific productivity. Biomass results used in calculations were measured as dried cell weight (DCW). E2-Crimson levels were measured in arbitrary units (A.U.) of fluorescence.

	X33	**E2-Crimson**	
	Wild-Type	**pre-α-MF**	**pre-Ost1**
Final E2-Crimson titer (A.U./mL)	-	879.65 ± 12	3096.2 ± 216
μ_{mean} (h^{-1})	0.081	0.050 ± 0.001	0.049 ±0.004
q_s (gMetOH)/gDCW*h)	0.30	0.20 ± 0.015	0.19 ± 0.006
Average q_p (A.U./gDCW*h)	-	740.0 ± 40	2357 ± 70
$Y_{(x/s)}$ (gDCW/gMetOH))	0.28	0.25 ± 0.005	0.25 ± 0.014
$Y_{(P/X)}$ (A.U./gDCW)	-	13620 ± 1000	44962 ± 2053
Volumetric productivity (A.U./(L*h))	-	19465 ± 303	66250 ± 3751
Specific productivity (A.U./(gDCW*h))	-	253 ± 5.5	901 ± 49.6

Hence, the missing fluorescence that accumulated presumably at the ER (as seen in Chapter 5) during the first hours of the induction phase, was either being secreted out of the cell more efficiently or degraded. In the first instance, there should be a higher E2-crimson production rate (q_p) at the late phase of induction of the pre-Ost1 strain. In order to assess this possibility, we plotted the maximum fluorescence in the reactor related to the total biomass and time for the E2-Crimson strains with either pre-Ost1 or pre-α-MF (Figure 2). The slope of the graphs corresponds to the average q_p. In the case of pre-α-MF, the q_p was rather constant during the entire induction phase of the fermentation, as the slope is kept unchanged during the entire fermentation (Figure 2A). This strain has a clear bottleneck at the translocation level, which greatly constrained

the amount of protein able to enter the ER and initiate its passage through the secretion pathway. The constant q_p value along the induction phase suggests that the translocation was the limiting factor and, as a result, protein trafficking once the protein was inside the ER was constant and without complications.

In the case of the strain with pre-Ost1, the average q_p was (2357±70 A.U./gDCW*h), which is 3-fold higher in comparison with the pre-α-MF (740±40 A.U./gDCW*h) (Table 1). However, the evolution of the q_p during the induction phase was not as constant as in the case of the pre-α-MF. Instead, there were two different periods with clearly different slope values (Figure 2B and 2B'): In the first 15 hours of the induction the slope was much higher (3761±225 A.U./gDCW*h) than the second one (1850±29 A.U./gDCW*h). Thus, average q_p was reduced twice in the second part of the fed-batch. Therefore, a plausible interpretation was that in the late phase of induction, the intracellular protein was being degraded.

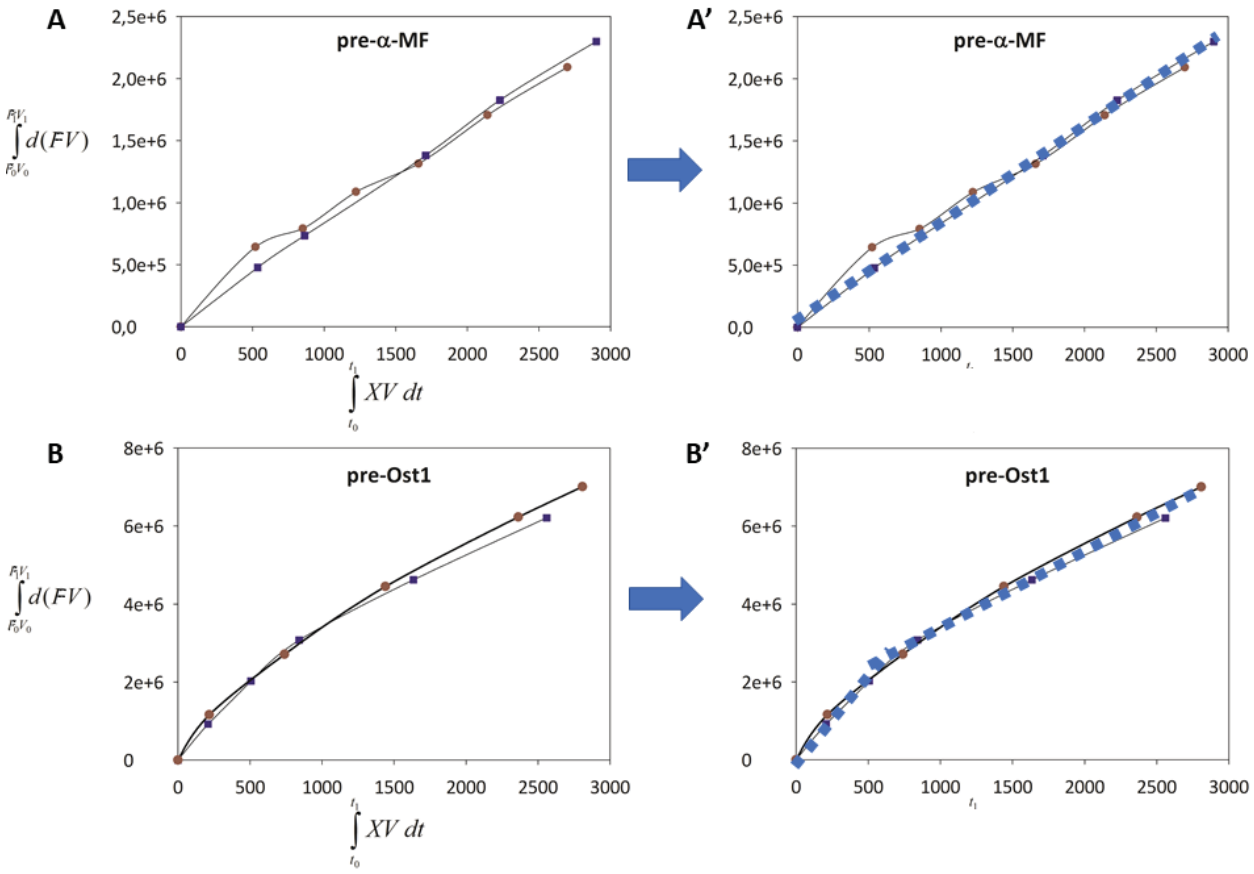

Figure 2. Variation of total fluorescence related to total biomass obtained for the fed-batch processes within the induction time. Y-axis represents the global fluorescence in the reactor and the X-axis represents the global biomass along time. **A** corresponds with the strain pre-α-MF and circles and squares represent two different biological replicates. **B** corresponds with the strain pre-Ost1 and circles and squares represent two different biological replicates. A' and B' are the same as A and B but with a trend line shown in blue. Specific production rates (q_P) can be estimated from the slope of the curves.

ERAD can be triggered by the UPR and reduce intracellular protein decreasing final production

To shed light on the underlying physiological responses occurring in the strain producing E2-Crimson with pre-Ost1 and formulate new cell engineering strategies to maximize protein production, we performed a transcriptional analysis from genes involved in the secretion pathway and its regulation. The chosen genes for this study (Table 2) were rationally selected based on two main reasons: First, the chosen genes were identified in previous transcriptomic studies to be strongly affected under protein production conditions triggering UPR, ER-stress or ERAD [25]. Second, some of these genes have

already been used as biomarkers for the UPR or ERAD in other studies [26–29]. Additionally, some of the chosen genes have been selected in this study as their overexpression or deletion conferred improved behavior in protein secretion [11,14,30]. Samples were taken at the end of the batch, at the end of the transition phase and every 5 hours after the induction phase until the end of the fermentation, where another sample was taken as well. For comparison, the reference strain (X33) was also studied using the same conditions as for the E2-Crimson strains. However, this strain reached the end of the fermentation sooner (15 hours after induction instead of 22 hours as in the E2-Crimson strains).

Table 2. Selected genes for the transcriptional analysis. The common name for the genes is shown in the first column. The genome locus corresponds to the strain GS115 (The parental strain of X33) was extracted from https://www.uniprot.org/proteomes/UP000000314 and corresponds to the genom locus where each gene is located [31]. The references shown are from relevant papers using the gene, either for transcriptional analyses purposes, or reported to improve protein secretion when the gene expression levels are modified in *P. pastoris*.

Gene	Genome locus	Function	Reference
ACT1	PAS_chr3_1169	Housekeeping gene	[32]
E2-Crimson		Model protein used for this study	-
AOX1	PAS_chr4_0821	Gene for the model's protein promoter	[27]
YAP1	PAS_chr4_0601	Transcription factor involved in the detoxification of reactive oxygen species	[33,34]
HAC1	PAS_chr1-1_0381	Transcription factor relevant for UPR activation	[35,36]
GCN4	PAS_chr1-4_0339	Transcription factor responsible for the activation of genes required for amino acid biosynthesis	[25]
KAR2	PAS_chr2-1_0140	ER resident chaperone	[27]
SSA1	PAS_chr4_0552	Cytosolic chaperone	[36]
HRD1	PAS_chr4_0156	Protein involved for the retro-translocation (ERAD-L)	[25]
VPS10	PAS_chr2-1_0625	Transmembrane protein present in the Golgi that sorts proteins to the vacuole	[3]
PEP4	PAS_chr3_1087	Vacuolar aspartyl protease (proteinase A)	[37]
ERV29	PAS_chr2-1_0287	Transmembrane protein that recognizes the pro alpha region and sends those proteins to the ERES	[30]
STE24	PAS_chr2-2_0164	Uncloggs the Sec61 translocon	[38]
ERO1	PAS_chr1-1_0011	Oxidoreductase that catalyzes the formation of disulfide bonds.	[25]

Most of the chosen genes can be organized into four main blocks: The first block corresponds to the genes *AOX1* and E2-Crimson, aiming to compare the transcription levels between both genes, as E2-Crimson expression is governed by P_{AOX1} (Figure 3). The second block contains Hac1, Yap1 and Gcn4, which are important transcription factors related with the proper balance of the cell and were chosen as they control the UPR, oxidative stress response and the amino acid biosynthesis, respectively (Figure 4). Hac1 is especially interesting as it controls the UPR and overexpression of the spliced form of *HAC1* has been reported to increase levels of secreted heterologous proteins [14] as well as *YAP1* [39]. The third block corresponds to the cytosolic chaperone Ssa1 and the ER resident chaperone Kar2 (BiP) (Figure 5). *Kar2/BiP* has been also used as a biomarker to study the UPR levels [40]. The fourth group corresponds to genes encoding proteins involved in the protein degradation. Hrd1 is involved in the ERAD, Vps10 is present in the Golgi and serves as a protein sorter to send proteins to the vacuole, and Pep4 is a protease present in the vacuole (Figure 6). Finally, three more genes were analyzed and were not possible to be included in a specific group. Erv29 recognizes the pro-α-MF at the ER and allows a rapid export from the ER to the Golgi of our model protein. In addition, overexpression of *ERV29* has been reported to improve protein secretion rates in *S. cerevisiae* [30]. *STE24* encodes a protein located in the translocon and it has been reported in *S. cerevisiae* [38] that it helps clogged translocons to get unclogged. *ERO1* encodes an ER protein essential to maintain the redox balance in the ER and can be an interesting biomarker to study the ROS levels (Figure 7). In addition, *ERO1* has been overexpressed in different studies and conferred improved secretion levels of recombinant proteins [37,41]. To normalize the results, expression levels of all genes have been normalized as relative expression levels to *ACT1*, the housekeeping gene commonly used to normalize results from transcriptional analysis [24,26,27].

Regarding expression of *AOX1* and E2-Crimson, their expression levels remained below detection level at the end of the batch, as P_{AOX1} is strongly repressed with glycerol and inducible in presence of methanol. Coherently, the expression of both genes was then induced during the entire methanol-fed fermentation stages, starting with the transition. *AOX1* expression was generally higher in the reference strain than in the strains expressing E2-Crimson (Figure 3). Both E2-Crimson-expressing strains showed

similar evolution of the E2-Crimson and *AOX1* expression profiles, except for the pre-Ost1/E2-Crimson strain, which showed significantly higher E2-Crimson transcriptional levels towards the end of the fermentation. The levels of E2-Crimson in both producing strains started being slightly smaller than *AOX1* during the first hours of induction but ended up being significantly higher than *AOX1* after 15 hours of induction (Figure 3).

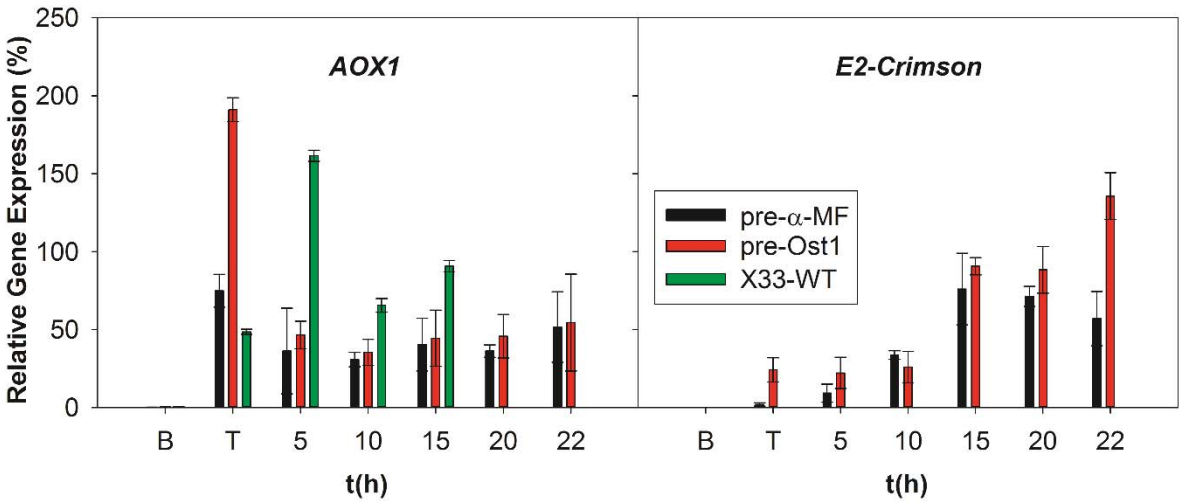

Figure 3. Analysis of P_{AOX1}-driven E2-Crimson encoding gene and endogenous *AOX1* expression levels. Relative gene expression of *AOX1* (left panel) and E2-Crimson encoding gene (Right panel) normalized with *ACT1*. **B** corresponds to the sample taken at the end of the batch phase and **T** corresponds to the sample taken at the end of the transition phase. Numbers correspond to the number of hours from the start of induction. Black bars are samples taken from strain pre-α-MF. Red bars are for the strain pre-Ost1 and green bars are for the wild-type strain. Error bars correspond to the standard deviation (±SD) between biological replicates for the E2-Crimson strains and technical triplicates for the wild-type strain.

The transcription factors are key elements to start a signaling cascade in response to any environmental perturbation during the fermentation. All strains showed increased *YAP1* relative expression levels upon shift from glycerol batch phase to methanol-fed phases. Thereafter, transcriptional levels remained essentially constant and comparable in all strains except for a transient increase at 10 hours of induction, where the relative expression levels were significantly higher for the pre-Ost1 and the reference strains. As for Hac1, which is the activator of UPR, relative expression of total *HAC1* (spliced and unspliced mRNA) was consistently increased over time for all the strains, reaching a maximum between 10 and 15 hours of induction. Upregulation of *HAC1* was especially noted in the pre-Ost1/E2-Crimson strain after 15h induction. This pointed at a higher UPR activation in the pre-Ost1 strain compared to pre-α-MF strain at this point.

Although HAC1 transcriptional levels decreased thereafter this point, they remained consistently higher in the pre-Ost1 strain until the end of the fermentation. Notably, *HAC1* levels in the reference strain were always significantly lower in comparison to the E2-Crimson producer strains (Figure 4).

Finally, the expression of *GCN4* was higher in presence of methanol compared to the glycerol batch, particularly at the end of the transition phase. However, no significant differences were observed between strains (Figure 4).

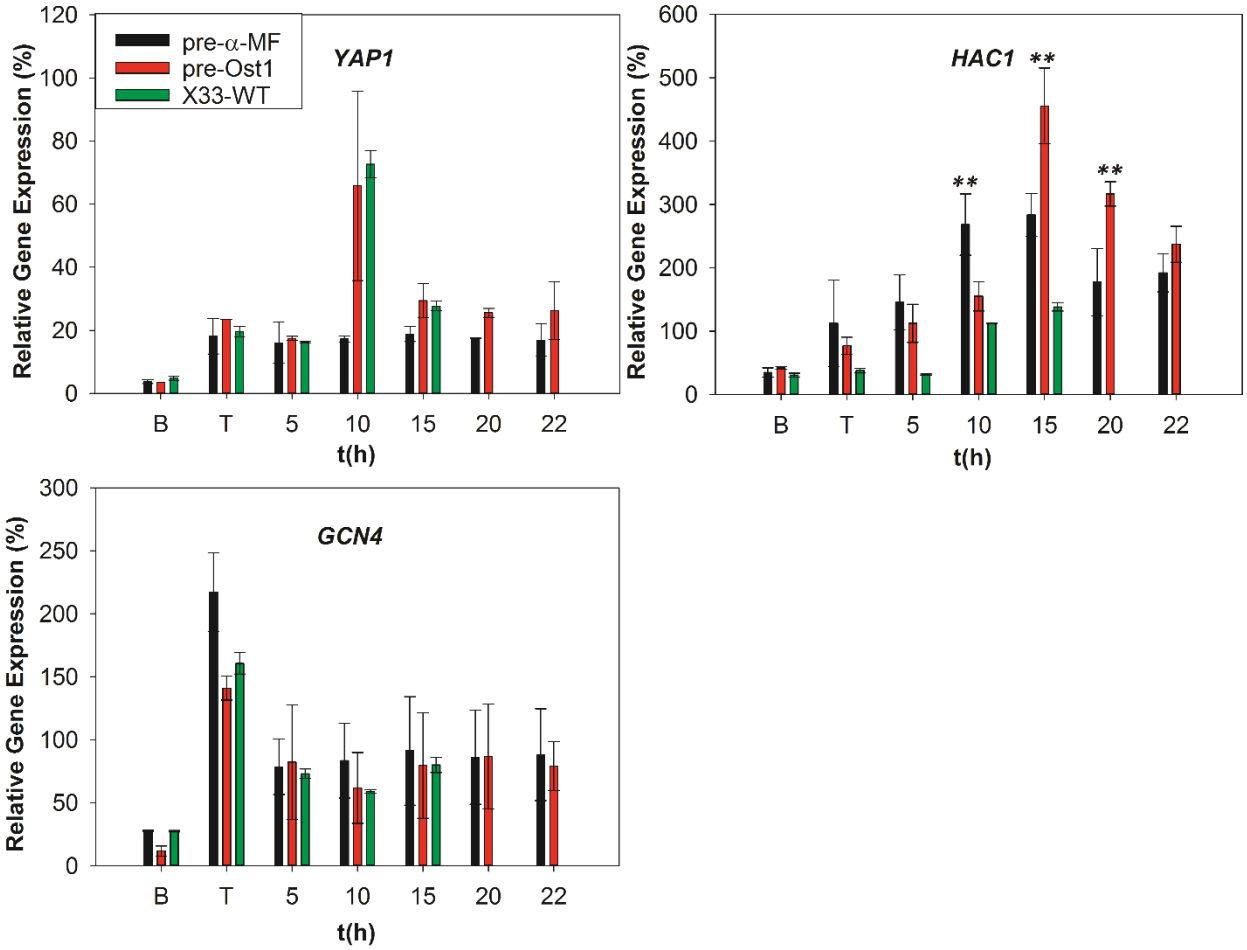

Figure 4. Relative expression levels for different transcription factors. Relative Gene expression of *YAP1* (upper-left panel), *HAC1* (upper-right panel) and *GCN4* (lower-left panel) normalized with *ACT1*. **B** corresponds to the sample taken at the end of the batch phase and **T** corresponds to the sample taken at the end of the transition phase. Numbers correspond to the number of hours from the start of induction. Black bars are samples taken from strain pre-α-MF. Red bars are for the strain pre-Ost1 and green bars are for the wild-type strain. Error bars correspond to the standard deviation (±SD) between biological replicates for the E2-Crimson strains and technical triplicates for the wild-type strain. Samples with (**) were compared with a t-student and the observed difference is statistically significant with a p-value below 0.01.

The relative expression levels of *KAR2* increased over time until it reached the maximum relative expression after 15 hours of induction and was maintained at this level until the end of the fed-batch for all the strains. Higher levels of expression were seen for those strains expressing E2-Crimson in comparison with the reference strain. However, no significant differences were observed between the E2-Crimson-expressing strains (Figure 5).

In the case of the cytosolic chaperone Ssa1, no significant differences were observed between strains and the expression levels were maintained similarly over time (Figure 5).

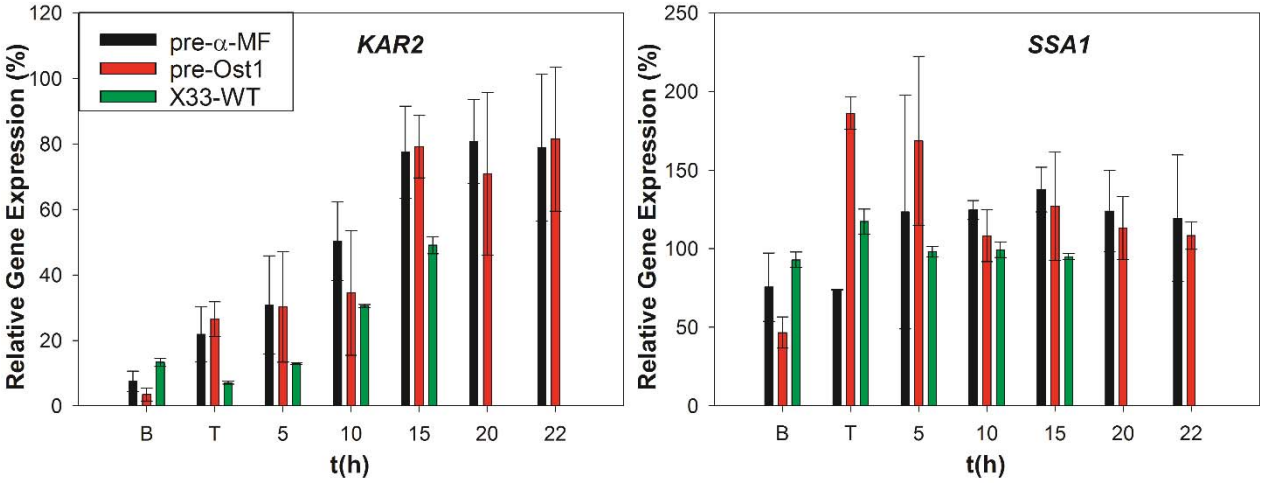

Figure 5. Relative expression levels for different chaperones. Relative Gene expression of *KAR2* (left panel) and *SSA1* (right panel) normalized with *ACT1*. **B** corresponds to the sample taken at the end of the batch phase and **T** corresponds to the sample taken at the end of the transition phase. Numbers correspond to the number of hours from the start of induction. Black bars are samples taken from strain pre-α-MF. Red bars are for the strain pre-Ost1 and green bars are for the wild-type strain. Error bars correspond to the standard deviation (±SD) between biological replicates for the E2-Crimson strains and technical triplicates for the wild-type strain.

Relative expression levels from genes involved in protein degradation is shown in Figure 6. The ERAD in the ER lumen is associated with the retrotranslocation of proteins by the Hrd1 complex. Expression of the gene encoding Hrd1 increased for all the strains after 10 hours of induction and was steadily maintained until the end of the fermentation. These increased expression levels are more relevant for the pre-Ost1 strain, which

doubled the levels of expression in comparison with the pre-α-MF strain in the final phase of induction (Figure 6). Similar behavior was detected for the *VPS10* gene, except that at 15 hours induction the reference strain seemed to have higher levels of expression in comparison with the E2-Crimson producer strains. As for the gene *PEP4*, the levels of expression were higher in presence of methanol than at the end of the batch. However, no significant differences were seen between strains (Figure 6).

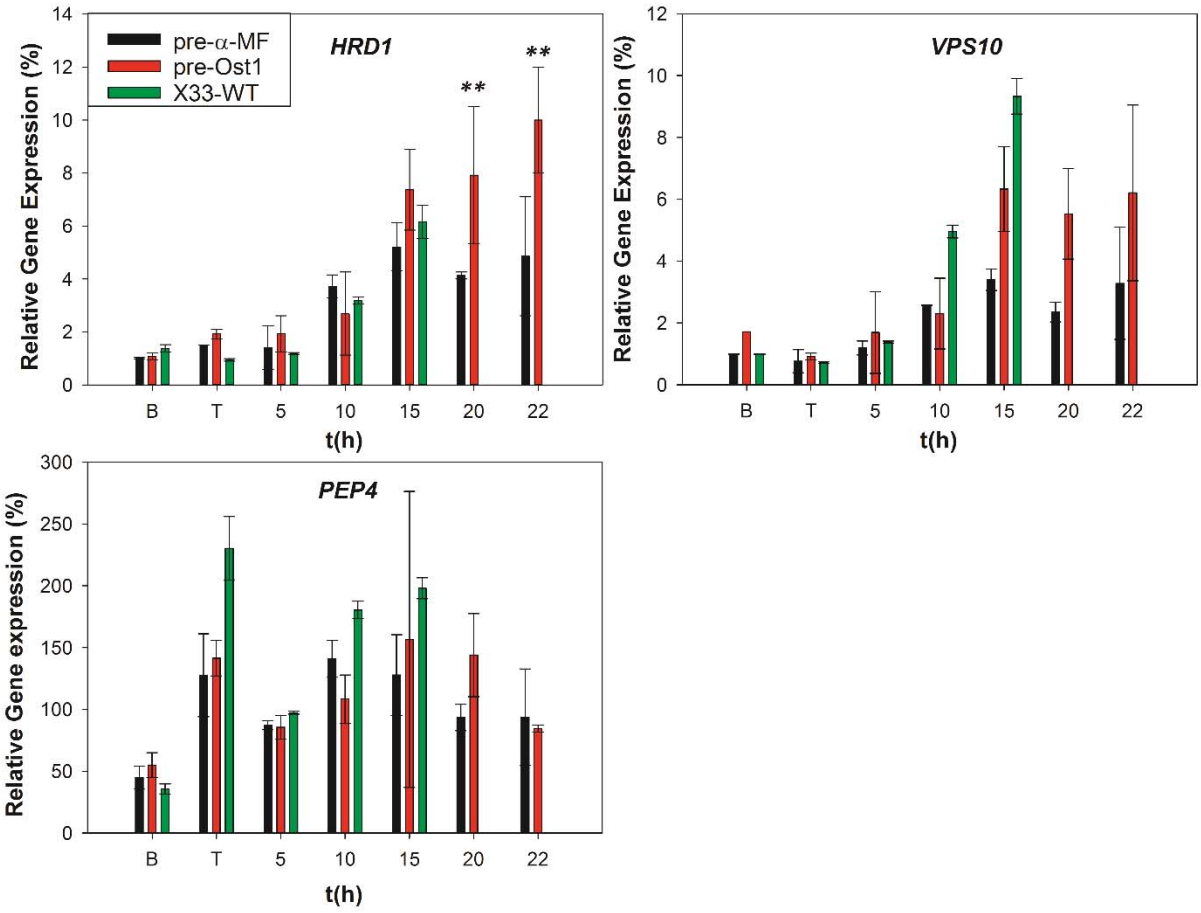

Figure 6. Relative expression for genes involved in protein degradation. Relative Gene expression of *HRD1* (upper-left panel), *VPS10* (upper-right panel) and *PEP4* (lower-left panel) normalized with *ACT1*. **B** corresponds to the sample taken at the end of the batch phase and **T** corresponds to the sample taken at the end of the transition phase. Numbers correspond to the number of hours from the start of induction. Black bars are samples taken from strain pre-α-MF. Red bars are for the strain pre-Ost1 and green bars are for the wild-type strain. Error bars correspond to the standard deviation (±SD) between biological replicates for the E2-Crimson strains and technical triplicates for the wild-type strain. Samples with (**) were compared with a t-student and the observed difference is statistically significant with a p-value below 0.01.

Expression of the gene *ERV29* reached the highest levels after 15 hours of induction and was maintained high until the end of the fermentation. No significant differences in terms of expression were seen between strains. As for the gene *STE24*, the expression levels were maintained similarly during the entire fermentation and between strains with no significant differences. In the case of Ero1, increased levels of *ERO1* were observed after 10-15 hours from the start of the induction phase. The expression of *ERO1* seemed to be slightly higher for the pre-Ost1 and the reference strain, although these results are not statistically significant (Figure 7).

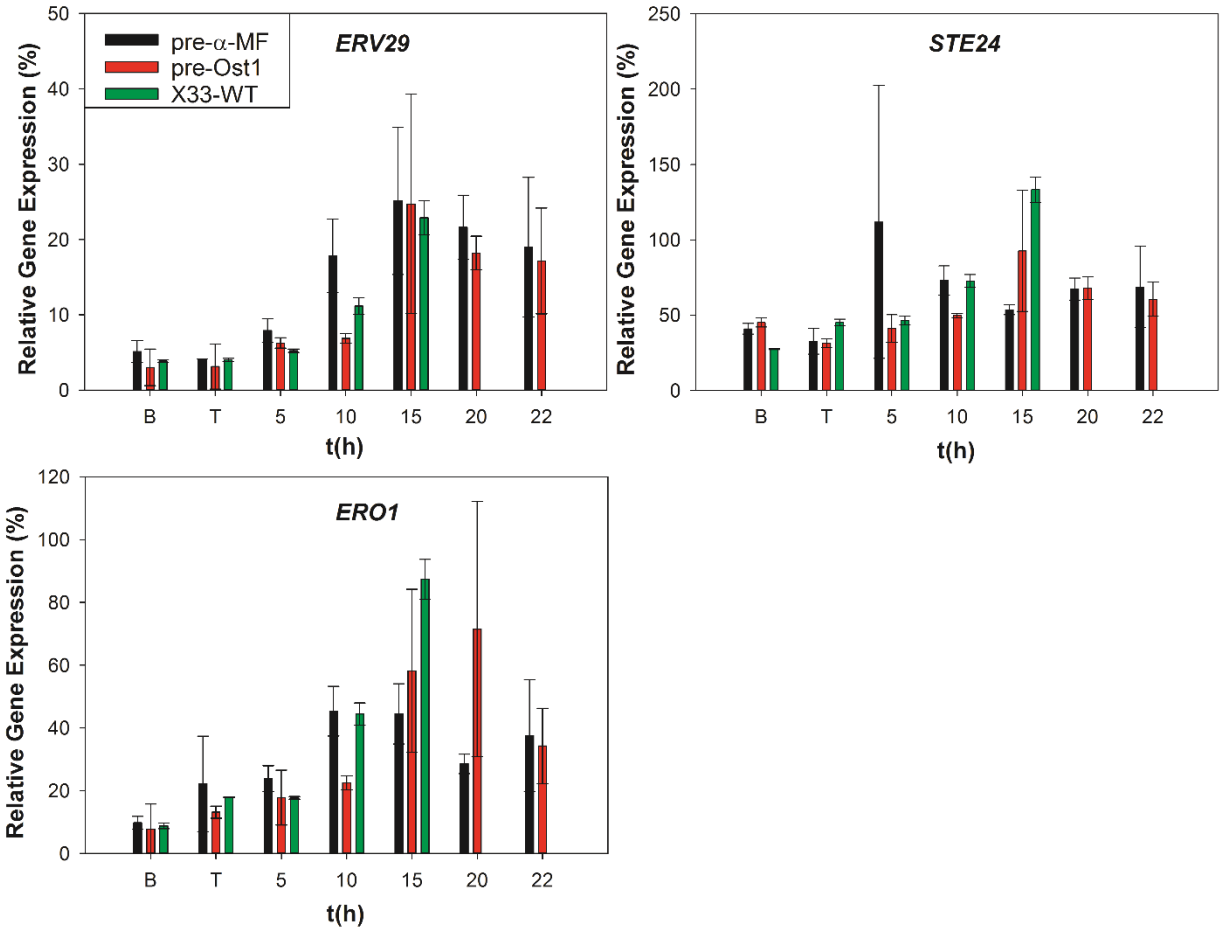

Figure 7. Relative expression for different genes. Relative Gene expression of *ERV29* (upper-left panel), *STE24* (upper-right panel) and *ERO1* (lower-left panel) normalized with *ACT1*. **B** corresponds to the sample taken at the end of the batch phase and **T** corresponds to the sample taken at the end of the transition phase. Numbers correspond to the number of hours from the start of induction. Black bars are samples taken from strain pre-α-MF. Red bars are for the strain pre-Ost1 and green bars are for the wild-type strain. Error bars correspond to the standard deviation (±SD) between biological replicates for the E2-Crimson strains and technical triplicates for the wild-type strain.

5. Discussion

In this study, we have further analyzed the physiological effects derived from the use of the signal sequence pre-Ost1 in comparison with the most commonly used signal sequence pre-α-MF for heterologous protein secretion. E2-Crimson was the chosen reporter gene, as it had been already used in previous studies. Furthermore, *P. pastoris* seemed to have some limitations to sustain E2-crimson secretion levels at high rates, especially over the later phases of fed-batch fermentations, i.e. after long induction periods (Chapter 5). The fed-batch cultivation strategy was based on a closed control loop in which 3 g/L of methanol was maintained until the biomass reached a concentration of around 70 g/L of DCW. Different samples were taken at different time-points and a transcriptional analysis was carried out for the genes listed in Table 2. For comparison, the X33 reference strain was also cultivated with the same strategy and samples were taken as well.

Our transcriptional analysis started with a comparison of our model gene E2-Crimson and the *AOX1*. As expected, the E2-Crimson and AOX1 transcriptional levels in the producer strains were positively correlated to each other at each sampling point, except for the last sample (22 hours of induction), where E2-crimson expression levels in the pre-Ost1 strain were significantly higher than in the pre-α-MF strain, but *AOX1* levels were similar in both strains. Interestingly, both *AOX1* and E2-Crimson relative expression levels in the transition phase were significantly higher in the pre-Ost1 strain compared to the pre-α-MF strain. (Figure 3). In the transition phase, cells are adapting from a glycerol-only metabolism to a dynamic mixed glycerol-methanol feeding and, therefore, a high variability in the cell's physiological state during this phase may be expected over a short period. However, the difference observed between *AOX1* and E2-crimson expression levels at the end of the pre-Ost1 strain fermentation was unexpected.

In principle, since P_{AOX1} was used to express E2-Crimson, similar relative expression levels should be observed for both, E2-Crimson and *AOX1*, at each of the sampling points. However, this behavior was not observed in any of the E2-crimson producing strains. In particular, E2-Crimson expression levels were generally lower than *AOX1* during the first 10 hours of induction, but thereafter this tendency was inverted (Figure

3). These results were unexpected, as expression levels of both genes should be similar to each other at each sampling point over the entire induction phase. This could reflect differences in the half-lives of the corresponding mRNAs.

It is also worth comparing *AOX1* expression levels between E2-Crimson strains and the reference strain. In the latter case, *AOX1* expression levels tended to be higher than those observed for the production strains. Coherently, the q_s of the reference strain was a 30% higher than those of the E2-Crimson strains (Table 1). A similar behavior has been recently observed in a series of strains expressing different dosages of a *Rhizopus oryzae* lipase gene under the control of the *AOX1* promoter, where a transcription factor titration effect caused by additional copies of *AOX1* promoters was observed, leading to reduced transcriptional levels of the endogenous *AOX1* gene in the recombinant strains compared to the reference strain [42].

The transcription factors are key elements to detect perturbations or changes in levels of transcription, as they usually are the first actuators to produce a cascade of expression from different genes. In this case, *HAC1* has been the gene with more distinguished differences during the fed-batch cultivations (Figure 4). *HAC1* is the first activator of the UPR signal response and is constitutively expressed under non stress conditions [14,43]. The mRNA of *HAC1* contains an intron that prevents Hac1 to be translated an active transcription factor. However, under stress conditions, a protein called Ire1 removes the intron and Hac1 is translated active [14]. This active Hac1 interacts as a transcription factor and enhances the overexpression of several genes related with the secretion pathway including chaperones, proteins involved in ERAD and Hac1 itself [44]. The *HAC1* transcript levels represented in Figure 4, correspond to the total amount of transcripts, i.e. with and without the intron. Increased levels of *HAC1* expression after 15 hours of induction for the strain pre-Ost1, denotates an increased level of active Hac1 prior to the 15 hours of induction, and the activation of the UPR response in the pre-Ost1 strain. This effect was also seen in the E2-Crimson producer strain pre-α-MF at 10 hours induction, but such upregulation seemed to be more moderate (Figure 4).

Coherently, both E2-Crimson-producing strains showed a clear upregulation of *KAR2*, especially noted after 15 hours of induction, and thereafter maintained upregulated

until the end of the fermentation. Interestingly, the *KAR2* expression levels were similar in both, pre-Ost1 and pre-α-MF strains during this period, despite the fact that HAC1 levels were significantly different between strains. Kar2 is commonly used by researchers as an UPR marker to study different effects from cellular stress [40]. A "plateau" effect in Kar2 levels could reflect the existence of an upper limit in Kar2 levels, regardless active Hac1 may transiently reach higher levels. However, further experiments should be performed to test whether this hypothesis is correct or not. Moreover, *KAR2* seemed to be constantly upregulated at the late phase of induction despite *HAC1*'s relative expression levels started to decrease after 15 hours of induction. A possible reduction of *KAR2*'s relative expression might have happened a few hours later from the end of the fermentation, as reported in other studies with longer induction times [29,40].

On the other hand, the Ssa1 cytosolic chaperone did not seem to be affected by the UPR response (Figure 5). Kar2 and Ssa1 are heat-shock proteins which expression can be affected by different temperature changes [45]. Although this effect can be corelated with the UPR, not all the genes are affected for both conditions [46], and the gene *SSA1* did not seem to be affected by the UPR in this study.

Another gene clearly affected by the UPR response was *HRD1*, a gene encoding a protein involved in the ERAD (Figure 6). This gene was almost 2-fold upregulated after 15 hours of induction in pre-Ost1, the strain with more *HAC1* expression (Figure 4). This upregulation of *HRD1* was also observed in the pre-α-MF strain, but more moderated. This is coherent with the observation that intracellular fluorescence of E2-Crimson in the pre-Ost1 strain started being accumulated along the early induction phase but was reduced at the late phase of induction (Figure 1C). Similar results were seen with the vacuolar sorter Vps10 (Figure 6).

An increased level of *HRD1* and *VPS10* was also seen in the reference strain. Notably, although this strain was grown at the same conditions as the E2-Crimson-producing strains (i.e. same methanol concentration set point during the induction phase), the growth rate was $0.081h^{-1}$, which is much faster than the observed rate of $0.05h^{-1}$ for the E2-Crimson strains. Here, our initial idea was to use this strain as a reference to assess the effect of protein production. However, this effect could be masked by a higher

growth rate of the reference strain. Therefore, although the results from the reference strain can be informative, they are not fully comparable to those of the producing strains.

Another gene related to the UPR that was significantly upregulated was the gene encoding for the transmembrane protein Erv29 (Figure 7) [46]. However, such upregulation was similar in both E2-Crimson-producing strains, as also observed for *KAR2*.

The mRNA is a rather labile molecule and a transcriptional analysis in such a dynamic system as a fed-batch is complex, as measured mRNA levels depend both on the synthesis and degradation rates, and the lifetime of a molecule of mRNA is usually of a few minutes [47]. In addition, in a fed-batch system cells are not in a steady-state and a mixed population of cells is usually present. However, in this study a clear trend on different gene expression levels have been observed and, from these results, different strategies to maximize protein production can be envisioned.

In principle, overexpression of Hac1 could be a good candidate to activate the UPR from the very beginning of the induction phase, thereby supporting increased protein production, as it has been previously described [37]. Increased expression levels of active *HAC1* can increase the levels of helper proteins such as Kar2 and Erv29, which could also increase a fast export of cargo protein to the Golgi [30]. However, enhanced expression of *HAC1* also leads to higher expression of genes involved in ERAD. Therefore, *HAC1* overexpression by itself could be even detrimental for an improved behavior of our pre-Ost1 strain. A good strategy would be then to overexpress *HAC1* in combination with knocking-out a gene related to ERAD-L such as *DER1* to avoid retrotranslocation of our protein of interest [11,30]. Alternatively, instead of disrupting the ERAD machinery, a non-conventional strategy we envision is to remove from the promoter of *HRD1* the CIS binding site for Hac1, thereby creating a strain unsensitive to higher levels of Hac1. This new strategy could create a strain not able to increase the levels of *HRD1* when the UPR is activated, preventing the ERAD machinery to overshoot active protein. This effect could be maximized if other Hac1 binding sites from other genes related to the ERAD machinery besides *HRD1* were removed.

Additionally, an overexpression of *ERV29* should be combined with the genetic modifications mentioned before to increase the export of proteins from the ER to the Golgi and avoid protein accumulation [30]. This overexpression could be performed with the endogenous *ERV29* from *P. pastoris* or even with the *ERV29* from *S. cerevisiae*. As the pro-α-MF commonly used for protein production is from *S. cerevisiae*, the *ERV29* from *S. cerevisiae* might work better and improve protein export.

In this study, we used strains harboring just a single gene copy of E2-Crimson. Therefore, the observed protein accumulation (Figure 1C) and subsequent effects could be multiplied in a system with increased gene dosages. Consequently, an increased gene dosage combined with the genetic engineering strategy proposed above could boost final protein production. Researchers have been previously attempted to increase gene dosage to increase protein yields [42]. However, some reports suggest certain limitations occur at some point of the secretory pathway [28,48]. We envision that these limitations could be minimized by using this combined genetic engineering approach.

Notably, we have shown that a mere change in the secretion signal may lead to a measurable physiological impact. A genome-wide transcriptomic study with microarrays or RNAseq could be interesting to find novel biomarkers allowing to fine-tune the analysis performed in fed-batch cultivations and obtain a better and more comprehensive diagnosis of the physiological response and adaptation to increased translocation efficiencies under dynamic conditions, as it has been previously reported for different environmental changes [49].

6. Conclusions

As we expected, the use of a signal sequence that alleviated the bottleneck present at the translocation step, yielded in other bottlenecks downstream the secretion pathway that could not allow to fully exploit the increased efficiency of the translocation step and, consequently, maximize the final protein production and the cellular performance. However, this transcriptional analysis has given us a diagnosis on potential physiological bottlenecks in our system under dynamic process-like conditions, therefore providing a rational basis for further improvement of the strain performance and robustness for protein production.

7. References

[1] Brake AJ, Merryweather JP, Coit DG, Heberlein UA, Masiarz FR, Mullenbach GT, et al. Alpha-factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 1984;81:4642–6. https://doi.org/10.1073/pnas.81.15.4642.

[2] Barrero JJ, Casler JC, Valero F, Ferrer P, Glick BS. An improved secretion signal enhances the secretion of model proteins from Pichia pastoris. Microb Cell Fact 2018;17:161. https://doi.org/10.1186/s12934-018-1009-5.

[3] Fitzgerald I, Glick BS. Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting. Microb Cell Fact 2014;13:125. https://doi.org/10.1186/s12934-014-0125-0.

[4] Ng DTW, Brown JD, Walter P. Signal sequences specify the targeting route to the endoplasmic reticulum membrane. J Cell Biol 1996;134:269–78. https://doi.org/10.1083/jcb.134.2.269.

[5] Braakman I, Bulleid NJ. Protein Folding and Modification in the Mammalian Endoplasmic Reticulum. Annu Rev Biochem 2011;80:71–99. https://doi.org/10.1146/annurev-biochem-062209-093836.

[6] Tannous A, Brambilla Pisoni G, Hebert DN, Molinari M. N-linked sugar-regulated protein folding and quality control in the ER. Semin Cell Dev Biol 2015;41:79–89. https://doi.org/10.1016/j.semcdb.2014.12.001.

[7] Otte S, Barlowe C. Sorting signals can direct receptor-mediated export of soluble proteins into COPII vesicles. Nat Cell Biol 2004;6:1189–94. https://doi.org/10.1038/ncb1195.

[8] Yin Y, Garcia MR, Novak AJ, Saunders AM, Ank RS, Nam AS, et al. Surf4 (Erv29p) binds amino-terminal tripeptide motifs of soluble cargo proteins with different affinities, enabling prioritization of their exit from the endoplasmic reticulum. PLoS Biol 2018;16:e2005140. https://doi.org/10.1371/journal.pbio.2005140.

[9] Stevenson J, Huang EY, Olzmann JA. Endoplasmic Reticulum–Associated Degradation and Lipid Homeostasis. Annu Rev Nutr 2016;36:511–42. https://doi.org/10.1146/annurev-nutr-071715-051030.

[10] Baldridge RD, Rapoport TA. Autoubiquitination of the Hrd1 Ligase Triggers Protein Retrotranslocation in ERAD. Cell 2016;166:394–407. https://doi.org/10.1016/j.cell.2016.05.048.

[11] Zahrl RJ, Mattanovich D, Gasser B. The impact of ERAD on recombinant protein secretion in Pichia pastoris (syn Komagataella spp.). Microbiology 2018;164:453–63. https://doi.org/10.1099/mic.0.000630.

[12] Pfeffer M, Maurer M, Stadlmann J, Grass J, Delic M, Altmann F. Intracellular interactome of secreted antibody Fab fragment in Pichia pastoris reveals its routes of secretion and degradation. Appl Microbiol Biotechnol 2012;93:2503–12. https://doi.org/10.1007/s00253-012-3933-3.

[13] Young CL, Robinson AS. Protein Folding and Secretion: Mechanistic Insights Advancing Recombinant Protein Production in S. cerevisiae. Curr Opin Biotechnol 2014;30:168–77. https://doi.org/10.1038/jid.2014.371.

[14] Guerfal M, Ryckaert S, Jacobs PP, Ameloot P, Van Craenenbroeck K, Derycke R, et al. The HAC1 gene from Pichia pastoris: Characterization and effect of its overexpression on the production of secreted, surface displayed and membrane

proteins. Microb Cell Fact 2010;9:49. https://doi.org/10.1186/1475-2859-9-49.

[15] Travers KJ, Patil CK, Wodicka L, Lockhart DJ, Weissman JS, Walter P. Functional and Genomic Analyses Reveal an Essential Coordination between the Unfolded Protein Response and ER-Associated Degradation. Cell 2000;101:249–58. https://doi.org/10.1016/s0092-8674(00)80835-1.

[16] Delic M, Rebnegger C, Wanka F, Puxbaum V, Haberhauer-Troyer C, Hann S, et al. Oxidative protein folding and unfolded protein response elicit differing redox regulation in endoplasmic reticulum and cytosol of yeast. Free Radic Biol Med 2012;52:2000–12. https://doi.org/10.1016/j.freeradbiomed.2012.02.048.

[17] Cos O, Ramon R, Montesinos JL, Valero F. A simple model-based control for Pichia pastoris allows a more efficient heterologous protein production bioprocess. Biotechnol Bioeng 2006;95:145–54. https://doi.org/10.1002/bit.21005.

[18] Cos O, Serrano A, Montesinos JL, Ferrer P, Cregg JM, Valero F. Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in Pichia pastoris fed-batch cultures. J Biotechnol 2005;117:321–35. https://doi.org/10.1016/j.jbiotec.2004.12.010.

[19] Arnau C, Ramon R, Casas C, Valero F. Optimization of the heterologous production of a Rhizopus oryzae lipase in Pichia pastoris system using mixed substrates on controlled fed-batch bioprocess. Enzyme Microb Technol 2010;46:494–500. https://doi.org/10.1016/j.enzmictec.2010.01.005.

[20] 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. Microb Cell Fact 2006;5:1–20. https://doi.org/10.1186/1475-2859-5-17.

[21] Barrigón JM, Montesinos JL, Valero F. Searching the best operational strategies for Rhizopus oryzae lipase production in Pichia pastoris Mut+ phenotype: Methanol limited or methanol non-limited fed-batch cultures? Biochem Eng J 2013. https://doi.org/10.1016/j.bej.2013.03.018.

[22] Arnau C, Casas C, Valero F. The effect of glycerol mixed substrate on the

heterologous production of a Rhizopus oryzae lipase in Pichia pastoris system. Biochem Eng J 2011;57:30–7. https://doi.org/10.1016/j.bej.2011.08.004.

[23] Valero F, Ramon R, Feliu JX, Berthet FX, Montesinos JL, Cos O. Improving the monitoring of methanol concentration during high cell density fermentation of Pichia pastoris. Biotechnol Lett 2004;26:1447–52. https://doi.org/10.1023/b:bile.0000045649.60508.c5.

[24] Landes N, Gasser B, Vorauer-Uhl K, Lhota G, Mattanovich D, Maurer M. The vitamin-sensitive promoter PTHI11 enables pre-defined autonomous induction of recombinant protein production in Pichia pastoris. Biotechnol Bioeng 2016;113:2633–43. https://doi.org/10.1002/bit.26041.

[25] Gasser B, Maurer M, Rautio J, Sauer M, Bhattacharyya A, Saloheimo M, et al. Monitoring of transcriptional regulation in Pichia pastoris under protein production conditions. BMC Genomics 2007;8. https://doi.org/10.1186/1471-2164-8-179.

[26] Wang XD, Jiang T, Yu XW, Xu Y. Effects of UPR and ERAD pathway on the prolyl endopeptidase production in Pichia pastoris by controlling of nitrogen source. J Ind Microbiol Biotechnol 2017;44:1053–63. https://doi.org/10.1007/s10295-017-1938-8.

[27] Cámara E, Monforte S, Albiol J, Ferrer P. Deregulation of methanol metabolism reverts transcriptional limitations of recombinant Pichia pastoris (Komagataella spp) with multiple expression cassettes under control of the AOX1 promoter. Biotechnol Bioeng 2019;116:1710–20. https://doi.org/10.1002/bit.26947.

[28] Garrigós-Martínez J, Nieto-Taype MA, Gasset-Franch A, Montesinos-Seguí JL, García-Ortega X, Valero F. Specific growth rate governs AOX1 gene expression, affecting the production kinetics of Pichia pastoris (Komagataella phaffii) P AOX1 -driven recombinant producer strains with different target gene dosage. Microb Cell Fact 2019;18:187. https://doi.org/10.1186/s12934-019-1240-8.

[29] Resina D, Bollók M, Khatri NK, Valero F, Neubauer P, Ferrer P. Transcriptional response of P. pastoris in fed-batch cultivations to Rhizopus oryzae lipase

production reveals UPR induction. Microb Cell Fact 2007;6:21. https://doi.org/10.1186/1475-2859-6-21.

[30] Besada-Lombana PB, Da Silva NA. Engineering the early secretory pathway for increased protein secretion in Saccharomyces cerevisiae. Metab Eng 2019;55:142–51. https://doi.org/10.1016/j.ymben.2019.06.010.

[31] De Schutter K, Lin YC, Tiels P, Van Hecke A, Glinka S, Weber-Lehmann J, et al. Genome sequence of the recombinant protein production host Pichia pastoris. Nat Biotechnol 2009;27:561–6. https://doi.org/10.1038/nbt.1544.

[32] Landes N, Gasser B, Vorauer-Uhl K, Lhota G, Mattanovich D, Maurer M. The vitamin-sensitive promoter PTHI11 enables pre-defined autonomous induction of recombinant protein production in Pichia pastoris. Biotechnol Bioeng 2016;113:2633–43. https://doi.org/10.1002/bit.26041.

[33] Delic M, Gö R, Mattanovich D, Gasser B. Engineering of Protein Folding and Secretion-Strategies to Overcome Bottlenecks for Efficient Production of Recombinant Proteins. Antioxid Redox Signal 2014;21:414–37. https://doi.org/10.1089/ars.2014.5844.

[34] Delic M, Rebnegger C, Wanka F, Puxbaum V, Haberhauer-Troyer C, Hann S, et al. Oxidative protein folding and unfolded protein response elicit differing redox regulation in endoplasmic reticulum and cytosol of yeast. Free Radic Biol Med 2012;52:2000–12. https://doi.org/10.1016/j.freeradbiomed.2012.02.048.

[35] Hesketh AR, Castrillo JI, Sawyer T, Archer DB, Oliver SG. Investigating the physiological response of Pichia (Komagataella) pastoris GS115 to the heterologous expression of misfolded proteins using chemostat cultures. Appl Microbiol Biotechnol 2013;97:9747–62. https://doi.org/10.1007/s00253-013-5186-1.

[36] Vanz AL, Nimtz M, Rinas U. Decrease of UPR- and ERAD-related proteins in Pichia pastoris during methanol-induced secretory insulin precursor production in controlled fed-batch cultures. Microb Cell Fact 2014;13. https://doi.org/10.1186/1475-2859-13-23.

[37] Puxbaum V, Mattanovich D, Gasser B. Quo vadis? The challenges of recombinant protein folding and secretion in Pichia pastoris. Appl Microbiol Biotechnol 2015;99:2925–38. https://doi.org/10.1007/s00253-015-6470-z.

[38] Ast T, Michaelis S, Schuldiner M, States U. The protease Ste24 clears clogged translocons. Cell 2017;164:103–14. https://doi.org/10.1016/j.cell.2015.11.053.The.

[39] Delic M, Graf AB, Koellensperger G, Haberhauer-Troyer C, Hann S, Mattanovich D, et al. Overexpression of the transcription factor Yap1 modifies intracellular redox conditions and enhances recombinant protein secretion. Microb Cell 2014;1:376–86. https://doi.org/10.15698/mic2014.11.173.

[40] Roth G, Vanz AL, Lünsdorf H, Nimtz M, Rinas U. Fate of the UPR marker protein Kar2/Bip and autophagic processes in fed-batch cultures of secretory insulin precursor producing Pichia pastoris. Microb Cell Fact 2018;17:123. https://doi.org/10.1186/s12934-018-0970-3.

[41] Baumann K, Adelantado N, Lang C, Mattanovich D, Ferrer P. Protein trafficking, ergosterol biosynthesis and membrane physics impact recombinant protein secretion in Pichia pastoris. Microb Cell Fact 2011;10:93. https://doi.org/10.1186/1475-2859-10-93.

[42] Cámara E, Landes N, Albiol J, Gasser B, Mattanovich D, Ferrer P. Increased dosage of AOX1 promoter-regulated expression cassettes leads to transcription attenuation of the methanol metabolism in Pichia pastoris. Sci Rep 2017;7. https://doi.org/10.1038/srep44302.

[43] Rüegsegger U, Leber JH, Walter P. Block of HAC1 mRNA translation by long-range base pairing is released by cytoplasmic splicing upon induction of the unfolded protein response. Cell 2001;107:103–14. https://doi.org/10.1016/S0092-8674(01)00505-0.

[44] Ogawa N, Mori K. Autoregulation of the HAC1 gene is required for sustained activation of the yeast unfolded protein response. Genes to Cells 2004;9:95–104. https://doi.org/10.1111/j.1365-2443.2004.00704.x.

[45] Yu X-W, Sun W-H, Wang Y-Z, Xu Y. Identification of novel factors enhancing recombinant protein production in multi-copy Komagataella phaffii based on transcriptomic analysis of overexpression effects. Sci Rep 2017;7:16249. https://doi.org/10.1038/s41598-017-16577-x.

[46] Liu Y, Chang A. Heat shock response relieves ER stress. EMBO J 2008;27:1049–59. https://doi.org/10.1038/emboj.2008.42.

[47] Kim CH, Warner JR. Messenger RNA for ribosomal proteins in yeast. J Mol Biol 1983;165:79–89. https://doi.org/10.1016/S0022-2836(83)80243-5.

[48] Nieto-Taype MÁ, Garrigós J, Sánchez M, Valero F, Montesinos-Seguí JL, García - Ortega X. Rational-based selection of optimal operating strategies and gene dosage impact on recombinant protein production in Komagataella phafii (Pichia pastoris) 2019;13:315–27. https://doi.org/10.1111/1751-7915.13498.

[49] Burgard J, Valli M, Graf AB, Gasser B, Mattanovich D. Biomarkers allow detection of nutrient limitations and respective supplementation for elimination in Pichia pastoris fed-batch cultures. Microb Cell Fact 2017;16:117. https://doi.org/10.1186/s12934-017-0730-9.

8. Appendix

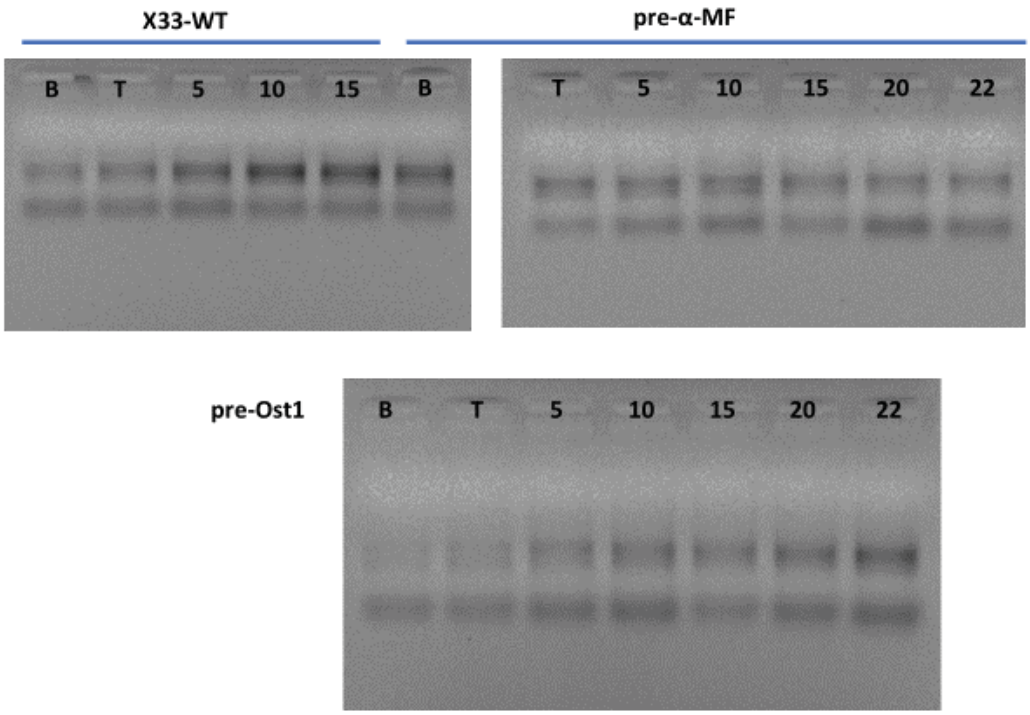

Figure S1. RNA integrity from the samples extracted in this study. An agarose gel with 100ng of RNA from each of the analyzed samples. The two bands correspond to the Ribosomal RNA 18S and 28S for all the samples analyzed in this study. **B** corresponds to the sample taken at the end of the batch phase and **T** corresponds to the sample taken at the end of the transition phase. Numbers correspond to the number of hours from the start of induction.

Gene	Primers	Sequence	Length (bp)	Annealing TºC	GC content	Amplicon Length (bp)	Efficiency (%)
Act1	Act1_FW	CCTGAGGCTTTGTTCCACCCATCT	24	61.1	54.2%	148	90.11
	Act1_REV	GGAACATAGTAGTACCACCGGACATAACGA	30	60.4	46.7%		
AOX1	AOX1_FW	CGTCTTGGGTGGTGGTTCTTCTATC	25	59	52%	145	87.50
	AOX1_REV	GCAAGCTCTTTGGTAGGTCTCAGTC	25	59.2	52%		
E2-Crimson	E2-Cri_FW	GCCAAGCTGCAAGTGACCAAGG	22	61.4	59.1%	136	86.34
	E2-Cri_REV	CGGGGAAGGACTGCTTGAGGTAG	23	61.2	60.9%		
Hac1	Hac1_FW	GGTTGGAAGCCTTAGGTGGTACCG	24	61.4	58.3%	156	84.98
	Hac1_REV	CCTCAGTCAAAGATCTGCGAGTGG	24	59.3	54.2%		
Yap1	Yap1_FW	CAGATGTACTAAGCCTACTGCCGACAAG	28	60.1	50%	144	94.23
	YAP1_REV	CTCCACTACTAGATGACGAACGCAGAG	27	59.7	51.9%		
GCN4	GCN4_FW	TACTGATGCTGACAACTCCCCCATG	25	60.5	52%	157	96.28
	GCN4_REV	CTTCAACGAAGGGACTACTGGAGAGG	26	60	53.8%		
Ssa1	Ssa1_FW	CAGATCACTATCACCAACGACAAGGG	26	58.9	50%	134	88.08
	Ssa1_REV	GACTCCAGAGCGTTTCTAGCTTGG	24	59.3	54.2%		
Kar2	KAR2_FW	GGTAGTCATAACGCCACCAGTAGTCTC	27	59.9	51.9%	148	90.82
	KAR2_RE	GGAATTAACCCAGATGAAGCTGTCGC	26	59.8	50%		
HRD1	HRD1_FW	GCCACATCATCCACTTGAGTTGCTTG	26	60.3	50%	149	95.87
	HRD1_REV	AGGCGGATTGTTATCGGGAAGAGG	24	60.6	54.2%		
Vps10	Vps10_FW	GGTAACAGAATAGCCTACGACGTTGAG	27	58.5	48.1%	147	89.08
	Vps10_REV	GGTGATCTTGTCAACGCTTTGC	22	57.1	50%		
PEP4	PEP4_FW	CCGGCGAAGGTTAAAGTTAAGTCTGG	26	59.5	50%	155	85.33
	PEP4_REV	GCCATCGACACTGGAACCTCATTG	24	60	54.2%		
Erv29	Erv29_FW	CTCATCTTTTCTGGTGGCCGTCG	23	60.1	57%	182	95.98
	Erv29_REV	CCTTGTGCTCCATAAGAGGCAATCC	25	59.7	52%		
Ste24	Ste24_FW	CCATTCAATGCTGTCCTTTCGTTTGC	26	59.2	46.2%	132	87.82
	Ste24_REV	CTCCTTATGAAGGCTGATTAGAGCACTGC	29	60.3	48.3%		
Ero1	Ero1_FW	GGAATGGTGATGAGGGATTCTGCAAG	26	59.6	50%	160	91.82
	Ero1_REV	GAGGACAGCTCATTTTCATCTTGGCCC	27	61.4	51.9%		

Table S1. Primers used for the transcriptional analysis. In this table is all the information from the primers utilized in this study.

7

General conclusions and future outlook

The secretory pathway has been long considered as a major bottleneck in recombinant protein production in eukaryotic systems, hampering the full capacity of well-established cell factories such as the yeast *P. pastoris*. For this reason, the development of new cell engineering strategies and genetic tools to increase the efficiency of the secretory pathway to secrete heterologous proteins have received much attention by the scientific community.

The study presented has addressed in a systematic way a bottleneck in the early steps of the secretory pathway that had been largely neglected in *P. pastoris*, i.e. the translocation step. Our starting hypothesis was that a big portion from the produced protein was not even able to enter the secretory pathway through the Sec61 translocon, and this effect could specially be noted if the produced protein tends to fold in a fast manner. This effect could be increased as the secretion signal commonly used for protein production (α-MF) was driving the proteins through the post-translational translocation and translation and translocation were occurring separately. Based on this hypothesis, we implemented the following engineering strategy for debottlenecking the early steps of the secretory pathway:

- First, we created a chimeric secretion signal containing a signal sequence that was driving proteins through the co-translational translocation, so proteins would be able to be translated and translocated inside the ER simultaneously. The so-called hybrid pre-Ost1-pro-α-MF contained the Ost1 signal sequence as well as the pro region from the α-MF, which allows a rapid export of the protein located at the ER to the Golgi. In addition, we pinpointed an aggregation prone region in the pro sequence that was easily removed by exchanging the Leucine to a Serine in the position 42 (Ser42). All these modifications drastically increased the secretion efficiency at shake flask scale of two different heterologous proteins, E2-Crimson and a microbial lipase from *Bacillus Thermocatenulatus* (BTL2). Notably, these results revealed that a big portion of the produced proteins was being trapped at the cytoplasm and failed entering into the secretion pathway. The creation of a hybrid secretion signal that drives proteins through the co-translational translocation enabled the release of such bottleneck.

- Second, we validated the improved performance of the new secretion signal at a bioreactor scale. To do so, the strains obtained in the first part of the thesis containing either the secretion signal pre-Ost1-pro-α-MF or the pre-pro-α-MF as a reference were cultured in bioreactors operated in fed-batch mode. In addition, the set of structurally different model proteins was extended to a third protein in this study from *Rhizopus oryzae* (ROL). The outcome of this series of cultivations corroborated the positive impact of the hybrid secretion signal on protein secretion for all the model proteins used, to a similar extent to that observed in preliminary small-scale experiments. Most interestingly, the maximum specific growth rate observed for the strains using the improved secretion signal was also increased, reducing the metabolic burden caused by the over secretion of the model proteins with the conventional signal peptide. Thereby, we demonstrated that the hybrid secretion signal supported an improvement not just on protein production, but an increased cellular performance as well. However, the secretion fold improvement was different for each model protein, ranging from 3.5-fold to just a 50% improvement in terms of final protein production. This meant that the effect of the secretion signal is clearly protein dependent. Thereby, we postulate that, if a model protein tends to fold fast, the potential secretion enhancement effect to be seen is going to be much pronounced. Conversely, if the protein tends to fold slowly into its tertiary structure and needs disulfide bonds, the improvement effect in terms of protein secretion would be more moderated.
- Third, tracking intracellular protein down the secretory pathway with the use of the fluorescent model protein E2-Crimson under dynamic conditions, i.e. in cells growing in a fed-batch bioreactor system, revealed new unexpected insights on the physiological impact of the hybrid secretion signal. When using the improved hybrid secretion signal, the model protein rapidly accumulated in the ER during the first few hours of induction. Thereafter, intracellular E2-Crimson fluorescence levels were rapidly decreased. More interestingly, such reduction of intracellular protein was not translated into an increased amount of secreted protein in the extracellular broth, suggesting that the protein was being degraded inside the cells. To investigate further the underlying mechanisms of this observation, a transcriptional analysis was performed along the fed-batch induction phase. The outcome of this

analysis indicated that the UPR response was upregulated early in the induction phase, thereby activating the ERAD pathway as well. Precisely, upregulation of the genes involved with the ERAD machinery correlated with the drastic reduction of intracellular E2-crimson levels. A plausible explanation is that, by means of the improved secretion signal, much more recombinant protein is entering into the secretion pathway, thereby overloading the ER capacity, and triggering the transcriptional activation of the UPR and, consequently, the ERAD. These findings have guided us to propose a new cycle of genetic engineering targets to overcome these new limitations.

Overall, in this study we have shown the importance of the secretion signal as a key element to produce recombinant proteins. Moreover, early testing and validation of new elements such as the hybrid secretion signal at a bioreactor-scale, under process-like conditions (high cell density fed-batch cultivation) has been proven to be key to accelerate translation of new developments of synthetic biology to potential applications in industrial biotechnology. Interestingly, simple changes on the secretion signal conferred a big physiological impact that was just noticeable when studying the secretion signal under bioprocess-like conditions. Specifically, although the improved secretion signal favored the entrance of recombinant protein into the secretory pathway and reduced bottlenecks first observed in the early steps of the secretory pathway , this led to an accumulation of proteins in the ER and generated a new bottleneck that was more perceptible in cultures with longer induction phases (e.g. fed-batch condition). Based on the physiological characterization of fed-batch cultures, we were able to propose a novel strategy to improve protein production under bioprocess-like conditions, pointing at the overexpression or deletion of some genes involved in the UPR and/or ERAD, as well as genes involved in the export of the recombinant proteins from the ER to the Golgi. Alternatively, the pro region from the hybrid secretion signal could be engineered to be recognized more efficiently by Erv29 or even other cargo protein receptors and increase the export rate of heterologous protein to the Golgi.