Understanding Tat export stress and improving production and secretion of heterologous protein.

Isabel Guerrero Montero

A thesis submitted for the degree of Doctor in Philosophy

in Cell Biology

University of Kent

Department of Bioscience

2019

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

Name: Isabel Guerrero Montero

Signature: ……

Date:

Acknowledgements

I would like to thank all of the people that have helped to make this research possible.

Firstly, to Professor Colin Robinson for giving me the opportunity to carry out these projects and his continued support to shape the manuscripts and turn them into papers.

To the CR lab, both past and present, for making these past few years fly by and teaching me so much on the way. To my seatmates Kelly Frain and Alex Jones, for always keeping things lively and interesting, to Julie Zedler and Doris Gangl for paving the way and helping us decipher the bureaucracy of the grant and to Wayne Miller, Andrew Dean, Kirsty Richards, Kelly Walker and Emi Nemoto-Smith, for going well beyond what a postdoc should do and saving my experiments more than once. To Jo Roobol and Ally Walters for making the lab run as smoothly as it did even though we were a handful, to Conner Webb for always making me laugh and to Amber Peswani for the most important job of keeping the biscuit tin well stocked. To Sarah Bischoff and Chillel Jawara for keeping my sanity intact during the final stretch, and finally to our adopted CR members Matt Stewart, Joe Carroll and Connor Sampson for sharing labs, coffee and good cheer and to Tim Kinner, for all those discussions while crafting.

I would also like to thank my fellow ESRs for embarking on this adventure with me. To Daphne Mermans, for being there from the beginning. To Gilles Malherbe and all those at UCB that helped during my secondments. In particular, to Kasia Dolata and the team at Greifswald, for the dozens of discussions in person and via Skype that allowed us to get the manuscripts out there.

To my family, for all their support and finally, to Fran, thank you for coming there and staying here and following me wherever I go.

Abstract

The Twin-arginine translocation (Tat) pathway of *Escherichia coli* has great potential for the export of biopharmaceuticals to the periplasm due to its ability to transport folded proteins, and its proofreading mechanism that allows correctly folded proteins to translocate. The translocation of protein to the periplasm will greatly aid the downstream processing of said biopharmaceuticals, bringing down costs and time necessary for their production. Coupling the Tat-dependent protein secretion with the capability to form disulfide bonds in the cytoplasm of *E. coli* CyDisCo provides a powerful platform for the production of industrially challenging proteins.

To learn more about the Tat pathway and its potential various proteomic studies were conducted: both to study the importance of the Tat pathway and to investigate the effects of exporting a folded substrate (scFv) to the periplasm using a Tat signal peptide, and the effects of expressing an export-incompetent misfolded variant on the *E. coli* cells. These studies identified characteristic changes occurring as a result of a lack of Tat pathway and the production of both a folded and a misfolded protein. Countering and compensating for these changes may result in higher yields of pharmaceutically relevant proteins exported to the periplasm.

To prove that this platform is viable for industrial use scale up experiments were performed. The protein of interest (hGH) was successfully and efficiently exported to the periplasm during extended fed-batch fermentation, to the extent that it was the most abundant protein in the periplasm. The protein was shown to be homogeneous, disulphide bonded and active, and bioassays showed that the yields of purified periplasmic hGH were 5.4 g/L culture.

Taken together these data suggest that the Tat pathway and CyDisCo are attractive platforms for biotechnology. The proteomic studies gave us great insight into the changes occurring in the cells and various targets that would enhance protein production. Scale-up experiments proved to be successful at transferring the technique from the bench to industry.

Table of Contents

Abbreviations

%	Percent
°C	Degrees Celsius
µL	Microlitre
µM	Micromolar
µm	Micrometer
aas	Amino acids
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
bp	Base pair
C	Cytoplasmic fraction
CyDisCo	Cytroplasmic Disulphide formation in *E. coli*
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotides
ddNTP	Dideoxynucleotides
E. coli	*Escherichia coli*
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
Erv1p	Yeast mitochondrial thiol oxidase
g	Gram
H	Histidine (his)
hGH	Human growth hormone
hPDI	Human Protein Disulphide Isomarase
h	Hours
IB	Inclusion bodies
IPTG	Isopropyl β-D-thiogalactoside
kDa	Kilodaltons
L	Litre
LB	Luria Bertani broth
M	Membrane fraction
M	Molar
MBP	Maltose Binding Protein
Min	Minute
mL	Millilitre
mM	Millimolar
ng	Nanogram
OD	Optical density
OMV	Outer membrane vesicles
P	Periplasmic fraction
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PBS-T	Phosphate buffered saline and Tween
PCR	Polymerase chain reaction

PMF	Proton motive force
RNA	Ribonucleic acid
rpm	Revolutions per minute
RR	Twin-Arginine motif
SDS	Sodium dodecylsulphate
scFv	Single-chain variable-fragment from an antibody
Sec	General Secretory pathway
SR	SRP receptor
SRP	Signal recognition particle
Tat	Twin-Arginine Translocase
TorA	Trimethylamine-N-oxide reductase
TorA-	TorA signal peptide fused to following protein
Tris	Tris (hydroxymethyl) aminomethane
Tween20	Polyoxythylenesorbitan monolaurate
UV	Ultraviolet
V	Volts
W	Wash
WT	Wild-type strain
NTP	Nucleoside triphosphates

1. Introduction

1.1 Recombinant proteins and their history

Proteins are large nitrogen-based molecules composed of one or more long amino acid chains. They are an essential part of all living organisms and their study is of the utmost importance since proteins are responsible for a great variety of indispensable functions; from providing structural support to the cells and catalysing biochemical reactions, to acting as signalling molecules that allow the coordination of biological processes between different cells and organs. They are especially involved in defensive responses such as the immune response against viruses and bacteria where they bind to specific foreign particles (antibodies) or other stress responses where they are secreted to the exterior. Given their importance and the many roles that proteins are involved with, problems in their production, by over or underproduction, or in their correct localization can have catastrophic consequences for the organism.

In many cases these protein-based problems are the direct cause of many documented diseases. To treat these diseases the production of that protein which is either lacking or incorrect, be it in structure or localization, is necessary. Initially during the 1950s they were obtained from natural sources such as animals or humans that already produced said protein, which was the case of insulin that was primarily purified from cows and pigs (Lens & Evertzen 1952). However, using proteins from natural sources had many unwanted side effects. For example, Human Growth Hormone (hGH) is a 22 kDa protein consisting of 191 amino acids (Lewis et al. 2018) used to treat burn injuries, wound healing, hipopituitarism and obesity (Isaksson et al. 1985) since it is safe to use even in high doses (van Loon 1998). hGH is naturally synthesized in humans by the somatotropic

cells of anterior pituitary, and historically used to be harvested from the pituitaries of cadavers. The use of this hGH led to several cases of Creutzfeld-Jacob disease, a prion-related disease that causes irreversible nerve damage and in consequence a loss of memory, changes in personality, progressive loss of brain functions and mobility and finally death within a year (Abrams et al. 2011). This led to the necessity to produce safe prion free hGH with a preserved native structure.

To circumvent these problems most biopharmaceutical industries currently employ recombinant proteins as their final commercialized product. Recombinant proteins are proteins obtained by the use of recombinant DNA that has been cloned in a system that supports expression and translation of that gene. Both the sequences of the recombinant DNA and the host system can be manipulated to produce large quantities of protein that is later purified. The high protein production not only facilitates the management of different diseases for the patients at a reasonable cost, but having more protein available per batch allows more assays and research to be performed without protein availability being an issue.

Although recombinant proteins are the keystone of many biopharmaceutical industries, the technology that became its foundation is under fifty years old. It wasn't until 1971 (research published in 1972) that the first recombinant DNA experiments took place: Berg et al showed that they had developed a method for covalently joining DNA molecules to each other and re-circularizing the vector by successfully introducing the galactose operon of *E. coli* into SV40 DNA (a DNA segment containing the lambda phage genes) (Jackson et al. 1972). However, although the technology and knowledge was then available another five years passed before scientists at Genentech managed to express the

first protein from a chemically synthetic gene in E. coli: the somatostatin hormone peptide consisting of 14 amino acids. (Itakura et al. 1977). Since then the number of proteins produced in this manner has grown in an exponential manner and is now the main method for producing commercial proteins of biopharmaceutical worth.

The rapid extension of this technique is also due to the critical advance of another two very interconnected techniques: chain-termination sequencing and polymerase chain reactions. Chain-termination sequencing, or Sanger sequencing, is the addition of ddNTPs (dideoxynucleotides), which lack the 3' hydroxyl group required for the extension of the DNA chain, with dNTPs (deoxyribonucleotides) that do allow for the extension and a DNA polymerase along with the template and initial primer. This mixture will give a final product of varying lengths where the ddNTPs have been inserted and by separating each type of ddNTPs into four different mixtures and running them on a gel side by side the final sequence of the DNA may be easily obtained (Sanger et al. 1977).
The first polymerase chain reactions on the other hand consisted in the use of a polymerase purified from the extreme thermophile *Thermus aquaticus* that had an optimum active temperature at 80 °C (Chien et al. 1976). This allowed for the exponential amplification of DNA by exposing the template, primers, dNTPs and polymerase to repeated cycles of heating and cooling to allow both DNA melting and DNA replication to occur in short periods of time (Mullis et al. 1986).

It quickly became a race to commercialise and improve these techniques, competition driving down prices and making them affordable for many research laboratories. Radioactively or fluorescent labelling for automated detection of ddNTPs for Sanger sequencing and the development of thermocyclers for automated PCRs were soon

commercially available. Although these techniques are now considered essential in any research laboratory interested in protein production and have quickly evolved into efficient procedures, their development did not occur until quite a few years after the first recombinant DNA and protein experiments took place, giving us insight into how fast the field has changed and evolved in such a short period of time. It is not a surprise that for their discoveries both Berg and Sanger shared the Nobel prize in 1980, Mullis following in 1993, since their research laid the foundation for modern biotechnology and protein production.

Finally, another important step necessary for the development of production strains was the study of the vectors themselves. A vector is a DNA molecule used to carry foreign genetic material into another cell where it can be replicated and/or expressed. Our main concern during this study was the use of plasmids, which are circular, double-stranded DNA molecules that are independent from the cell's chromosomal DNA and occur naturally in bacteria, yeast and higher eukaryotic cells. They hold an either parasitic or symbiotic relationship with their host cell and are duplicated by hijacking the cell's own replication machinery before cell division so that during segregation a minimum of one copy of plasmid DNA is present in each daughter cell. They can range in size from a few thousand base pairs to over 100 kb (Lodish et al. 2000). Out of all their characteristics the most important and useful is that they can be engineered for use as cloning and expression vectors by using a multi-cloning site and a selectable marker introduced in the plasmid to assure maintenance.

These plasmids are many and varied, however another key breakthrough for protein production was the development of the *tac* promoters in 1982. The *tac* promoters are a

hybrid derived from the *trp* and *lac* promoters that allowed not only high levels of protein expression (between 7 and 11 times more than the *lac* promoter and between 2 and 3 times more efficient than the *trp* promoter on its own), but the use of the *lac* promoter repression system allows for a tight control of expression (de Boer et al. 1983). The *lac* repression system can be derepressed with isopropyl β-D-thiogalactoside (IPTG), a molecular mimic of allolactose. IPTG binds to the *lac* repressor and releases the tetrameric repressor from the *lac* operator in allosteric manner. However unlike allolactose, IPTG contains a sulphur atom that creates a chemical bond that is non-hydrolyzable by the cells, preventing thus the metabolization and degradation of the inducer and maintaining a constant concentration of IPTG throughout the experiment (Beckwith 2013).

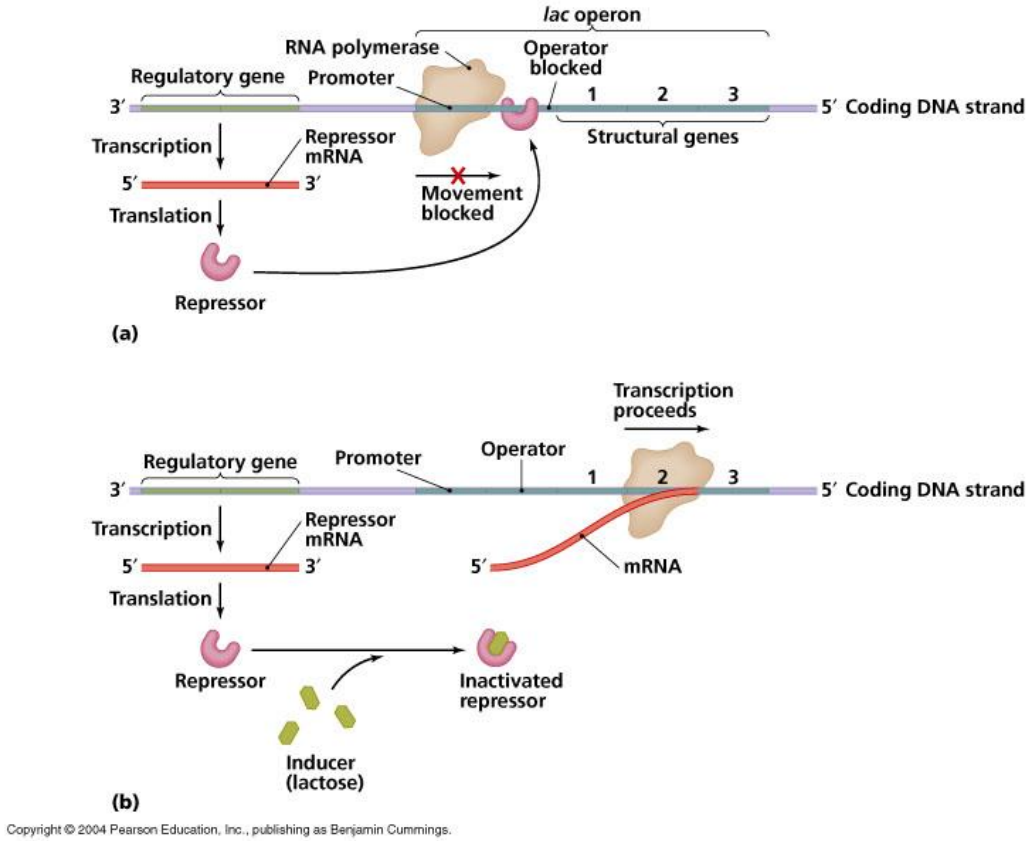


Figure 1: **Mechanism of the (a) repression and (b) induction of the lac promoter in the presence of IPTG.** Taken from (Campbell *et al.*, 2004).

All of these discoveries in unison provided the foundation for the protein production technology we use today and have helped to shape the current biopharmaceutical landscape. Thanks to the development of these vectors protein expression in a large variety of hosts is possible, although each of these hosts carry their own advantages and disadvantages for production and must be chosen carefully and specifically for a determined task depending on the characteristics of the protein of interest.

1.2 Host Systems

Currently almost all biopharmaceuticals are produced in one of three host systems: bacteria (*Escherichia coli*), yeast (*Saccharomyces cerevisiae, Pichia pastoris*) and mammalian cells (Berlec & Štrukelj 2013) among others. Each of these hosts presents their own advantages and challenges for protein production, from lack of post-translational modifications of the protein in bacteria to the ease of contamination and slow doubling time of mammalian cells. In this section, we will look more in depth at the different methods for protein production in each of these hosts.

1.2.1 Mammalian systems

The first host examined will be mammalian cells. Mammalian cells are the most often used expression system making up over half of the market of biopharmaceuticals approved in the United States and the European Union (Walsh 2014). The most commonly used cells are Chinese hamster ovary (CHO), human embryonic kidney (HEK-293), human retina derived (PER-C6), baby hamster kidney (BHK), and mouse myeloma (NS0) (Zhu 2012). Of all the mentioned cells the most widespread in terms of protein production are CHO cells, which account for over 68% of the approved products

produced in mammalian systems as of 2014 (Walsh 2014) since not only do they have the additional benefit of being adaptable to not requiring supplementation with serum, such as fetal calf serum or bovine serum, making them safer since there is less risk of viral and prion contamination as well as making them more cost effective, but they also have been found to not allow the replication of important human viral pathogens (Bandaranayake & Almo 2014). However, although CHO cells present high yields of protein production, in the order of 2-6 g/L of antibody product, human cell lines are demonstrating incredible potential, with the PER-C6 cell line achieving over 25 g/L using high-density cell cultures and perfusion technology (Kuczewski et al. 2011).

The main advantage of the mammalian cells is their ability to create post-translational modifications like gamma-carboxylation, hydroxylation, o-sulphation, amidation and most importantly, disulphide bond formation, proteolytic processing and glycosylation, which in many cases is necessary for the correct functioning of the biopharmaceutical (Walsh & Jefferis 2006). Since both proteolytic processing and disulphide bond formation can also be obtained with other expression systems, glycosylation has thus received the most attention especially for their importance in protein folding and stability, protein targeting and trafficking, ligand recognition and binding, biological activity, protein half-life and in evoking an immune response (Lee et al. 2015). Although glycosylation can be achieved using other eukaryotic systems such as yeast or insect cells, their abilities to replicate the correct glycosylation patterns necessary for human therapeutics are limited (Higel et al. 2016), making mammalian cells the best candidate for the more complex proteins requiring post-translational modifications.

However, the use of mammalian cell lines for recombinant protein production also presents unique challenges. The volumetric yields of protein produced from processes are generally low in comparison to the use of microorganisms (Walsh 2014), making the construction and maintenance of large and costly facilities necessary. These cultures also have a slower growth rate, meaning that to obtain high cell densities the process needs to be extended over time, increasing the chances of foreign contamination. The development of the cell lines are also laborious and time consuming: although transient gene expression is possible in which the heterologous protein is expressed from transfected DNA vectors that have not been integrated into the host genome, these tend to be unstable and not reliably replicated and passed down to the progeny during cell division, meaning that protein expression lasts for typically 10 days post-transfection with variable results between batches (Baldi et al. 2007). As of now, no transient gene expression systems have been approved for manufacturing of biopharmaceuticals since they have not yet proven that the method is apt for large scale production or that it can be effectively replicated between batches to the same standard. To avoid these problems and maintain a uniform performance across batches stable gene expression by integrating the transfected DNA in the genome is essential. However, due to the random nature of integration, there is great variability between clones and a lengthy screening process is necessary to select the best candidates. These candidates will need to pass various phases of screening before establishing a cell bank and evaluating their efficacy for production, increasing the time of development of the biopharmaceutical.

1.2.2 Yeast systems

To decrease the time needed for the development of the biopharmaceutical other eukaryotic systems are further explored. One of the most interesting and lucrative

eukaryotic hosts are yeast: they combine the advantages of both prokaryotes (fast growth, easy genetic manipulation and scale-up as well as inexpensive growth media) and eukaryotes (ability to perform post-translational modifications). Historically, among yeast, *Saccharomyces cerevisiae* is the most widely used thanks to its well-known biochemistry, genetics and molecular biology (Siggers & Lesser 2008), as well as its history of producing heterologous protein for industry thanks to its natural adaptability to the harsh industrial conditions and its ability to correctly produce and secrete biologically active proteins (Huang et al. 2014; Jozala et al. 2016). Nevertheless, in recent years, other yeast are being used as hosts due to their various capabilities. Among these, *Pichia pastoris* (recently reclassified as *Komagataella phaffii* (Kurtzman 2009)), has gained remarkable attention due to the fact that it is an obligate aerobic yeast that can use methanol as a carbon source. This characteristic has allowed for the development of an expression system based on the utilization of the inducible AOX1 promoter (Cregg et al. 2000). When compared to *S. cerevisiae*, *P pastoris* gives higher protein titres since it is not affected by the Crabtree effect (in which respiratory activity is repressed by the presence of free glucose (Deken, De 1966)) and so does not lose carbon by producing ethanol resulting in higher biomass formation and thus more recombinant protein (Mattanovich et al. 2012). *P. pastoris* has been used to produce human insulin, human serum albumin, hepatitis B vaccine, interferon-alpha 2b, trypsin and collagen among others (Vieira Gomes et al. 2018).

Unfortunately, yeast hosts also present a series of disadvantages. Mainly, their glycosylation methods differ slightly from their mammalian counterparts in the way they form both N and O linked oligosaccharide structure on target proteins (Jung & Williams 1997). Particularly, *S. cerevisiae* often adds additional sugars leading to the hyper-

glycosylation of proteins and its N-glycans are highly mannose-rich whereas mammalian proteins are more complex and contain monosaccharides rarely present in yeast including galactose, fucose and sialic acid (Stanley et al. 2017). This limitation of its glycosylation abilities makes them unsuitable for the production of a number of glycosylated human therapeutic proteins. Advances in glyco-engineering of yeast and the expression of recombinant protein have shown significant promise (Hamilton & Gerngross 2007) (Wildt & Gerngross 2005)however the technology is still in the early research stages (Laurent et al. 2016) and it is not possible to predict whether a particular heterologous domain will be effective, although genetic libraries for construct screening and being developed (Nett et al. 2011). Another disadvantage is that expression levels of protein tend to be low in comparison with bacterial hosts and secretion of the protein to the media is ineffective and often limited to the periplasmic space or inclusion bodies, having a significant impact in the recovery of the recombinant protein.

Overall, yeast are a good choice for difficult proteins that cannot be correctly made in bacterial hosts but that are flexible enough in their configuration to not need complex post-translational modifications.

1.2.3 Other eukaryotic systems

Before moving on to prokaryotes there are other lesser used eukaryotic systems of great interest for biopharmaceutical production that we will touch upon briefly. These systems are not as extended as the previously mentioned, however they have great untapped potential and with further development could become competitive host systems in the future market.

The first of these systems is that composed of filamentous fungi. Fungal systems such as Aspergillus and Tricoderma have a high secretion power due to their decomposer lifestyle in which the secretion of determined enzymes is indispensable, and using this to our advantage have allowed us to obtain yields of over 100 g/L (Landowski et al. 2016). However, their complexity and our relatively little understanding of their molecular biology and genetics (Su et al. 2012) have slowed down the development of potential competitive strains for protein production, and more basic research must be done in these areas to achieve efficient producing strains.

The next system are insect cultures. A large number of cell lines have been established for the production of vaccines and viral pesticides, as well as for basic research in biology. The most extended system is the baculovirus system, in which a modified baculovirus (which has a restricted host range and is not infectious to vertebrates making it safe to work with) is used as a vector and infects the insect cells. The virus is typically prepared by inserting the DNA of the protein of interest under the polyhedron growth promoter (*polh* promoter) so that once infected the virus will produce the recombinant protein instead of the protective protein coat (Martínez-Solís et al. 2016). The production levels in this system tend to be very high, however, as the host cell infected with the virus will eventually die the recombinant gene cannot be continuously expressed meaning that they are difficult to use for continuous fermentation as constant re-infection is required (Balamurugan et al. 2006). Another drawback is that, as occurred with the yeast systems, insect cells have different glycosylation patterns such as in the lengths of oligosaccharides and in mannose content (Kost & Condreay 1999) from mammalian cells which can lead to improper recognition of the protein depending on the essentially of said glycosylation.

The final eukaryote host systems are transgenic plants and animals. They offer interesting alternatives to traditional fermentation with both advantages and disadvantages. Protein expression can be localized to different organs by controlling the tissue-specific regulatory sequences (to leaves or seeds in plants or to mammary glands in animals, for example), allowing for easier and constant recuperation of the protein of interest by processing the leaves and seeds or the milk among other strategies. However, the impact of disease and variable cultivation and growth conditions between specimens, as well as the time it takes for said plant or animal to reach a producing age (most usually adulthood) give us great variations in product yield between specimens over time, which does not make them ideal for the production of biopharmaceuticals in which annual targets must be met, although they can be used for medical purposes (Yin et al. 2007).

1.2.4 Bacterial systems

Finally, the last host system we will examine are bacterial hosts. Some bacterial hosts such as *Bacillus subtilis* and *Bacillus megaterium*, Gram positive and catalase positive bacterium found in soil, and *Pseudomonas fluorescens* and *Pseudomonas putida*, Gram negative bacterium found in plants and soil respectively, are being studied for the production of heterologous proteins at a large scale (Retallack et al. 2006) (Biedendieck 2016) (Yue et al. 2017). However, out of all possible bacterial hosts the preferred on the market is *Escherichia coli*, for its rapid growth rate, as short as 20-30 minutes, its capacity for continuous fermentation at high densities and relatively low cost to maintain, and especially its well-known genetics, biochemistry and physiology which makes it easier to be manipulated into ideal hosts (Gibson et al. 2018). For this reason, we will examine the characteristics of *E. coli* in more detail further down.

1.3 Post-translational modifications in *E. coli*

Unfortunately *E. coli* also has some major limitations for protein production. As commented before bacterial hosts have difficulties making the post translational modifications such as glycosylations, which can affect the bioactivity, function, structure, solubility, stability and half-life of functional proteins, acylations, phosphorylations and disulphide-bond formation.

For the first kind of modification efforts are being made to produce strains capable of making mammalian like glycosylations (Ding et al. 2017), however our main interest lies in the last modification mentioned: disulphide-bonds. This modification is especially important for the correct folding of the protein in question and its lack can cause the unfolded protein to become insoluble and form inclusion bodies (Thomas & François Baneyx 1996). There are a few strategies to circumvent this problem, the first would be the export of the protein from the cytoplasm of *E. coli,* where it is made, to the periplasmic space of the cell which has an oxidizing environment and is rich in disulphide bond oxidoreductases, that catalyse the disulphide bridge formation as can be seen in the schematic representation of the process in Fig. 2. This system is ideal for proteins that can be translocated by the Sec translocation system, which facilitates the passage of unfolded protein, one amino acid at a time (Natale et al. 2008) from the cytoplasm to the periplasm. However, since the list of biopharmaceutically relevant proteins is long and varied, many of these proteins will have a quick folding mechanisms that will disqualify them for export via the Sec system.

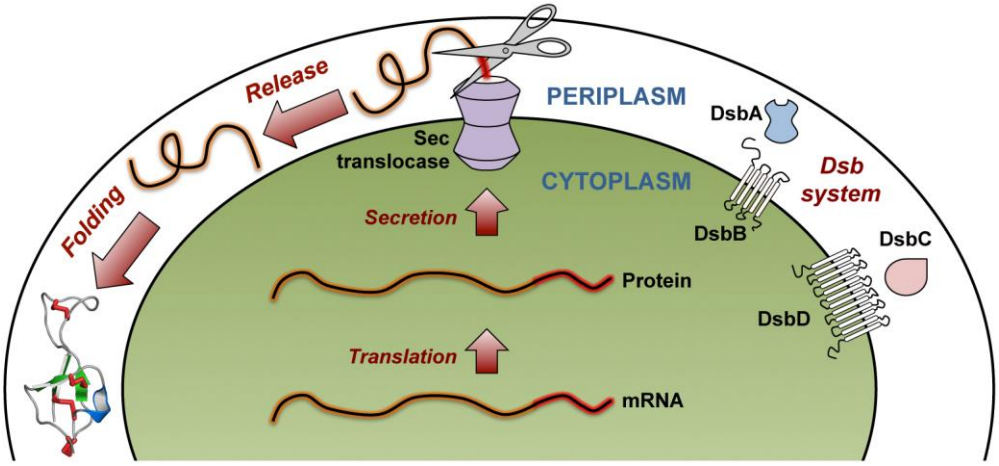


Figure 2: Schematic of the periplasmic expression system for production of disulphide bonded proteins in *E. coli*. After translation, the fusion protein is transported to the periplasm via the Sec translocase system. The MalE signal sequence (red) is removed during this process, releasing the fusion protein (orange) into the periplasm where the Dsb machinery (DsbA, DsbB, DsbC, and DsbD) can assist with disulfide-bond formation. Figure adapted from (Klint et al. 2013).

For those proteins that cannot be exported to the periplasm with the Sec translocation system the disulphide-bonds need to be formed in the cytoplasm. This has been originally done by genetically altering the host system, such as by the removal of the gor and trxB genes which are two naturally occurring reducing pathways as can be seen in Fig. 3, to make a Δgor/ΔtrxB strain which are commercially available from many suppliers. However, this is inefficient since there is no catalyst for the disulphide bond formation (Prinz et al. 1997). In recent years, the development of the CyDisCo system (cytoplasmic disulphide bond formation in *E. coli*) has allowed the formation of disulphide bonds in the cytoplasm while maintaining the reducing pathways intact by expressing a catalyst of disulphide bond formation, sulfhydryl oxidase (Erv1p), and a catalyst of disulphide isomerization, human disulphide isomerase (PDI) (Hatahet et al. 2010). This allows for the efficient production of disulphide bonded eukaryotic proteins in the cytoplasm of *E. coli* (Nguyen et al. 2011). The mechanism of this reaction can be seen in Fig. 4.

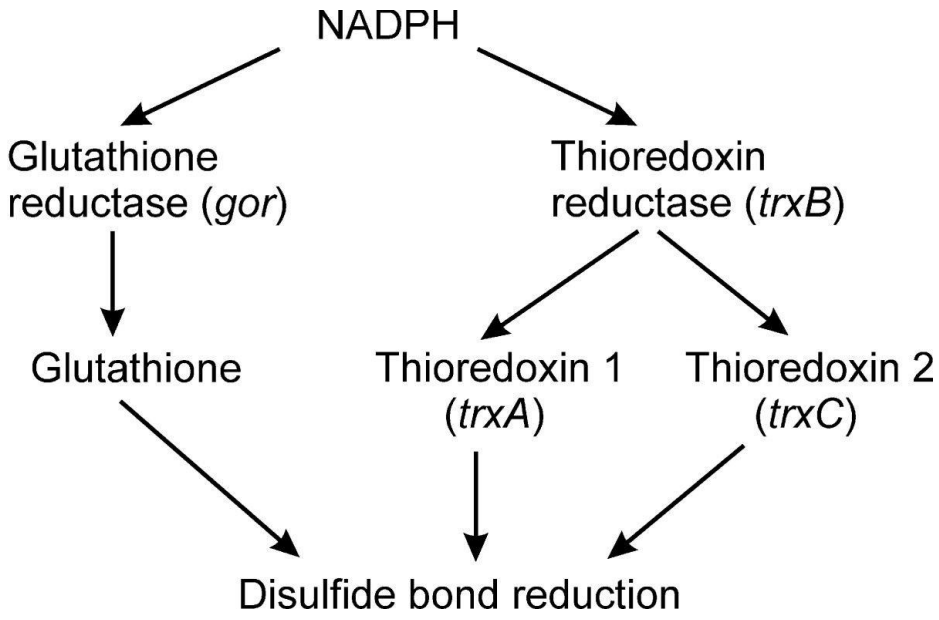


Figure 3: **Dual pathways for disulfide bond reduction in the cytoplasm of *E. coli***. Dual pathways for disulfide bond reduction in the cytoplasm of *E. coli*. On the knockout of both pathways the reduction of disulfide bonds is inhibited, but there is no active system to catalyze their formation. Figure replicated from (Hatahet et al. 2010).

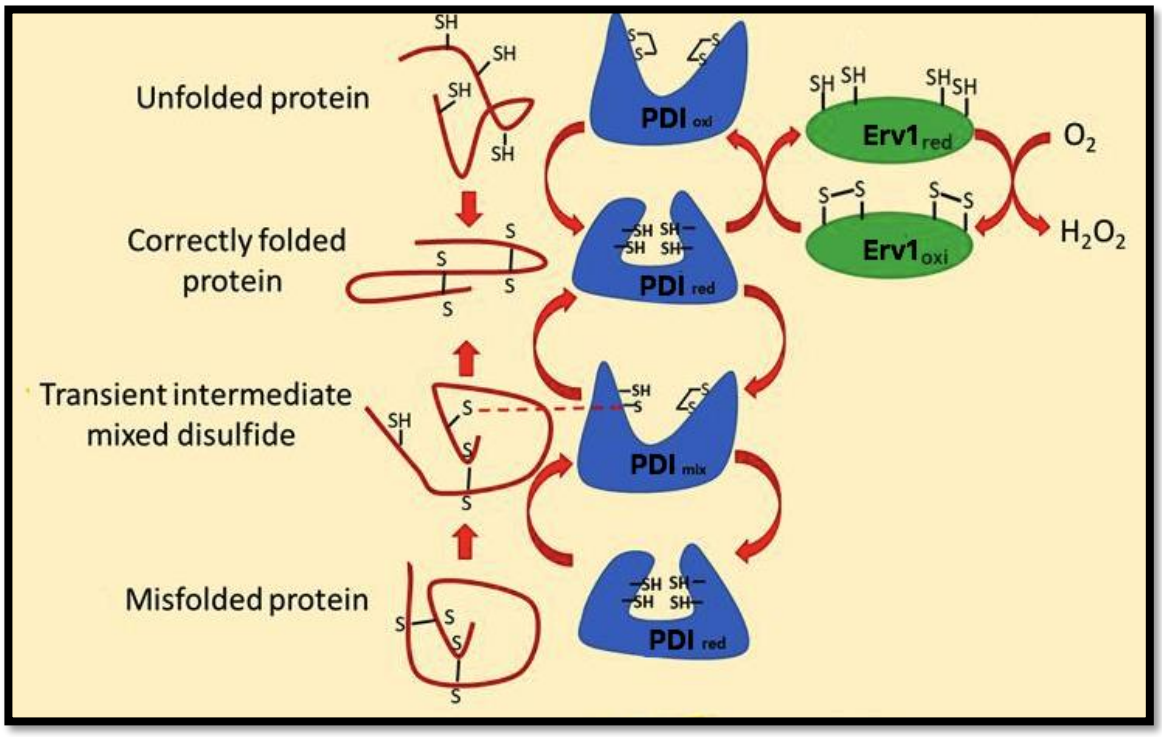


Figure 4: **Schematic showing the correct and incorrect folding of a protein by the CyDisCo system depending on the availability of oxidised PDI.** To maintain PDI in an oxidized form Erv1p must be present. Figure adapted from Tiphany de Bessa, 2018.

1.4 Protein Harvest

1.4.1 Harvest from the Cytoplasm

After obtaining correctly folded and disulphide bonded protein the next step is the harvest and purification of said protein. The first attempts to produce heterologous proteins in bacteria found the first obstacles as protein accumulation in insoluble deposits called inclusion bodies (Thomas & François Baneyx 1996). These inclusion bodies (IB) may be formed for a number of reasons such as high local protein concentration resulting in precipitation, insufficient chaperones to assist with protein folding or absence of PTMs of the foreign protein especially disulphide bridges due to the reducing environment of the cytoplasm. Depending on the method of harvest inclusion bodies can either be avoided or encouraged.

To improve the solubility of the protein and avoid the formation of inclusion bodies a common strategy is to fuse the protein of interest to a soluble protein such as glutathione S-transferase (GST) or maltose binding protein (MBP). GST is a 26 kDa protein from *Schistosoma japonicum* that also acts as a purification tag as it binds to immobilized glutathione that can be used in an affinity chromatography system (Schäfer et al. 2015). MBP is a 42 kDa *E. coli* protein that enhances both solubility and expression, and also has the possibility of being used as a purification tag as it binds to amylose resin, however this fusion has shown low binding capacity for some of the proteins tested, making an additional affinity chromatography steps to achieve purity (Reuten et al. 2016). These fusion proteins can afterwards be easily removed thanks to specific protease cleavage sites engineered for downstream processing.

Taking the opposite approach, by decreasing solubility and placing the protein in inclusion bodies it is possible to harvest said inclusion bodies by differential centrifugation (Palmer & Wingfield 2004). Separating the protein of interest from the rest of the cytoplasmic content is a good way to protect the cells from the effects of toxic products or vice versa by protecting proteins that are particularly susceptible to degradation. The downstream processing of inclusion bodies (IBs) can be a simple operation in which the IBs are solubilized by the use of detergents and chaotropic agents (Yang et al. 2011). This technique is then only suitable for small, easily refolded proteins that will regain their biological activity even after this harsh treatment.

1.4.2 Secretion to the Periplasm via the Sec and Tat Pathways

As we have seen the protein can be harvested directly from the cytoplasmic space in the cells, however secretion of the protein to the periplasm has significant advantages over its accumulation in the cytoplasm (Yoon & Kim 2010; Overton 2014) since the secreted protein is no longer vulnerable to degradation by intracellular proteases and downstream recovery is aided by the lack of contaminating cellular components.

As previously mentioned and shown in Fig 2 the general secretory mechanism to transport proteins from the cytoplasm to the periplasm is by use of the Sec translocation pathway. Sec translocation can occur through in two different ways: through co-translational or post-translational targeting. During co-translational targeting a signal recognition particle (SRP) binds to the signal sequence of the secretory protein while it emerges from the ribosome and the SRP/ribosome/protein chain is targeted to the Sec-translocase (Luirink et al. 2012). A signal recognition particle receptor (SR or FtsY) helps to transfer the SRP to the Sec-translocase. For post-translational targeting the protein is released from the

ribosome and is guided to the translocase by SecB, which recognizes an exportable pre-protein by its N-terminal signal peptide and maintains the proteins in a translocation-competent unfolded state (Driessen 2001). Sec translocase is composed of a protein conducting channel embedded in the membrane. This channel consists of three integral membrane proteins, SecY, SecE and SecG, and an ATPase, SecA, that drives the translocation at the expense of ATP (Mori & Ito 2001). SecA associates to the protein conducting channel and to either the ribosome or SecB to pass the unfolded protein through a narrow transmembrane channel formed by SecYEG (Veenendaal et al. 2004). As we have already discussed this method is only suitable for proteins with slow folding mechanics since Sec only facilitates the passage of unfolded protein. In cases where the protein of interest needs to be folded in the cytoplasm due to its inherent folding mechanics an alternative transport system must be used: the Tat pathway.

The Tat pathway, short for twin-arginine pathway due to the characteristic structure found in its signal peptides which will be touch upon later on, is highly selective for the transport of fully folded proteins (Berks 2015). The Tat system can not only transport folded proteins but it also discriminates against misfolded proteins (Sutherland et al. 2018), implying that there is a quality control or proofreading mechanism that recognizes correctly when proteins are correctly folded and have inserted necessary cofactors or metal ions (Natale et al. 2008)

In *E. coli* the Tat system is made up of four functionally different membrane proteins: TatA, TatB, TatC and TatE (Patel et al. 2014). TatC is the most conserved of the Tat proteins and consists of six transmembrane helices (Behrendt et al. 2004), whereas TatA, TatB and TatE are sequence-related and possess a single N-terminal transmembrane α-

helix (Baglieri et al. 2012). TatC and TatB form a complex that acts as the initial recognition site for Tat substrates, whereas TatA and TatE are putative pore forming components that mediate the actual protein translocation event (Fröbel et al. 2012). A model for the transport mechanism can be seen in Fig 5. The amino acid sequences of TatA and TatE are more than 50% identical and they have overlapping functions (Sargent et al. 1998). It has been shown that the stoichiometry of the TatA/B/C components is crucial for the Tat pathway functioning (Leake et al. 2008), and an increase in the Tat-dependent protein flux is possible only when tatABC are jointly overexpressed (Browning et al. 2017). TatB and TatC interact in a 1:1 stoichiometry (Bolhuis et al., 2001) and low resolution EM structures of TatBC revealed a hemispherical morphology, with an internal cavity where the signal peptide would be inserted (Tarry et al. 2009). Up to 7 copies of TatBC fit into the 11- to 17-nm reconstruction, and apparently, more than one substrate can bind at once (Ma and Cline, 2010). Cross-linking showed TatC is the primary interaction site of the signal peptide (Zoufaly et al. 2012) which inserts deeply into TatC by adopting a hairpin-like conformation (Ramasamy et al. 2013). Moreover, a range of tat mutants have been constructed to study the roles of Tat pathway components and determinate their phenotypes (Lee et al. 2002) Wexler et al. 2000). These studies suggest that both TatB and TatC are essential components of the Tat system and their deletion leads to complete mislocalization of selected Tat substrates. In contrast, single ΔtatA or ΔtatE mutations exhibit versatile defects, but no complete dysfunction in Tat-transported protein localization.

TatBC forms a receptor that binds Tat substrates at the membrane. This triggers TatA recruitment and oligomerization to form an active translocase (Berks et al. 2014). This oligomerization allows TatA complexes to be in the 100- to 500-kDa size range, allowing

for even the largest native Tat substrate, a formate dehydrogenase of approximately 150 kDa (Berks et al. 2000), to pass. Low-resolution images of TatA have revealed a pore-like complex of various diameters (8.5 to 13 nm) that would accommodate various substrate sizes (Gohlke et al., 2005). A lidlike feature was identified which presumably resides on the cis side of the inner membrane, as if the transporter were a trap door (Gouffi et al., 2004). Upon substrate docking onto TatBC, TatA protomers would assemble to suit the substrate size in an oligomeric ring format (Walther et al., 2013). A second model proposes TatA complexes destabilize the membrane when Tat substrate is bound (Rodriguez et al., 2013), by restricting the thickness to its own length. When no substrate is bound, TatA has an amphipathic helix (APH) that immerses itself into the membrane and elongating it. Upon substrate association, TatA reorients its APH outside of the membrane and this switch in conformation would be responsible for the destabilization. (Taubert et al., 2015; Blümmel et al., 2015).

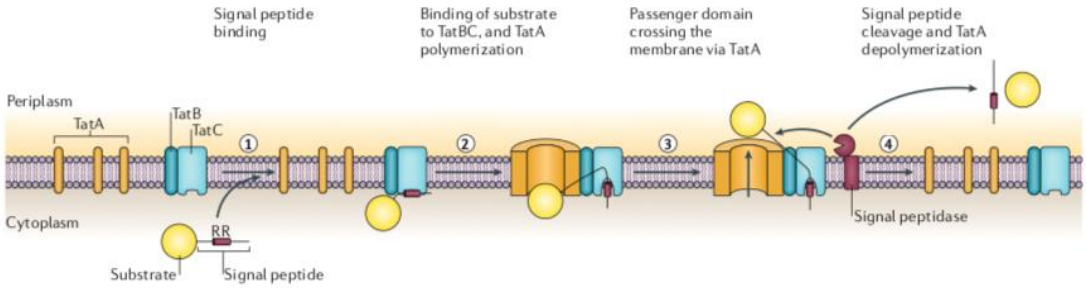


Figure 5: **Model for Tat translocation in *E. coli*.** Step 1: the TatBC complex binds the signal peptide of a substrate protein in an energy-independent step; the twin-arginine (RR) consensus motif in the signal peptide is specifically recognized by a site in TatC. The remainder of the signal peptide and the substrate passenger domain are close to TatB. Step 2: TatA protomers are recruited to the TatBC complex and polymerized. The resulting TatABC complex is the active translocation site. At this stage, the signal peptide is in contact with all three Tat components. Step 3: the passenger domain of the substrate protein crosses the membrane via the polymerized TatA component, and the signal peptide remains bound to the TatBC complex. Step 4: when the passenger domain has reached the far side of the membrane, the signal peptide is normally proteolytically removed by a signal peptidase, and TatA dissociates from TatBC and depolymerizes back to free protomers. Figure adapted from (Palmer & Berks 2012).

Proteins are targeted to the Tat apparatus by an N-terminal signal sequence containing a highly-conserved RR motif that is critical for the efficient recognition of Tat substrates (Stanley et al. 2000). In *E. coli*, at least 27 secreted proteins, see table 1 for a complete list, are predicted to be Tat-dependent and these proteins are involved in important functions e.g. energy metabolism, substrate uptake, cell envelope structure and pathogenicity (Berks et al. 2005). Interestingly, some of the redox proteins are reported to be transported via the Tat pathway yet they do not have a signal sequence of their own. These proteins use a so called ‘’hitchhiker mechanism’’ and form a multimeric complex with another protein containing a Tat signal (Rodrigue et al. 1999). As well as being used for complex proteins, the Tat pathway may be used for the export of slowly folding proteins, that would be exposed to a fast degradation process in a harsh periplasmic environment (DeLisa et al. 2003).

Tat substrate	Uniprot Entry	Protein description
HyaA	P69739	[NiFe] hydrogenase-1 subunit
HybO	P69741	[NiFe] hydrogenase-2 subunit
HybA	P0AAJ8	Electron transfer from hydrogenase-2
NrfC	P0AAK7	Electron transfer to nitrite reductase
NapG	P0AAL3	Electron transfer to nitrate reductase
PaoA/YagT	P77165	Aldehyde oxidoreductase iron-sulfur-binding subunit
YdhX	P77375	Uncharacterized ferredoxin-like protein YdhX
TorA	P33225	TMAO reductase catalytic subunit
TorZ	P46923	TMAO reductase-2 catalytic subunit
NapA	P33937	Nitrate reductase catalytic subunit
YnfE	P77374	DMSO reductase homolog
YnfF	P77783	DMSO reductase homolog
DmsA	P18775	DMSO reductase catalytic subunit
FdnG	P24183	Formate dehydrogenase-N catalytic subunit
FdoG	P32176	Formate dehydrogenase-O catalytic subunit
MsrP/YedY	P76342	Protein-methionine-sulfoxide reductase catalytic subunit MsrP
CueO	P36649	Multi-copper oxidase
SufI/FtsP	P26648	Homologous to CueO [Cu ligands absent]

YahJ	P77554	Uncharacterized protein
WcaM	P71244	Biosynthesis of colanic acid
MdoD/YdcG	P40120	Glucans biosynthesis protein
EfeB/YcdB	P31545	Deferrochelatase/peroxidase
EfeO/YcdO	P0AB24	Iron uptake system component
YaeI	P37049	Phosphodiesterase
AmiA	P36548	Cell wall amidase
AmiC	P63883	Cell wall amidase
FhuD	P07822	Ferrichrome binding protein

Table 1: List of Tat substrates with their description and Uniprot database associated number. Table adapted from (Sargent et al. 2007).

Although the Tat protein export system is not essential for the growth of *E. coli*, the lack of a functioning pathway causes significant growth defects. Blocking the Tat pathway results in mislocalization of Tat-dependent proteins. Among those, amidases AmiA and AmiC were reported to be important for a functional cell envelope phenotype and cell division process (Bernhardt & de Boer 2003; Ize et al. 2003). Their mis-localization leads to formation of long chains of cells, which appear to be defective in cell separation (Stanley et al. 2001).

Recently, attention has been attracted by the application of Tat-dependent transport in protein expression and engineering. The Tat pathway in unison with the CyDisCo system already described is a promising tool for export of complex recombinant proteins and native *E. coli* proteins that fail to be secreted via the Sec pathway (Tinker et al. 2005). In particular, the proofreading mechanism allows for easier downstream purification as there is less protein heterogeneity in the periplasm sample.

1.4.3 A better Tat: TatExpress

As stated, the Tat pathway is used to export a very limited number of proteins to the periplasm and thus although it is constitutively expressed (Jack et al. 2001) the amount

of Tat protein available to form the pathway has always been limited. For this reason, a new strain was created with a pTac promoter introduced before the original promoter of Tat in the genome. This has shown notable increase in protein yields in the periplasm of *E. coli* and can be an interesting alternative to the other host systems we have reviewed in section 1.2.

1.5 Aims of the project

The main aim of this project is to harness the potential of the Tat pathway for its use as a production platform in an industrial setting. Tat offers several advantages over other platforms and its development alongside CyDisCo could be a welcome alternative for complex protein production in a quicker and cost-efficient system. We have set out to test whether native Tat and TatExpress could be scaled-up and produce competing quantities of heterologous protein in comparison with other platforms. To do so it is necessary to also acknowledge the importance of this pathway as well as study the effects its proofreading mechanism has on the integrity of the cell.

2. Publication Summary

2.1 Equal first author publication- Far-reaching consequences of *tat* deletion in *Escherichia coli* revealed by comprehensive proteome analyses.

Katarzyna M. Dolata, **Isabel Guerrero Montero**, Wayne Miller, Susanne Sievers, Thomas Sura, Christian Wolff, Rabea Schlüter, Katharina Riedel, Colin Robinson, Far-reaching cellular consequences of tat deletion in Escherichia coli revealed by comprehensive proteome analyses. Microbiological Research, 218, 97–107.

2.1.1 Contribution

For this project, I contributed in concept design, analysed the proteomic data, reviewed the literature available on proteins that showed important shifts in their regulation, made the figures and wrote the manuscript in collaboration with Katarzyna Dolata. The manuscript was revised by all the authors, especially Colin Robinson, Katharina Riedel and Susanne Sievers.

For this study, we carried out a comparative proteomic analysis of *E. coli* wild type and *ΔtatABCDE* strain and additional experiments to quantify colanic acid was done after the data showed an upregulation of most of the genes in the operon for its biosynthesis. Since there is a link between colanic acid and biofilm formation (Gottesman & Stout 1991) a biofilm assay was also performed to assess whether the *Δtat* strain had triggered this response.

The first set of proteins that was analysed was the effect of the Δtat mutation on the export of Tat-dependent substrates by reviewing how each protein interacts with the tat pathway. Unfortunately, only a handful of proteins were present in the samples, most likely due to the stress conditions in which the cells were grown, however we were able to establish sensible pattern to explain their behaviour and localization.

The next set of proteins analysed were stress response proteins. A set of 98 proteins related to stress responses were categorized into 10 functional groups using Figfam classification and then manually curated. The categories were: protein folding and degradation, stress response, cell envelope stress and cell division and iron and molybdenum homeostasis. Each of these categories was individually studied to determine

the relationship between those proteins and what effect they were having or protecting the cell from having.

Finally, electron micrographs were taken by Rabea Schlüter and interpreted by myself and Katarzyna Dolata.

2.1.2 Publication

The paper as was sent for publication is as follows:

2.1 Far-reaching consequences of *tat* deletion in *Escherichia coli* revealed by comprehensive proteome analyses.

Far-reaching cellular consequences of *tat* deletion in *Escherichia coli* revealed by comprehensive proteome analyses

Katarzyna M. Dolata[a*], **Isabel Guerrero Montero[b*]**, Wayne Miller[b], Susanne Sievers[a], Thomas Sura[a], Christian Wolff[a], Rabea Schlüter[c], Katharina Riedel[a], Colin Robinson[b]

[a] *Institute of Microbiology, University of Greifswald, Felix-Hausdorff-Straße 8, 17487 Greifswald, Germany*

[b] *School of Biosciences, University of Kent, Canterbury CT2 7NJ, U.K.*

[c] *Imaging Center of the Department of Biology, University of Greifswald, Friedrich-Ludwig-Jahn-Str. 15, 17487 Greifswald, Germany*

* Katarzyna M. Dolata and Isabel Guerrero Montero contributed equally to this work

Corresponding author:

Katharina Riedel, riedela@uni-greifswald.de, tel. +49 3834 420 5900[1]

[1]*Abbreviations:* Tat, twin-arginine translocation; Sec pathway, general secretory pathway; cps, capsular polysaccharide; psp, phage shock protein; Fur, ferric uptake regulator

2.1 Far-reaching consequences of *tat* deletion in *Escherichia coli* revealed by comprehensive proteome analyses.

Abstract

In *Escherichia coli*, the Twin-arginine translocation (Tat) pathway secretes a set of folded proteins with important physiological functions to the periplasm and outer membrane. The loss of Tat secretion impairs outer membrane integrity and leads to decreased cell growth. Only recently, the Tat pathway has gained more attention due to its essential role in bacterial virulence and applications in the production of fully folded heterologous proteins. In this study, we investigated the influence of the deletion of all active Tat pathway components on the *E. coli* cells. The comprehensive proteomic analysis revealed activation of several stress responses and experimentally confirmed the dependence of certain proteins on the Tat system for export. We observed that a *tat* deletion triggers protein aggregation, membrane vesiculation, synthesis of colanic acid and biofilm formation. Furthermore, the mislocalization of Tat-dependent proteins disturbs iron and molybdenum homeostasis and impairs the cell envelope integrity. The results show that the functional Tat pathway is important for the physiological stability and that its dysfunction leads to a series of severe changes in *E. coli* cells.

Keywords: *E. coli*, protein secretion, Twin-arginine translocation, proteomics, stress response

2.1 Far-reaching consequences of *tat* deletion in *Escherichia coli* revealed by comprehensive proteome analyses.

1. Introduction

The main systems transporting proteins across plasma membranes in bacteria are highly conserved. Most of the transported proteins are inserted into the membrane or exported to the periplasm via one of the three routes – the Sec machinery, the YidC insertase and the Tat system (Collinson et al., 2015; Lee et al., 2006; Petriman et al., 2018). In *E. coli*, most proteins are transported by the general secretory (Sec) pathway, yet a subset of periplasmic proteins are reported to be exported via the twin-arginine (Tat) pathway (Tullman-Ercek et al., 2007). These two systems use very different mechanisms to transport proteins to the periplasm; the Sec pathway transports proteins in an unfolded state, while the Tat pathway is highly selective for the transport of fully folded proteins (Berks, 2015; Müller and Klösgen, 2005). Remarkably, the Tat system can not only transport folded proteins, but also discriminate against misfolded proteins (Sutherland et al., 2018). This interesting feature implies the existence of a so-called proofreading and quality control mechanism that recognizes correctly folded proteins. However, despite growing research interest, not much is known so far about this mechanism.

Unlike the quality control mechanism, the composition of the Tat translocon is well known. In *E. coli*, the Tat system comprises of four functionally individual membrane proteins - TatA, TatB, TatC and TatE (Bogsch et al., 1998; Patel et al., 2014). TatC is the most conserved of the Tat proteins and contains six transmembrane helices (Behrendt et al., 2004), whereas TatA/B/E are sequence-related and possess a single N-terminal transmembrane α-helix (Baglieri et al., 2018; Koch et al., 2012). Proteins TatC and TatB form a complex, which acts as an initial recognition site for Tat substrates. TatA is a putative pore forming component that may mediate the actual protein translocation event (Fröbel et al., 2012). This complex is recruited to the TatBC complex upon substrate binding (Rose et al., 2013). The amino acid sequences of TatA and TatE are more than 50% identical but these proteins have overlapping functions, with TatB tightly bound to TatC in the substrate-binding complex. It has been shown that the stoichiometry of the TatA/B/C components is crucial for the Tat pathway functioning (Leake et al., 2008), and an appreciable increase in the Tat-dependent protein flux is possible only when *tatABC* are jointly overexpressed (Alami et al., 2002; Browning et al., 2017). Moreover, a range of *tat* mutants have been constructed to study the roles of Tat pathway components and determine their phenotypes (Lee et al., 2002; Wexler et al., 2000). These studies suggest that both TatB and TatC are essential components of the Tat system and their deletion

leads to complete mislocalization of selected Tat substrates. Furthermore, TatB has a function in stabilizing TatC since both proteins are suggested to form a complex. On the contrary, single Δ*tatA* or Δ*tatE* mutations exhibit diverse defects, but no complete dysfunction in Tat-transported protein localization. However, despite the number of studies analyzing *tat* mutants and focusing on the cellular consequences of the Tat secretion impairment in *E. coli*, none of them investigated the cell phenotype under simultaneous deletion of all *tatABCDE* genes.

The Tat pathway operates independently of the Sec and YidC post-translational protein transport systems. Proteins are targeted to the Tat apparatus by an N-terminal signal sequence containing a highly-conserved RR motif that is critical for efficient recognition of Tat substrates (Alami et al., 2003; Stanley et al., 2000). However, it has been shown that signal sequences with a single mutation of an arginine residue are still transported by the Tat system (Ize et al., 2002; Summer et al., 2000). In *E. coli*, at least 29 secreted proteins are predicted to be Tat-dependent. These proteins are involved in important functions e.g. energy metabolism, substrate uptake, cell envelope structure and pathogenicity (Berks et al., 2002; Tullman-Ercek et al., 2007). Interestingly, some of the redox proteins are reported to be transported via the Tat pathway yet they do not have a signal sequence of their own. These proteins use a so-called ‘’hitchhiker mechanism’’ and form a multimeric complex with another protein containing a Tat signal (Rodrigue et al., 1999).

Recently, attention has been attracted by the application of Tat-dependent transport in protein expression and engineering. The Tat pathway is a promising tool for export of complex recombinant proteins and native *E. coli* proteins that fail to be secreted via the Sec pathway (Tinker et al., 2005; Walker et al., 2015). In particular, the proofreading mechanism allows for easier downstream purification as there is less protein heterogeneity in the periplasm sample. Moreover, the role of the Tat pathway in bacterial pathogenesis is under investigation (Ball et al., 2016; Pradel et al., 2003; Ochsner et al., 2002). Although the Tat protein export system is not essential for the growth of *E. coli*, the lack of a functioning pathway causes significant growth defects.

Deletion of the Tat system results in mislocalization of Tat-dependent proteins. Among those, amidases AmiA and AmiC were reported to be important for a functional cell envelope phenotype and cell division process (Bernhardt et al., 2003; Ize et al., 2003). Their mislocalization leads to formation of long chains of cells, which appear to be

defective in cell separation (Stanley et al., 2001). Despite the clear importance of Tat-dependent secretion in *E. coli*, our understanding of its biophysical underpinnings remains vague.

In this study, we have sought to ascertain at a comprehensive level how *E. coli* responds to a defect in the Tat protein export pathway. The deletion mutant Δ*tatABCDE* has been used to determine the *E. coli* phenotype and proteome alterations under disruption of the Tat pathway. We describe several stress responses and novel phenotypes emerging from the Tat system's deletion. Moreover, we have experimentally identified a range of known Tat substrates and studied their localization.

2. Materials and methods

2.1. Cell culture and fractionation

E. coli strains utilized for this study were MC4100 and a variant of said strain with a full *tatABCD* deletion and a partial *tatE* deletion (Table 1). Five mL Luria Bertani (LB) medium (10 g/L sodium chloride, 10 g/L tryptone, 5 g/L yeast extract) pre-cultures were inoculated from glycerol stocks and grown aerobically overnight at 30 °C, 200 rpm. The next day cultures were diluted to an OD_{600} of 0.05 in 25 mL fresh LB. Cultures were then grown at 37 °C, 200 rpm in 250 mL Erlenmeyer flasks for 5 hours. Cells equivalent to a density of OD_{600} of 10 (~12 mL) were taken and fractioned into cytoplasmic (C), membrane (M) and periplasmic (P) samples. Periplasmic (P) fractions were collected using an EDTA/lysozyme/cold osmotic shock method previously described (Randall and Hardy, 1986), with modifications (Pierce et al., 1997). Spheroplasts were further fractionated into cytoplasmic (C) and insoluble/membrane (M) fractions as described earlier (Pierce et al., 1997). Fractions of three biological replicates from each stain were prepared.

Table 1. Strains and plasmids used in this work.

Strains	Description	Reference/source
MC4100	AraR, F2 araD139 DlacU169 rpsL150 relA1 flB5301 deoC1 ptsF25 rbsR	Wexler et al., 2000
ΔtatABCDE (Δtat)	MC4100 strain lacking tatABCDE genes, AraR	Wexler et al., 2000

2.2. Inclusion bodies preparation

Inclusion bodies (IBs) were separated from bacterial cultures grown in conditions described in the previous section. Cells were grown for 5 hours and harvested by centrifugation at 10.000 x g, 4 °C for 10 minutes. Pellets were resuspended in washing buffer (150 mM NaCl, 50 mM Tris-HCl, pH 8.0) and cells were disrupted by French Press (operated at 16.000 to 18.000 psi) followed by a high-speed centrifugation at 22.000 x g. Unbroken cells, large cellular debris and the inclusion body proteins were washed twice with a washing buffer and IBs were isolated as described previously (Joliff et al., 1986). In short, pellets were resuspended in lysis buffer (5 M urea, 50 mM Tris-HCl, pH 8.0), incubated 15 minutes at room temperature and centrifuged for 30 minutes at 22.000 x g. The supernatant containing denatured proteins of IBs was separated and concentrated using Amicon® Ultra-0.5 centrifugal filters (Millipore).

2.3. Proteomics sample preparation and LC−MS/MS analysis

Protein concentration was determined by the Bicinchoninic Acid (BCA) Protein Assay (Thermo Fisher Scientific). Cytoplasmic proteins (100 μg) were reduced with TCEP, alkylated with iodoacetamide and digested in-solution using trypsin (Muntel et al., 2012). Desalting of peptides prior to mass spectrometry analysis using Stage tips, C18 material (Thermo Fisher Scientific) was performed according to the protocol described earlier (Rappsilber et al., 2007). For absolute protein quantification, a tryptic digest of yeast alcohol dehydrogenase (ADH1, Waters, USA) was added into the samples to a final concentration of 50 fmol/μL. The nanoACQUITY^{TM} UPLC^{TM} system (Waters) was used to separate and introduce peptides into the Synapt G2 (Waters) mass spectrometer. Parameters for liquid chromatography and IMS^{E} were used as described previously (Zühlke et al., 2016).

Proteins from periplasmic and membrane fractions (30 μg) were separated via 1D SDS-PAGE, the entire gel lanes cut into ten pieces each. Proteins from isolated inclusion bodies (30 μg) were run via 1D SDS-PAGE 5-6 mm into the resolving gel and the bands containing proteins were cut out of the gel. Proteins were digested with trypsin (Promega, USA) overnight. Peptides were purified using ZipTip C18 tips (Millipore). The eluted peptides were subjected to LC-MS/MS analysis performed on a Proxeon nLC 1200 coupled online to an Orbitrap Elite (Thermo Fisher Scientific) mass spectrometer. Peptides were separated on in-house self-packed columns (id 100 μm, od 360 μm, length

200 mm; packed with 3.6 µm Aeris XB-C18 reversed-phase material (Phenomenex)) in an 80 min nonlinear gradient from 1% acetonitrile and 0.1% acetic acid to 95% acetonitrile, 0.1% acetic acid. To obtain a better separation of peptides of the IBs fraction, peptides were separated on in-house self-packed columns (id 100 µm, od 360 µm, length 200 mm; packed with ReproSil-Pur 120 C18-AQ, 3 µm material (Maisch)) in a 180 min nonlinear gradient. A full MS scan (resolution of 60,000) was acquired using the automatic data-dependent mode of the instrument. After acquisition of the full MS spectra, up to 20 dependent scans (MS/MS) were performed according to precursor intensity by collision-induced dissociation fragmentation (CID) in the linear ion trap. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD008803 (username: reviewer18233@ebi.ac.uk, password: FfgJJudJ).

2.4. Protein identification and quantification

MS/MS spectra of cytoplasmic samples were searched against an *E. coli* K12 MC4100 UniProt/Swissprot database (Proteome ID: UP000017782, 3985 protein entries, version January 2017) with added laboratory contaminants and yeast ADH1 sequence (Zühlke et al., 2016). MS/MS spectra of periplasmic and membrane samples were searched against the above-mentioned database using MaxQuant software (version 1.5.8.3) (Cox and Mann, 2006). The search was performed with the Andromeda search algorithms (Cox et al., 2011). Search parameters were set as follows: a minimal peptide length of six amino acids, up to two missed cleavages, carbamidomethylation of cysteine specified as a fixed modification, N-terminal protein acetylation and methionine oxidation were set as variable modifications. The false discovery rate (FDR) was estimated and protein identifications with FDR< 1% were considered acceptable. A minimum of two unique peptides per protein was required for relative quantification using the label free quantification (LFQ) algorithm provided by MaxQuant.

2.5. Protein functional analysis and subcellular localization prediction

The Student's t test requiring a p-value of < 0.05 was subsequently conducted to determine proteins with significant alterations in protein abundance. Changes in protein abundance of Δ*tat vs.* wild type (WT) were presented with a log2 fold change

(Supplementary Table 1). If the proteins were identified in only one of the strains in three biological replicates, they were added to the list of ON/OFF proteins. Heat maps were generated with Prism 7 (GraphPad) and the protein biological functions were assigned based on their FIGfam roles according to The SEED (http://pubseed.theseed.org/) (Overbeek et al., 2014) complemented with manual curation. The protein subcellular localization was assigned based on the *E. coli* specific STEP database (STEPdb) (Orfanoudaki and Economou, 2014).

2.6. Transmission electron microscopy (TEM)

Cells were fixed (1 % glutaraldehyde, 4 % paraformaldehyde, 0.2 % picric acid, 50 mM sodium azide in 5 mM HEPES buffer pH 7.4) at 4 °C and stored at the same temperature until further processing. Subsequently to embedding in low gelling agarose, cells were postfixed in 2 % osmium tetroxide in washing buffer (50 mM cacodylate buffer pH 7, 10 mM magnesium chloride, 10 mM calcium chloride) for 1 h at 4 °C. After dehydration in a graded series of ethanol (20 %, 30 %, 50 % for 10 min each step; 70 % ethanol with 0.5 % uranyl acetate for 30 min at 4 °C; 90 %, 96 %, 100 % ethanol) the material was embedded in AGAR 100 resin. Sections were cut on an ultramicrotome (Reichert Ultracut, Leica UK Ltd, Milton Keynes, UK), stained with 4 % aqueous uranyl acetate for 3 min followed by lead citrate for 1 min and analyzed with a transmission electron microscope LEO 906 (Zeiss, Oberkochen, Germany). Afterwards, the micrographs were edited by using Adobe Photoshop CS6.

2.7. Crystal violet biofilm assay

A quantitative biofilm assay was performed in 96-well polystyrene plates as described previously (O'Toole, 2011). Briefly, cells were inoculated from overnight cultures grown in LB and Belitsky minimal medium (Stülke et al., 1993), diluted in replicates in both media to an initial turbidity of 0.05 at 600 nm and grown at 37 °C without shaking. Cell density (turbidity at 600 nm) was measured after 24 and 48 hours in a microtiter plate reader (BioTek Synergy™ Mx) and analyzed with Gen5™ version 2.0 software. After removing the supernatants, biofilms were stained with crystal violet and absorbance at 550 nm was measured. Total biofilm was normalized by bacterial growth for each strain and represented as biofilm formation index (BFI). Each data point was averaged from at least 20 replicate wells.

2.8. Quantification of colanic acid

Colanic acid was quantified by measuring L-fucose, the sugar component which is exclusively found in this exopolysaccharide (EPS) (Stevenson et al., 1996). Samples of *E. coli* biofilm, growing in the wells of 6-well polystyrene plates for 48 h, were collected and colanic acid concentration was measured based on previously described method (Obadia et al., 2007). Biofilm samples were boiled for 15 minutes to inactivate EPS-degrading enzymes and release EPS from the cell surface. In the next step, samples were centrifuged at 14.000 x g for 30 minutes at 4 °C, after cooling down to room temperature. The supernatants were collected, and polysaccharides were precipitated with 70 % ethanol overnight. Samples were subjected to centrifugation at 14.000 x g for 30 minutes at 4 °C. The pellet was diluted to 0.2 mL with distilled water and mixed with 0.8 mL of H_2SO_4/H_2O (6:1; v/v). The mixture was heated at 100 °C for 20 min and cooled down to room temperature. For each sample, absorbance at 396 nm and 427 nm, before and after addition of 100 µL of cysteine hydrochloride, was measured. The absorption due to this reaction was subtracted from the total absorption of the sample, as described earlier (Obadia et al., 2007). A L-fucose (Sigma-Aldrich) calibration curve was used (0–100 µg/mL) to determine the fucose concentration. The values for *E. coli* biofilm samples were normalized by cell turbidity at 600 nm.

3. Results and discussion

3.1. Comparative proteomic analysis of E. coli subfractions

In this study, we carried out a comparative proteomic analysis of *E. coli* wild type and Δ*tatABCDE* strains. To efficiently identify and quantify proteomic changes, three different subcellular fractions were investigated, i.e. periplasmic and cytoplasmic protein fractions as well as an insoluble protein fraction containing mainly membrane proteins. The application of subcellular compartment fractionation was previously shown to improve protein identification, especially for membrane and periplasmic proteins (Brown et al., 2010). However, the common limitation of subfractionation methods are the impurities of highly abundant cytoplasmic proteins in periplasmic and membrane fractions (Brown et al., 2010; Wolff et al., 2008). To validate a high resolution and good depth coverage of periplasmic and membrane proteins in the subcellular fractions, we have calculated the abundance of protein impurities in all the fractions based on signal

intensities (Table 2). As expected, the analysis of relative protein abundances points to cross-contamination in all the subfractions of *E. coli* MC4100 and to significantly higher levels of cytoplasmic contaminants for membrane and periplasmic fractions of the *E. coli* Δ*tat* mutant. These results are not surprising, given the fact that some *tat* deletion strains have been reported to present cell envelope disruption and cytoplasmic protein leakage (Ize et al., 2003). Based on the high content of cytoplasmic proteins in the periplasmic fraction of the *E. coli* Δ*tatABCDE*, but not the control strain, we conclude that the complete deletion of the Tat system leads to a disruption of the inner membrane and cytoplasmic protein leakage during fractionation. Collectively, our results indicate that the fractionation protocol distinctively enriched subcellular fractions, however, due to the Δ*tat* membrane impairment, it was not possible to obtain a comparable protein enrichment rate for both strains.

Table 2. Distribution of enriched proteins and protein impurities observed in subcellular fractions via LC-MS/MS.

	E. coli MC4100		E. coli MC4100 Δtat	
Cell fraction	Total recovery [%][a]	Protein impurities [%][b]	Total recovery [%][a]	Protein impurities [%][b]
Cytoplasm	76.58	23.42	73.23	26.77
Periplasm	66.88	33.12	31.75	68.25
Membrane	66.88	33.12	50.27	49.73

[a] Total recovery shows the percentage of the total abundance, calculated from LFQ intensities, of proteins identified in a certain subfraction and assigned to this subfraction by STEPdb.
[b] Protein impurities shows the percentage of the total abundance, calculated from LFQ intensities, of proteins identified in a certain subfraction but assigned to other subfractions by STEPdb.

3.2. Pleiotropic effects in an E. coli tat mutant

Various reports have demonstrated that Δ*tatAE* and Δ*tatC E. coli* mutants display pleiotropic phenotypes. The most prominent effect of the Tat-system impairment is the formation of filamentous, chain like structures (Fig. 1), comprised of cells which fail to undergo a complete cell division (Stanley et al., 2001). This phenotype is assumed to be the result of mislocalization of amidases AmiA and AmiC. Both proteins are involved in splitting the murein septa and the separation of cells. It has been shown that the chain forming phenotype can be complemented by overexpression of AmiB (Bernhardt et al.,

2003). Here, we decided to investigate if the simultaneous deletion of *tatABCDE* will contribute to the discovery of novel phenotypes of the *E. coli tat* mutants.

In addition to previously reported findings, this study revealed an increase in inclusion body formation (Fig. 1B) and membrane vesiculation in the *E. coli tat* mutant (Fig. 1C) compared to the wild type (Fig. 1A). Inclusion body formation occurs naturally in bacteria as a result of the accumulation of unfolded proteins. In our study, the *tat* mutant strain formed inclusion body-like aggregates in moderately higher amounts than the wild type. Thus, we conducted a proteomic analysis to investigate whether there is any difference in the composition of the IB fraction between the wild type and the mutant (Supplementary Table 2). Several cytoplasmic proteins such as GroEL and DnaK, the main drivers of protein folding (in concert with their co-chaperones DnaJ-GrpE and GroES, respectively), IbpA and ClpB have been found to be more abundant in the IB fraction of the mutant strain compared to the WT. Together, these proteins have been already identified as real cytoplasmic components of *E. coli* IBs (Carrio et al., 2002) and GroEL itself may promote IBs formation via clustering of small aggregates. Interestingly, we could identify the cell division protein ZipA in Δ*tat* IBs. This protein is essential for cell division, as it binds with FtsZ protein and forms the septal ring structure that mediates cell division in *E. coli* (Brown et al., 2010). Together with the mislocalization of AmiA and AmiC, the lack of functioning of ZipA may contribute to the formation of long-chains of cells in the *E. coli* Δ*tat* mutant.

The second novel phenotype we observed in Δ*tat E. coli* cells was the formation of outer membrane vesicles (OMV). Several cells formed OMVs, especially the cells who failed to undergo complete division (Fig. 1C). This phenotype is most likely induced in response to the state of the cell envelope and the accumulation of overexpressed or misfolded proteins. In conclusion, the Δ*tatABCDE* mutation in *E. coli* results not only in the formation of long chains of cells, as previously reported for Δ*tatAE* and Δ*tatC* mutants, but also in the production of OMVs and increased formation of IBs, which may serve as a complementary mechanism for managing stress at the cell envelope.

2.1 Far-reaching consequences of *tat* deletion in *Escherichia coli* revealed by comprehensive proteome analyses.

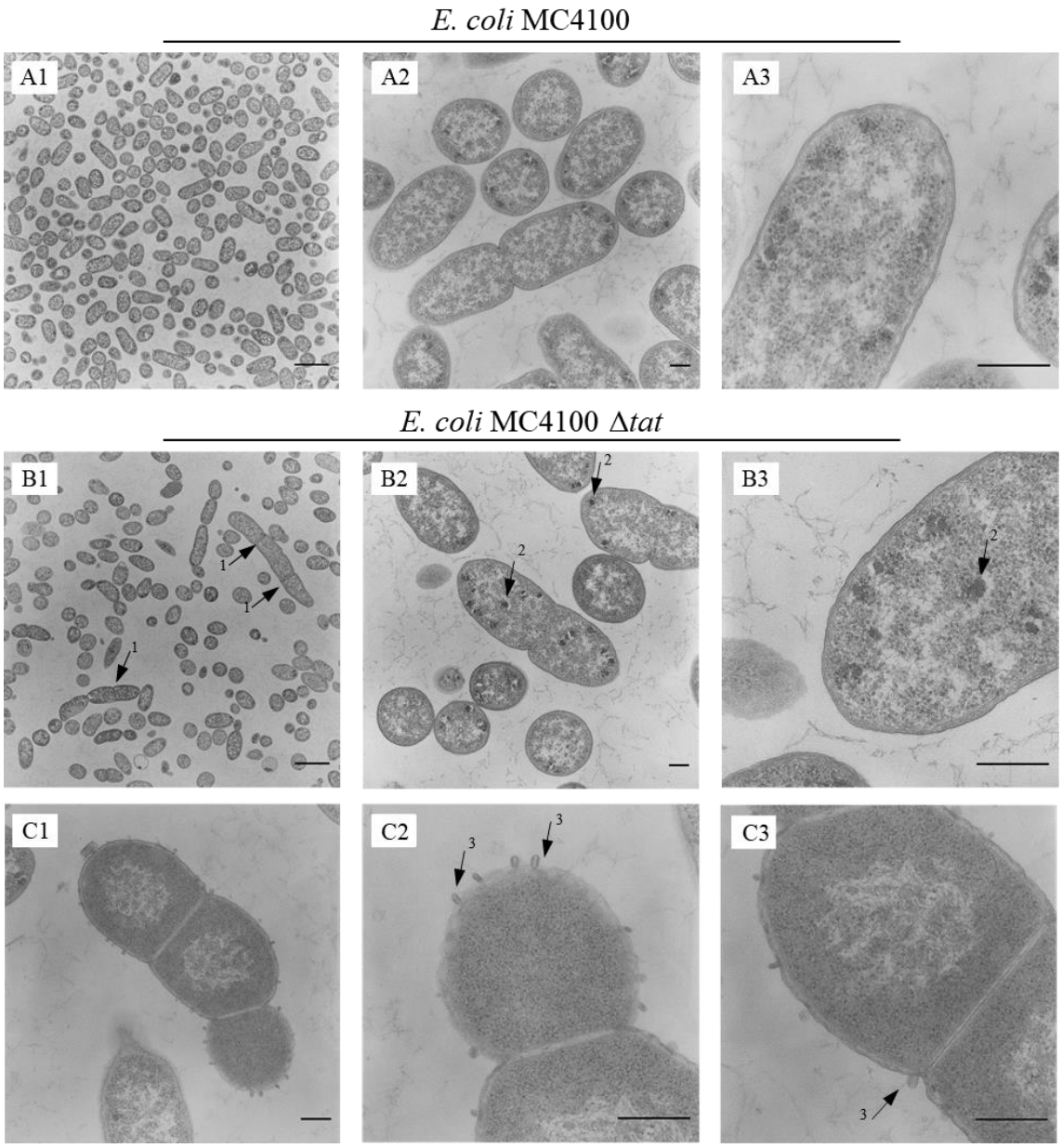


Figure 1. Transmission electron micrographs of *E. coli* MC4100 (WT) (A1-A3) and Δ*tat* (B1-B3, C1-C3) strains at different magnifications. The cells were grown aerobically in LB medium for 5 hours. The mutant strain shows morphological changes such as formation of 1. Long chains of cells, 2. Inclusion bodies and 3. Outer membrane vesicles. Scale bars A1, B1 = 2 µm; scale bars A2, A3, B2, B3,C1, C2, C3 = 250 nm.

3.3. The effect of Δtat mutation on the export of Tat-dependent substrates

The genome of *E. coli* is predicted to encode between 22 and 34 proteins with a Tat signal peptide (Bendtsen et al., 2005; Dilks et al., 2003; Robinson and Bolhuis, 2001; Tullman-Ercek et al., 2007). These predictions are primarily based on the properties of the signal peptides; however, confirming their status as Tat dependent substrates is difficult due to the fact that some of the signal peptides are also capable of directing export of heterologous proteins to the periplasm through the Sec pathway (Blaudeck et al., 2001; Blaudeck et al., 2003; Stanley et al., 2002) and this pathway is essential for the viability of the cells (Economou, 2005). Current studies have shown that the Tat system is used to directly export 27 proteins to the periplasm and indirectly export another 8 using the hitchhiker mechanism described before (Berks et al., 2005; Tullman-Ercek et al., 2007). Out of the 27 reported, we only identified 8 Tat substrates and 3 of the hitchhiker proteins (Fig. 2). The presence of so few Tat substrates is most likely due to the low stress conditions in which the cells were grown, as well as the phase in which they were harvested (exponential). Many of Tat-dependent proteins have been found to be expressed only under certain conditions such as anaerobic growth (HyaA, HybO, NrfC, TorA and TorZ) (Hussain et al., 1994; Méjean et al., 1994; Sawers and Boxer, 1986), presence of cinnamaldehyde at a low pH (YagT) (Neumann et al., 2009) or hydroxyurea (FhuD) (Davies et al., 2009). The expression of some of the Tat substrates can be also inhibited by other proteins like NarL (YnfE, YnfF, YnfG and once again TorA) and NarP (FdnH) (Constantinidou et al., 2005). The rest of the Tat substrates have been shown not to be essential for the growth conditions tested (SufI, YahJ, MdoD, YaeI, Ycb and FdoH) (Gerdes et al., 2003). Within these 8 identified Tat substrates and 3 hitchhiker proteins we could discern 3 different localization and expression profiles.

2.1 Far-reaching consequences of *tat* deletion in *Escherichia coli* revealed by comprehensive proteome analyses.

Category	Accession number UniProt	Protein biological function	Protein	Log2 fold change		
				C	M	P
Tat Substrates	P0AAJ8	Hydrogen oxidation	HybA			OFF
	P31545	Iron transport	EfeB			OFF
	P36548	Cell wall amidase	AmiA		1.2 *	OFF
	P63883	Cell wall amidase	AmiC		0.7	OFF
	P36649	Copper homeostasis	CueO		ON	-2.8 ***
	P33937	Nitrate reduction	NapA	0.7	1.2	-12.3 **
Membrane bound Tat substrates	P24183	Formate oxidation	FdnG		-2.7 *	
	P18775	DMSO reduction	DmsA		-7.7 *	
Hitchhiker proteins	P18776	DMSO reduction	DmsB		OFF	
	P0ACD8	Hydrogen oxidation	HyaB	OFF		
	P0ACE0	Hydrogen oxidation	HybC	-1.5		

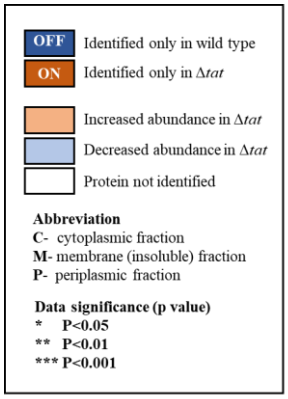


Figure 2. Protein abundance of Tat substrates in *E. coli* Δ*tat vs.* WT. The table shows Tat substrates their abundance which increased or decreased due to the Tat pathway deletion. Proteins are grouped according to their localization pattern. The functional classification and accession numbers were adopted from UniProt. The average ratios (log2 fold change) of the protein abundance in Δ*tat vs*. WT are shown. The fold change in the abundance is marked with shading; blue – decreased abundance, orange- increased abundance. Significantly different results were marked with asterisks, (*) P<0.05, (**) P<0.01, (***) P<0.001.

The first four proteins examined were HybA, EfeB, AmiA and AmiC as they were found to be exclusively present in the periplasm of the WT, but not in the Δ*tat* strain, strongly suggesting that they are exported exclusively by Tat. HybA and EfeB (previously YcdB) seem to be lacking entirely in all other fractions of Δ*tat* strain, presumably due to degradation in the absence of export. The up-regulation of *hybA* and *efeB* genes expression in the *E. coli* Δ*tatC* strain has been already shown (Ize, et al., 2004), thus we assume that the absence of HybA and EfeB in our Δ*tat* mutant might be due to degradation inside the cell. On the other hand, the levels of AmiA and AmiC accumulated in the membrane fraction of the WT as well as in the Δ*tat* mutant. This suggests that there is an inherent stability in these proteins that allows for an excess of protein to associate with the membrane without sequestration in insoluble bodies. When analyzing the content of IBs formed in the Δ*tat* strain, no traces of these proteins were determined. This suggests that their presence in the membrane fraction is due to the association of these proteins with the inner membrane of the cells, but not in inclusion bodies.

Another two Tat dependent substrates, CueO and NapA, were identified in the periplasm of the Δ*tat* strain. However, they were not as abundant and found in much lower quantity than in the WT, suggesting that instead of using Sec as a secondary translocation mechanism these traces of protein are artifacts of the fractionation process discussed before. When comparing the amount of protein present in WT and mutant, we found that both proteins accumulate in the membrane fraction of the Δ*tat* strain. In the case of CueO this accumulation only happens in Δ*tat*, implying that the protein either associates with the membrane or causes aggregation into inclusion bodies.

The next group contains two Tat substrates, FdnG and DmsA. Due to their tight association with inner membrane proteins they are not present in the periplasmic fraction. FdnG is associated with the transmembrane protein FdnI, whereas DmsA form a complex with DmsB and DmsC (Stanley et al., 2002). Both FdnI and DmsC are not present in our samples, probably because of the difficulty to identify transmembrane proteins using the proteomics methods described in this study (Rabilloud, 2012). We observe a dramatic decrease in abundance of FdnG and DmsA in Δ*tat* cells. The presence of FdnG and DmsA in the Δ*tat* cells is most likely due to their association to the inner membrane on the cytosolic side. Since we did not identify them in the IB fraction, we assume they do not aggregate.

The final group of Tat substrates are the hitchhiker proteins: DmsB, HyaB and HybC. DmsB is completely undetected in the Δ*tat* cells. Both HyaB and HybC form multimeric complexes with HyaA and HybO, respectively to be correctly co-transported across the membrane with the help of the chaperons HyaE and HybE (Berks et al., 2005; Palmer et al., 2005). Neither HyaA nor HybO were identified, although they are both located upstream in their operons, followed by their hitchhiker counterparts. However, both of these proteins have transmembrane domains (Hatzixanthis et al., 2003) meaning that, as described before, their identification is analytically challenging. HyaB was not present in Δ*tat* cells and HybC was downregulated, however the abundance values were not statistically significant. In both cases the amount of protein we expect to find are minimal since both operons are under promotors that are induced under anaerobiosis. It is likely then that both HyaB and HybC are either being produced on their own, or alongside HyaA and HybO, but these last two are degraded before they can be exported since the cells are not under anaerobiosis, meaning that their hitchhikers are not able to cross the inner

membrane to the periplasm and in the case of HyaB in *Δtat* cells begin being degraded as well.

3.4. Proteome changes linked to the loss of Tat export

Deactivation of the Tat apparatus causes profound stress in *E. coli* cells. The global analysis of gene regulation has shown the effect of *tatC* deletion on the expression of genes involved in cell envelope associated functions, iron and copper homeostasis and polysaccharide synthesis (Ize et al., 2004). However, to our knowledge, there have been no studies on the effects of *tatABCDE* deletion on the *E. coli* proteome. Thus, in this study we decided to characterize the effect of the loss of all known Tat secretion components on proteins located in the periplasm, membrane, and cytoplasm. A set of 98 proteins related to stress response were identified and represented in the heat map (Fig. 3) (Supplementary Table 3). Changes in protein abundance between the *E. coli tat* mutant and its parental strain were shown as log2 fold changes.

2.1 Far-reaching consequences of *tat* deletion in *Escherichia coli* revealed by comprehensive proteome analyses.

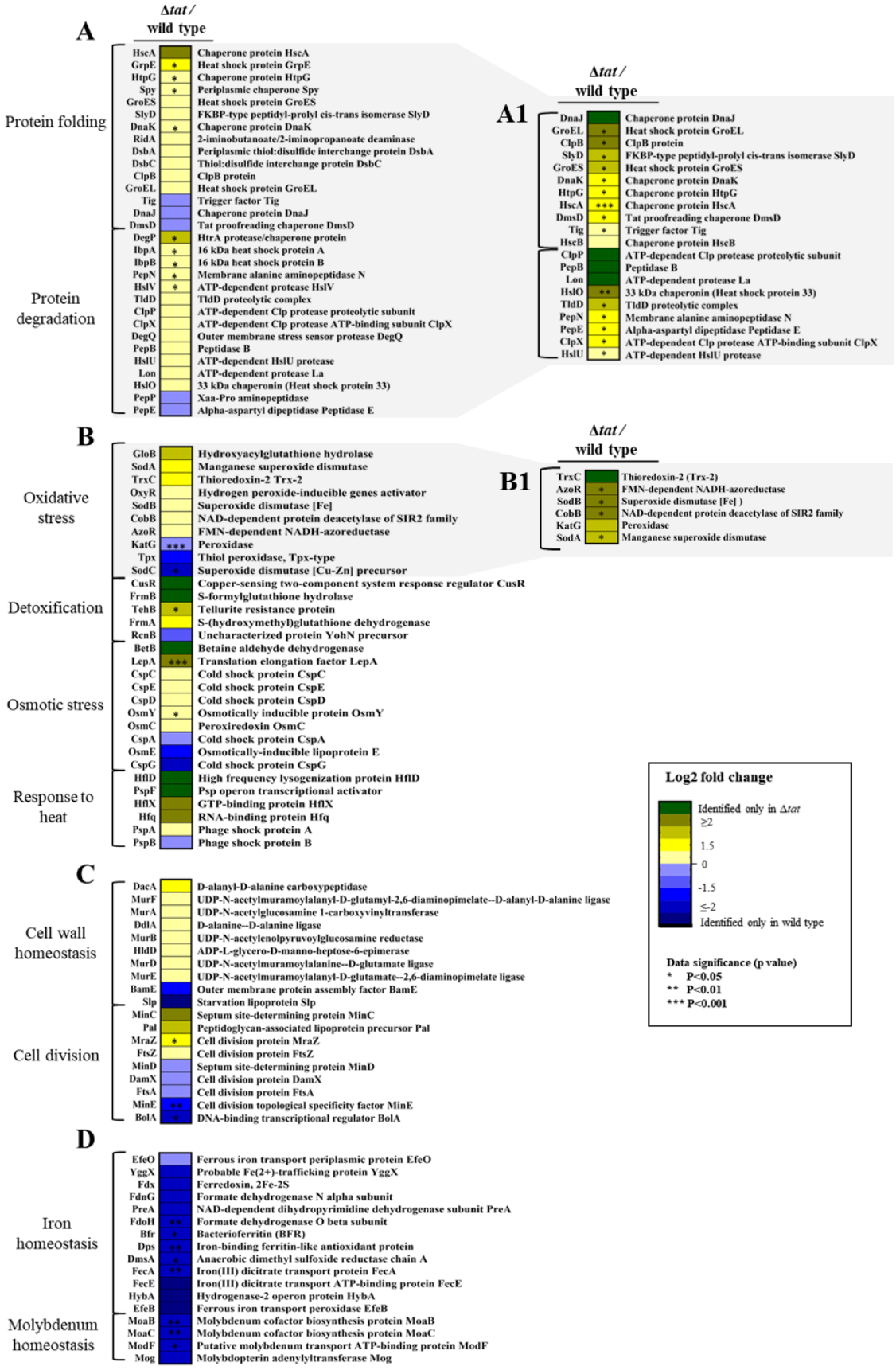


Figure 3. Heat map of changes in the abundance of stress response proteins between *E. coli* Δ*tat* and the WT. A set of 98 proteins related to stress responses were categorized

into 10 functional groups using Figfam classification. The heat map shows proteins involved in: protein folding and degradation (A, A1), stress responses (B, B1) cell envelope stress and cell division (C) and iron and molybdenum homeostasis (D). Proteins represented in panels A1 and B1 were identified in high abundance in the periplasmic fraction, however according to STEPdb they are cytoplasmic proteins. The changes in protein abundance between *E. coli* Δ*tat* and the parental strain are represented as mean log2 fold change and are depicted in a color gradient from blue (decreased protein abundance in Δ*tat*) to yellow-green (increased protein abundance in Δ*tat*). Significant results were marked with asterisks, (*) P<0.05, (**) P<0.01, (***) P<0.001.

3.4.1. Activation of chaperones and proteases in periplasm

E. coli cells upregulate the expression of chaperones and proteases in response to cellular and environmental stress. We observed that the abundances of several chaperones and proteases were increased in the cytoplasmic and periplasmic fractions of the *tat* mutant cells (Fig. 3A). Among the upregulated chaperones in the cytoplasm, the most prominent increase in the protein abundance was observed for the HscA chaperone which is required for the assembly of iron-sulfur clusters (Takahashi and Nakamura, 1999). Moreover, two of the major chaperone systems DnaK/J/GrpE and GroEL/GroES also become upregulated as a consequence of the *tat* deletion. DnaK is one of the most abundant and essential stress inducible chaperones (Bukau and Walker, 1989), it works together with its co-chaperones DnaJ and GrpE to facilitate de-novo protein folding in stressed *E. coli* cells. GroEL is accompanied by its cofactor GroES, and together they are essential for viability under all growth conditions tested (Favetet al., 1998). Furthermore, a number of additional proteins, including ClpB, HtpG, IbpA/B and HslO, are induced under stress conditions to prevent and/or repair stress-induced damage in the *E. coli* proteome (Hoffmann et al., 2004; Schröder et al., 1993; Thomas and Baneyx, 2000). Proteases work together with chaperones and form the cell quality control system of *E. coli*. These proteolytic enzymes degrade damaged, misfolded or unassembled polypeptides, which become harmful when accumulated in the cell. The main cytoplasmic proteases such as Lon, ClpP/X and HslV/U are present in the cytoplasmic fraction of *E. coli* Δ*tat* in high amounts. The periplasmic fraction shows activation of DegP/Q proteases. In addition, we observed the increase in the abundance of cytoplasmic proteases and chaperones in the periplasmic fraction of the Δ*tat* (Fig. 3A1). We assume that most of the cytoplasmic

proteins observed in the periplasmic fraction are technical impurities of the osmotic shock separation process (Brown et al., 2010; Yaagoubi et al., 1994). However, considering their upregulation in both the cytoplasmic and periplasmic fractions, the induction of chaperones and proteases in the *tat* mutant remains unquestionable. We also hypothesize that the upregulation of some cytoplasmic proteins such as GroEL and DnaK may suggest their actual presence in the periplasm. It has been previously reported that a high load of cytoplasmic aggregates can result in their secretion to the periplasm together with GroEL and DnaK (Mar Carrió and Villaverde, 2005). In conclusion, the observed increase in the abundance of chaperones and proteases is most likely a reaction to the perturbations caused by mislocalization and cytoplasmic accumulation of Tat substrates, and the loss of their function in the cell, which disturbs the whole network of proteins that are directly or indirectly involved in Tat secretion.

3.4.2. Cellular stress responses

Our analysis has also revealed the induction of proteins involved in the stress response to heat, oxidation, changes in osmolarity and cold (Fig. 3B). Interestingly, we observed that PspF, the transcriptional activator of the stress-induced *psp* operon, is present only in the *E. coli tat* mutant. The Psp response is induced by perturbations in the integrity and energization of the inner membrane. PspF initiates the Psp response which has been shown to become active with defects of the Sec secretion system (Jones et al., 2003) and enhances the efficiency of the Tat pathway function (Delisa et al., 2004) when the translocation apparatus becomes saturated. Furthermore, most of the upregulated oxidative stress related proteins were also identified in the periplasmic fraction (Fig. 3B1). For example, thioredoxin TrxC which reduces disulfide bridges and methionine sulfoxide formed by reactive oxygen species during the oxidative stress. TrxC is induced by OxyR in the cytoplasm, however it is also present in the periplasm of the *tat* mutant. It has been shown that some small cytoplasmic proteins such as thioredoxin and superoxide dismutase (SodA) can be released from the cytoplasm through the mechanosensitive ion channel of large conductance (MscL) during osmotic shock (Ajouz et al., 1998; Krehenbrink et al., 2011).

3.4.3. Cell envelope stress

It has been previously observed that *E. coli tat* mutants have impaired membrane integrity which results in outer membrane permeability and depletion of peptidoglycans (Ize et al., 2004). One possible explanation for this defect of cell envelope integrity is the mislocalization of amidases AmiA and AmiC, involved in cell wall metabolism. To combat stress resulting from the disruption of the peptidoglycan layer and compensate for its depletion, *E. coli* upregulates proteins involved in the cell wall biogenesis pathway. The proteomic analysis revealed an increased abundance of proteins MurA/B/D/E/F (Fig. 3C), which take part in the synthesis of murein. Most likely, the cell senses the depletion of peptidoglycans in the membrane and escalates their synthesis. In addition, we observed an increase of proteins involved in colanic acid synthesis. This polysaccharide is produced in response to changes in the environment, that may damage the outer membrane, and it protects the cell capsule of *E. coli* (Gottesman and Stout, 1991). The upregulation of proteins involved in colanic acid synthesis is described in the next section (3.5). The phenotypic analysis has shown that Δ*tatABCDE*, similar to Δ*tatC* (Ize et al., 2004), is impaired in the cell separation stage of cell division. Here, we identified some important proteins for the cell division process which may be co-responsible for the formation of long chains of cells in the *E. coli tat* mutant. We determined the upregulation of MinC, cell division inhibitor protein (Conti et al., 2015), and MraZ whose overproduction was reported to be lethal and perturb cell division (Eraso et al., 2014). Together with these findings we observed a dramatic downregulation of transcriptional regulator protein BolA which is involved in the coordination of genes that control cell physiology and cell division (Aldea et al., 1988).

3.4.4. Iron and molybdenum homeostasis

The Tat secretion pathway is involved in the transport of iron. Many of the known Tat substrates bind Fe-S cofactors, and EfeB is directly involved in the iron uptake by extracting iron from heme in *E. coli* cells. We believethat *tat* deletion regulatory influences the expression of the ferric citrate transport (*fec*) operon. We observed that FecA is not present in the periplasmic fraction of Δ*tat* and it is significantly downregulated (-10 log2 fold change) in the insoluble fraction. Since FecA binds $(Fe^{3+}$ citrate$)_2$ to the outer membrane and initiates transcription of the *fec* transport genes, its absence in the cell most likely results in cessation of the Fec-dependent transport, which

explains the absence of FecE transporter in the *E. coli tat* mutant. In addition, we found several other proteins e.g. YggX, Dps, Bfr, Fdx, PreA and FdoH, important for maintaining the cell iron homeostasis, to be downregulated in the of *E. coli* Δ*tat* (Fig. 3D). Proteins YggX and Dps are known for their protective role under iron imbalance conditions in *E. coli* (Almiron et al., 1992; Pomposiello et al., 2003), so their downregulation in the cytoplasm is clearly related to the stress response. We have also observed the downregulation of several proteins important for molybdenum uptake. This unbalance in molybdenum homeostasis may be due to the dramatic decrease in DmsA, which is the largest subunit that binds molybdenum at its active site in *E. coli* (Trieber et al., 1996).

3.5. Δtat mutation provokes synthesis of colanic acid capsular polysaccharide

It has previously been observed that the Δ*tatC* strain responds to envelope stress by upregulation of genes involved in the cell envelope synthesis. Expression of the *cps* operon was shown to be induced (more than 2-fold difference compared to parental strain) only under anaerobic conditions (Ize et al., 2004). The results of our study not only showed the induction of *cps* operon proteins in Δ*tatABCDE* under aerobic conditions, but also the exclusive presence of nearly all the *cps* proteins in the *tat* mutant when compared with the parental strain. In short, we observed an induction of most of the *cps* genes responsible for the production of colanic acid, an exopolysaccharide necessary for the formation of the three-dimensional structure of *E. coli* biofilms (Danese et al., 2000). Fourteen out of 20 proteins involved in this cluster are present in our dataset, notably 13 of them only in the Δ*tat* strain (Fig. 4A-B). Of all the proteins in the *cps* cluster only one of them has a Tat signal peptide: WcaM (Stevenson et al., 1996). However, WcaM was not present in the obtained proteomic data, and neither were Wzx or WcaB/D, which were assumed to be also upregulated. These proteins are inner transmembrane proteins, which makes them highly difficult to solubilize and detect using standard proteomics. In conclusion, the identification of proteins up- and downstream from WcaM, Wzx, WcaB/D may suggest an induction of the entire *csp* operon. The fact that most of the proteins were identified only in the Δ*tat* strain, but not in the WT, strongly suggests the dependence of colanic acid formation on the Tat secretion deficiency in the cell. Moreover, the upregulation of the colanic acid synthesis in Δ*tat* was also mirrored by an

increased amount of the exopolysaccharide produced by the mutant compared to WT (Fig. 4C).

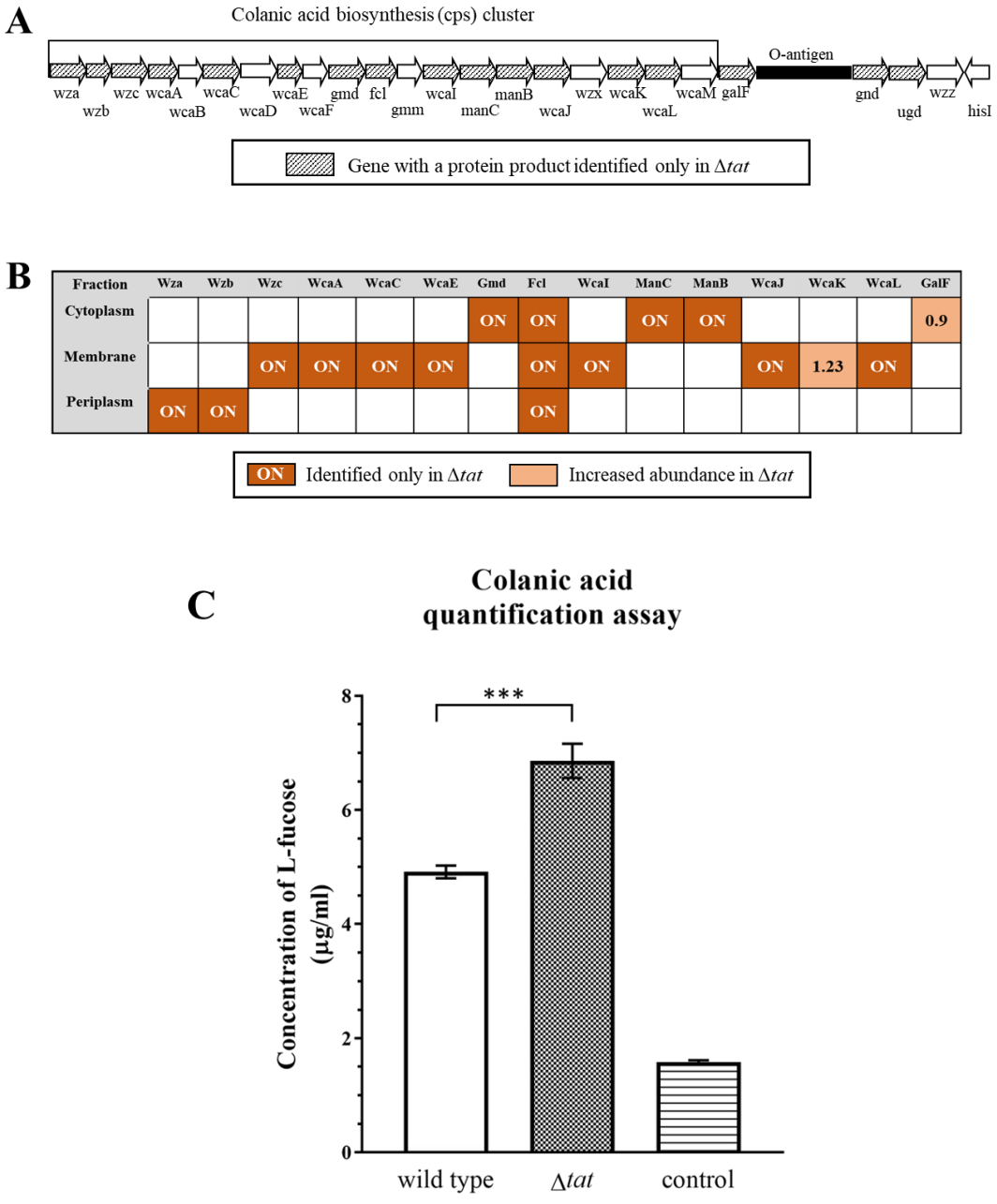


Figure 4. Induction of colanic acid synthesis in *E. coli* Δ*tat*. Genetic organization of *E. coli* colanic acid synthesis (*cps*) cluster (A). Genes which protein products were synthesized upon *tat* mutation were marked. Table representing induced proteins of colanic acid synthesis and their cellular localization (B). Concentrations of colanic acid in *E. coli* WT and Δ*tat* grown in LB liquid medium were compared (C). To quantify

colanic acid, concentration of its specific component, L-fucose, was measured. LB liquid medium was used as a control. Bars represent geometric means ± standard deviations from three parallel samples. A Student's t-test was applied to determine statistical significance, (***) P<0.001.

3.6. Colanic acid synthesis promotes biofilm formation in E. coli Δtat

In the previous section (3.5) we showed that the deletion of *tat* genes leads to induction of a set of proteins involved in the synthesis of colanic acid, which are not present in the wild type. Based on this observation and the fact that this exopolysaccharide is required for development of biofilm in *E. coli*, we decided to investigate biofilm formation of the *tat* mutant in comparison to wild type cells. A biofilm forms when bacteria adhere to surfaces in aqueous environments. Ultimately, a biofilm is formed to protect bacteria from stressful environmental factors, although it is known that certain cellular changes such as DNA damage, oxidative stress (Gier et al., 2008) and iron depletion (Garcia et al., 2015) also promote biofilm development. In fact, it is often believed that any stress may trigger biofilm formation as a means of protection. To comparatively quantify biofilm formation of mutant and WT, a biofilm assay based on the ability of bacteria to adhere to plastic surface of microtiter plates was performed. To this end, cells were grown in standard LB medium to mirror the conditions of the proteomic experiment. Biofilm formation was also assessed for cells grown in minimal medium, since it has been reported that a strong biofilm formation in *E. coli* can be observed mostly in minimal media (Naves et al., 2008). Biofilm formation was defined as the intensity of crystal violet stained cells divided by optical density of the upper culture. The obtained results showed enhanced biofilm formation ability of *E. coli* Δ*tat*, especially after 48 hours in both media (Fig. 5). These results may seem surprising, especially taking into account the previously reported decrease in biofilm formation of a Δ*tatC* mutant (Ize et al., 2004). Nevertheless, it has to be noted that biofilm formation is a highly complex phenomenon that is initiated via numerous pathways depending on the strain, medium and environment. Apart from the colanic acid, the presence of IBs in the *tat* mutant may account for the observed biofilm production. It has been shown that IBs formation is associated with an abnormal phenotype involving cell clustering and the development of biofilm-like characteristics (Lee at al., 2008). In conclusion, we assume that the higher biofilm-forming capacity of *E. coli* Δ*tat* is acquired as a response to stress generated by the loss of Tat-secretion and

can be attributed to the upregulation of colanic acid cluster proteins in the mutant, not observed in the wild type.

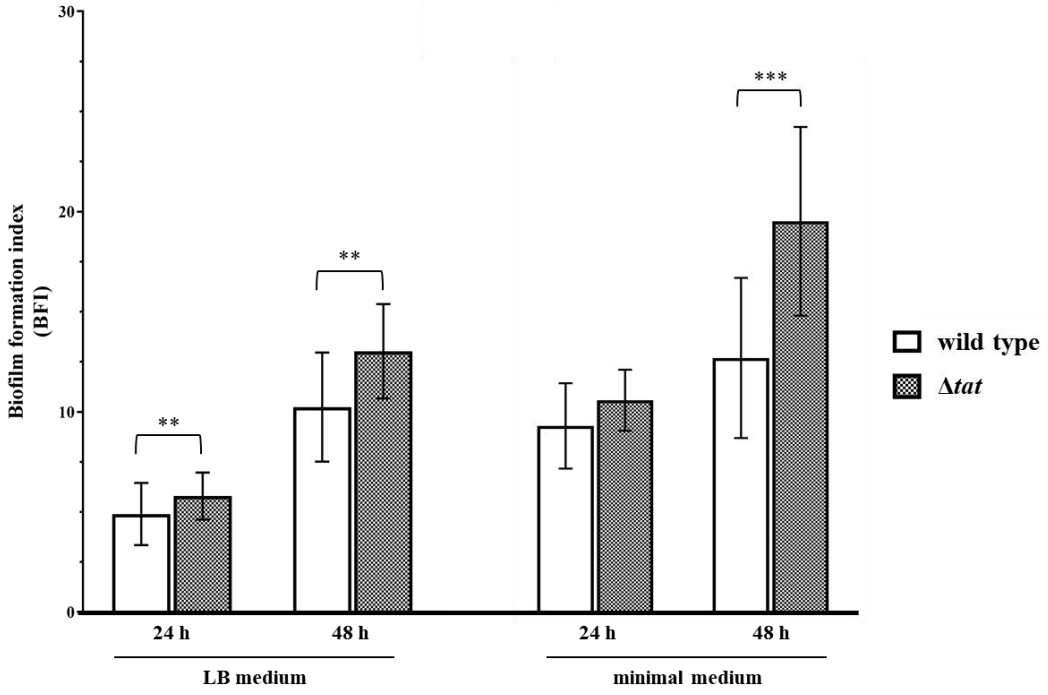


Figure 5. Biofilm formation of *E. coli* WT and Δ*tat*. The ability to form a biofilm was assessed using a polystyrene microtiter plate biofilm assay described under "Materials and methods". Cultures were grown anaerobically in two media: LB and Belitsky minimal medium. The biofilm absorbance was measured after 24 and 48 hours. Error bars indicate standard deviations calculated from 20 replicates.

4. Conclusions

We have presented a global view of cellular consequences of a *tatABCDE* deletion, that to the best of our knowledge, is so far the most comprehensive analysis of the proteome changes due to the Tat apparatus deficiency in *E. coli*. Apart from the already known filamentous chain-like phenotype of *E. coli* Δ*tat*, we identified the production of OMVs and increased inclusion body formation as novel phenotypes of *tat* mutants. Our proteomic analysis has revealed the upregulation of proteins involved in protein folding, degradation and response to heat, oxidation, osmolarity, and cold. In addition, the impairment of *E. coli* outer membrane resulted in the induction of proteins responsible for cell wall biogenesis. The *tat* deletion negatively affects the synthesis of iron and molybdenum transporters and imbalances their homeostasis in the cell. Finally, we have

shown that a simultaneous deletion of *tatABCDE* genes leads to the activation of proteins responsible for colanic acid synthesis, which contributes to the biofilm formation in *tat* mutant.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642836.

Authors contributions

K.D. and I.G. designed and performed the experiments, analyzed the data and wrote the manuscript. W.M and C.R. devised the project. R.S. prepared specimens for microscopy and provided micrographs. T.S. and C.W. performed mass spectrometric measurements. C.R, K.R. and S.S were involved in planning and supervision of the work. All authors provided critical feedback and helped shape the research, analysis and manuscript.

References

Ajouz, B., Berrier, C., Garrigues, A., Besnard, M., Ghazi, A., 1998. Release of Thioredoxin via the Mechanosensitive Channel MscL during Osmotic Downshock of *Escherichia coli* Cells. J. Biol. Chem. 273(41), 26670-4.

Alami, M., Trescher, D., Wu, L.F., Müller, M., 2002. Separate analysis of twin-arginine translocation (Tat)-specific membrane binding and translocation in *Escherichia coli*. J. Biol. Chem. 277(23), 20499-503.

Alami, M., Luke I., Deitermann, S., Eisner, G., Koch, H.G., Brunner, J., Müller, M., 2003. Differential interactions between a twin-arginine signal peptide and its translocase in *Escherichia coli*. Mol. Cell. 12, 937-946.

Aldea, M., Hernandez-Chico, H., De la Campa, A.G., Kusnher, S.R., Vicente, M., 1988. Identifcation, cloning, and expression of bolA, an ftsZ-dependent morphogene of *E. coli*. J. Bacteriol. 170, 5169-5176.

Almiron, M., Link, A.J., Furlong, D., Kolter, R., 1992. A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. Genes. Dev. 6, 2646-2654.

Baglieri, J., Beck, D., Vasisht, N., Smith, C.J., Robinson, C., 2018. Structure of TatA paralog, TatE, suggests a structurally homogeneous form of Tat protein translocase that transports folded proteins of differing diameter. J. Biol. Chem. 287, 7335-7344.

Ball, G., Antelmann, H., Imbert, P.R.C., Gimenez, M.R., Voulhoux, R., Ize, B., 2016. Contribution of the Twin Arginine Translocation system to the exoproteome of *Pseudomonas aeruginosa*. Scientific Reports. 6, 27675

Behrendt, J., Standar, K., Lindenstrauß, U., Bruser, T., 2004. Topological studies on the twin-arginine translocase component TatC. FEMS Microbiol. Lett. 15, 303-8.

Bendtsen, J.D., Nielsen, H., Widdick, D., Palmer, T., Brunak. S., 2005. Prediction of twin-arginine signal peptides. BMC Bioinformatics. 6, 167.

Berks, B.C, Palmer, T., Sargent, F., 2005. Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. Curr. Opin. Microbiol. 8, 174-181.

Berks, B.C., 2015. The twin-arginine protein translocation pathway. Annu. Rev. Biochem. 84, 843-64.

Bernhardt, T.G., de Boer, P.A., 2003. The *Escherichia coli* amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. Mol. Microbiol. 48(5), 1171–1182.

Blaudeck, N., Sprenger, G.A., Freudl, R., Wiegert, T., 2001. Specificity of signal peptide recognition in Tat-dependent bacterial protein translocation. J. Bacteriol. 183(2), 604-610.

Blaudeck, N., Kreutzenbeck, P., Freudl, R., Sprenger, G.A., 2003. Genetic analysis of pathway specificity during posttranslational protein translocation across the *Escherichia coli* plasma membrane. J. Bacteriol. 185(9), 2811-2819.

Bogsch, E.G., Sargent, F., Stanley, N.R., Berks, B.C., Robinson, C., Palmer, T., 1998. An essential component of a novel bacterial protein export system with homologues in plastids and mitochondria. J. Biol. Chem. 273, 18003–18006.

Brown, R.N., Romine, M.F., Schepmoes, A.A., Smith, R.D., Lipton, M.S., 2010. Mapping the subcellular proteome of *Shewanella oneidensis* MR-1 using sarkosyl-based fractionation and LC-MS/MS protein identification. J. Proteome Res. 9, 4454-63.

Browning, D.F., Richards, K.L., Peswani, A.R., Roobol, J., Busby, S.J.W., Robinson, C., 2017. *Escherichia coli* 'TatExpress' strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway. Biotech. Bioeng. 114, 2828-2836.

Bukau, B., Walker, G.C., 1989. Cellular defects caused by deletion of the *Escherichia coli dnaK* gene indicate roles for heat shock protein in normal metabolism. J. Bacteriol. 171, 2337-2346.

Carrio, M.M., Villaverde, A., 2002. Construction and deconstruction of bacterial inclusion bodies. J. Biotechnol. 96, 3-12.

Carrió, M.M., Villaverde, A., 2005. Localization of Chaperones DnaK and GroEL in Bacterial Inclusion Bodies. J. Bacteriol. 187(10), 3599–3601.

Collinson, I., Corey, R.E., Allen, W.J., 2015. Channel crossing: how are proteins shipped across the bacterial plasma membrane? Philos. Trans. R. Soc. Lond., B, Biol. Sci. 370, 1679.

Constantinidou, C., Hobman, J.L., Griffiths, L., Patel, M.D., Penn, C.W., Cole, J.A., Overton, T.W., 2005. A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as *Escherichia coli* K12 adapts from aerobic to anaerobic growth. J. Biol. Chem., 281(8), 4802–4815.

Conti, J., Viola, M.G., Camberg, J.L., 2015. The bacterial cell division regulators MinD and MinC form polymers in the presence of nucleotide. FEBS Letters. 589(2), 201-206.

Cox, J., Mann, M., 2006. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein 98 quantification. Nature biotechnology 26(12), 1367-1372.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A, Olsen, J.V., Mann, M., 2011. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. 10(4), 1794-805.

Danese, P.N., Pratt, L.A., Kolter, R., 2000. Exopolysaccharide production is required for development of *Escherichia coli* K-12 biofilm architecture. J. Bacteriol. 182(12), 3593-6.

Davies, B.W., Kohanski, M.A., Simmons, L.A., Winkler, J.A., Collins, J.J., Walker, G.C., 2009. Hydroxyurea Induces Hydroxyl Radical-Mediated Cell Death in *Escherichia coli* Molecular Cell, 36(5), 845–860.

DeLisa, M.P., Tullman, D., Georgiou, G., 2003. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. Proc. Natl. Acad. Sci. USA. 100, 6115–6120.

DeLisa, M.P., Lee, P., Palmer, T., Georgiou, G., 2004. Phage shock protein PspA of *Escherichia coli* relieves saturation of protein export via the Tat pathway. J. Bacteriol. 186, 366-373.

Dilks, K., Rose, R.W., Hartmann, E., Pohlschröder, M., 2003. Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey. J. Bacteriol. 185(4), 1478-83.

Economou, A., 2005. Sec, drugs and rock'n'roll: antibiotic targeting of bacterial protein translocation. Emerging Therapeutic Targets, 5(2), 141-153.

Eraso, J.M., Markillie, L.M., Mitchell, H.D., Taylor, R.C., Orr, G., Margolin, W., 2014. The highly conserved MraZ protein is a transcriptional regulator in *Escherichia coli*. J. Bacteriol. 196, 2053-2066.

Fayet, O., Ziegelhoffer, T., Georgopoulos, C., 1998. The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. J. Bacteriol. 171(3), 1379-85.

Fröbel, J., Rose, P., Müller, M., 2012. Twin-arginine-dependent translocation of folded proteins. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 367, 1029-1046.

García, C.A., Alcaraz, E.S., Franco, M.A., Passerini de Rossi, B.N., 2015. Iron is a signal for *Stenotrophomonas maltophilia* biofilm formation, oxidative stress response, OMPs expression, and virulence. Front. Microbiol. 6, 926.

Geier, H., Mostowy, S., Cangelosi, G.A., Behr, M.A., Ford, T.E., 2008. Autoinducer-2 Triggers the Oxidative Stress Response in *Mycobacterium avium*, Leading to Biofilm Formation, Appl Environ Microbiol. 74(6), 1798–1804.

Gerdes, S.Y., Scholle, M.D., Campbell, J.W., Balázsi, G., Ravasz, E., Daugherty, M.D., Osterman, A.L., 2003. Experimental Determination and System Level Analysis of Essential Genes in *Escherichia coli* MG1655. J. Bacteriol. 185(19), 5673–5684.

Gottesman, S., Stout, V., 1991. Regulation of capsular polysaccharide synthesis in *Escherichia coli* K12. Mol. Microbiol. 5(7), 1599-606.

Hatzixanthis, K., Palmer, T., Sargent, F., 2003. A subset of bacterial inner membrane proteins integrated by the twin-arginine translocase. Mol. Microbiol.,49, 1377-1390.

Hoffmann, J.H., Linke, K., Graf, J.H., Lilie, H., Jakob, U., 2004. Identification of a redox-regulated chaperone network. EMBO J. 23(1), 160-8.

Hussain, H., Grove, J., Griffiths, L., Busby, S., Cole, J., 1994. A seven-gene operon essential for formate-dependent nitrite reduction to ammonia by enteric bacteria. Mol. Microbiol. 12, 153-163.

Ize, B., Gerard, F., Zhang, M., Chanal, A., Voulhoux, R., Palmer, T., Filloux, A., Wu, L.F., 2002. In vivo dissection of the Tat translocation pathway in *Escherichia coli*. J. Mol. Biol. 317, 327–35.

Ize, B., Stanley, N.R., Buchanan, G., Palmer, T., 2003. Role of the *Escherichia coli* Tat pathway in outer membrane integrity. Mol. Microbiol. 48, 1183–1193.

Ize, B., Porcelli, I., Lucchini, S., Hinton, J.C., Berks, B.C., Palmer, T., 2004. Novel phenotypes of *Escherichia coli* tat mutants revealed by global gene expression and phenotypic analysis. J. Biol. Chem. 279(46), 47543–54.

Joliff, G., Beguin, P., Juy, M., Millet, J., Ryter, A., Poljak, R., Anbert, J.P., 1986. Isolation, crystallization and properties of a new cellulase of *Clostridium thermocellum* overproduced in *Escherichia coli*. Bio. Technol. 4, 896-900.

Jones, S.E., Lloyd, L.J., Tan, K.K., Buck, M., 2003. Secretion defects that activate the phage shock response of *Escherichia coli*. J. Bacteriol. 185, 6707-6711.

Koch, S., Fritsch, M.J., Buchanan, G., Palmer, T., 2012. *Escherichia coli* TatA and TatB Proteins Have N-out, C-in Topology in Intact Cells. J Biol Chem. 287(18), 14420–14431.

Krehenbrink, M., Edwards, A., Downie, J.A., 2011. The superoxide dismutase SodA is targeted to the periplasm in a SecA-dependent manner by a novel mechanism. Mol. Microbiol. 82(1),164-79.

Leake, M.C., Greene, N.P., Godun, R.M., Granjon, T., Buchanan, T., Chen, S., Berry, R.M., Palmer, T., Berks, B.C., 2008. Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by in vivo single- molecule imaging. PNAS. 105(40), 15376-15381.

Lee, K.K., Jang, C.S., Yoon, J.Y., Kim, S.Y., Kim, T.H., Ryu, K.H., Kim, W., 2008. Abnormal cell division caused by inclusion bodies in *E. coli*; increased resistance against external stress. Microbiol. Res. 163(4), 394-402.

Lee, P.A., Buchanan, G., Stanley, N.R., Berks, B.C., Palmer, T., 2002. Truncation Analysis of TatA and TatB Defines the Minimal Functional Units Required for Protein Translocation. J Bacteriol. 184, 5871-5879.

Lee, P.A., Tullman-Ercek, D., Georgiou, G., 2006. The Bacterial Twin-Arginine Translocation Pathway. Annu. Rev. Microbiol. 60, 373–395.

Méjean, V., Lobbi-Nivol, C., Lepelletier, M., Giordano, G., Chippaux, M., Pascal, M., 1994. TMAO anaerobic respiration in *Escherichia coli*: involvement of the tor operon. Mol. Microbiol., 11, 1169-1179.

Müller, M., Klösgen R.B., 2005. The Tat pathway in bacteria and chloroplasts (review). Mol. Membr. Biol. 22, 113-21.

Muntel, J., Hecker, M., Becher, D., 2012. An exclusion list based label-free proteome quantification approach using an LTQ Orbitrap. Rapid Commun. Mass Spectrom. 26, 701–709.

Naves, P., del Prado, G., Huelves, L., Gracia, M., Ruiz, V., Blanco, J., Rodríguez-Cerrato, V., Ponte, M.C., Soriano, F., 2008. Measurement of biofilm formation by clinical isolates of *Escherichia coli* is method-dependent. J. Appl. Microbiol. 105(2), 585-90.

Neumann, M., Mittelstädt, G., Iobbi-Nivol, C., Saggu, M., Lendzian, F., Hildebrandt, P., Leimkühler, S., 2009. A periplasmic aldehyde oxidoreductase represents the first molybdopterin cytosine dinucleotide cofactor containing molybdo-flavoenzyme from *Escherichia coli*. FEBS Journal, 276 (10), 2762–2774.

Obadia, B., Lacour, S., Doublet, P., Baubichon-Cortay, H., Cozzone, A.J., Grangeasse, C., 2007. Influence of tyrosine-kinase Wzc activity on colanic acid production in *Escherichia coli* K12 cells. J. Mol. Biol. 367, 42–53.

Ochsner, U.A., Snyder, A., Vasil, A.I., Vasil, M.L., 2002. Effects of the Twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis. Pro. Nat. Ac. Sci. 99(12), 8312-8317.

Orfanoudaki, G., Economou, A., 2014. Proteome-wide subcellular topologies of *E. coli* polypeptides database (STEPdb). Mol. Cell Proteomics. 13(12), 3674-87.

O'Toole, G.A., 2011. Microtiter Dish Biofilm Formation Assay. J Vis Exp. 47, 2437.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R., 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42, 206-214.

Palmer, T., Sargent, F., Berks, B.C., 2005. Export of complex cofactor-containing proteins by the bacterial Tat pathway. Trends Microbiol. 13(4), 175–180.

Patel, R., Smith, S.M., Robinson, C., 2014. Protein Transport by the bacterial Tat pathway, Biochim. Biosphys. Acta – Mol. Cell. Res. 1843, 1620-1626.

Petriman, N.A., Jauß, B., Hufnagel, A., Franz, L., Sachelaru, I., Drepper, F., Warscheid, B., Koch, H.G., 2018. The interaction network of the YidC insertase with the SecYEG translocon, SRP and the SRP receptor FtsY. Sci. Rep. 8 (578).

Pierce, J.J., Turner, C., Keshavarz-Moore, E., Dunnill, P., 1997. Factors determining more efficient large-scale release of a periplasmic enzyme from *E. coli* using lysozyme. J Biotechnol. 58(1), 1-11.

Pomposiello, P.J., Koutsolioutsou, A., Carrasco, D., Demple, B., 2003. SoxRS-regulated expression and genetic analysis of the yggX gene of *Escherichia coli*. J. Bacteriol. 185, 6624-6632.

Pradel, N., Ye, C., Livrelli, V., Xu, J., Joly, B., Wu, L., 2003. Contribution of the twin arginine translocation system to the virulence of enterohemorrhagic *Escherichia coli* O157:H7. Infect. Immun. 71, 4908–4916.

Rabilloud, T., 2012. Limits of Proteomics: Protein Solubilisation Issues. eLS.

Randall, L.L., Hardy, J.S., 1986. Correlation of competence for export with lack of tertiary structure of the mature species: A study in vivo of maltose-binding protein in *E. coli*. Cell. 12, 921-8.

Rappsilber, J., Mann, J., Ishihama, Y., 2007. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc. 2, 1896–1906.

Robinson, C., Bolhuis, A., 2001. Protein targeting by the twin-arginine translocation pathway. Nat. Rev. Mol. Cell Biol. 2, 350-356.

Rodrigue, A., Chanal, A., Beck, K., Muller, M., Wu, L.F., 1999. Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial Tat pathway. J. Biol. Chem. 274(19), 13223–28.

Sargent, F., Bogsch, E.G., Stanley, N.R., Wexler, M., Robinson, C., Berks, B.C., Palmer, T., 1998. Overlapping functions of components of a bacterial Sec-independent protein export pathway. EMBO J. 17, 3640–3650.

Sawers, R.G., Boxer, D.H., 1986. Purification and properties of membrane-bound hydrogenase isoenzyme 1 from anaerobically grown *Escherichia coli* K12. European J. Biochem. 156, 265-275.

Schröder, H., Langer, T., Hartl, F.U., Bukau, B., 1993. DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. EMBO J. 12(11), 4137-44.

Stanley, N.R., Palmer, T., Berks, B.C., 2000. The Twin arginine consensus motif of Tat signal peptides is involved in Sec- independent protein targeting in *Escherichia coli*. J. Biol. Chem. 275, 11591-11596.

Stanley, N.R., Findlay, K., Berks, B.C., Palmer, T., 2001. *Escherichia coli* Strains Blocked in Tat-Dependent Protein Export Exhibit Pleiotropic Defects in the Cell Envelope. J. Bacteriol. 183, 139-144.

Stanley, N.R., Sargent, F., Buchanan, G., Shi, J., Stewart, V., Palmer, T., Berks, B.C., 2002. Behaviour of topological marker proteins targeted to the Tat protein transport pathway. Mol. Microbiol. 43, 1005-1021.

Stevenson, G., Andrianopoulos, K., Hobbs, M., Reeves, P.R., 1996. Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. J. Bacteriol. 178(16), 4885-4893.

Stülke, J., Hanschke, R., Hecker, M., 1993. Temporal activation of beta-glucanase synthesis in *Bacillus subtilis* is mediated by the GTP pool. J. Gen. Microbiol. 139, 2041-2045.

Summer, E.J., Mori, H., Settles, A.M., Cline, K., 2000. The thylakoid ΔpH-dependent pathway machinery facilitates RR-independent N-tail protein integration. J. Biol. Chem. 275, 23483–23490.

Sutherland, G.A., Grayson K.J., Adams, N.B.P., Mermans, D.M.J., Jones, A.S., Robertson, A.J., Auman, D.B., Brindley, A.A., Sterpone, F., Tufferye, P., Derreumaux, P., Dutton, P.L., Robinson, C., Hitchcock, A., Hunter, C.N., 2018. Probing the quality control mechanism of the *Escherichia coli* twin-arginine

translocase with folding variants of a de novo-designed heme protein. J. Biol. Chem.

Takahashi, Y., Nakamura, M., 1999. Functional assignment of the ORF2-iscS-iscU-iscA-hscB-hscA-fdx-ORF3 gene cluster involved in the assembly of Fe-S clusters in *Escherichia coli*. J. Biochem. 126(5), 917-26.

Thomas, J.G., Baneyx, F., 2000. ClpB and HtpG facilitate de novo protein folding in stressed *Escherichia coli* cells. Mol Microbiol. 36(6), 1360-70.

Tinker, J.K., Erbe, J.L., Holmes, R.K., 2005. Characterization of fluorescent chimeras of cholera toxin and *Escherichia coli* heat-labile enterotoxins produced by use of the twin arginine translocation system. Infect. Immun. 73(6), 3627-35.

Trieber, C.A., Rothery, R.A., Weiner, J.H., 1996. Consequences of Removal of a Molybdenum Ligand (DmsA-Ser-176) of *Escherichia coli* Dimethyl Sulfoxide Reductase. J. Biol. Chem. 271(44), 27339-45.

Tullman-Ercek, D., DeLisa, D., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., Georgiou, G., 2007. Export Pathway Selectivity of *Escherichia coli* Twin Arginine Translocation Signal Peptides. J. Biol. Chem. 282, 8309-8316.

Walker, K.L., Jones, A.S., Robinson, C., 2015. The Tat pathway as a biotechnological tool for the expression and export of heterologous proteins in *Escherichia coli*. Pharm Bioprocess. 3(6), 387-396.

Wexler, M., Sargent, F., Jack, R.L., Stanley, N.R., Bogsch, E.G., Robinson, C., Berks, B.C., Palmer, T., 2000. TatD is a cytoplasmic protein with DNase activity. No requirement for TatD family proteins in sec-independent protein export. J. Biol. Chem. 275(22), 16717-22.

Wolff, S., Hahne, H., Hecker, M., Becher, D., 2008. Complementary analysis of the vegetative membrane proteome of the human pathogen *Staphylococcus aureus*. Mol. Cell Proteomics. 7(8), 1460-8.

Yaagoubi, A., Kohiyama, M., Richarme, G., 1994. Localization of DnaK (chaperone 70) from *Escherichia coli* in an osmotic-shock-sensitive compartment of the cytoplasm. J. Bacteriol. 176(22), 7074–7078.

Zühlke, D., Dörries, K., Bernhardt, J., Maaß, S., Muntel, J., Liebscher, V., Pané-Farré, J., Riedel, K., Lalk, M., Völker, U., Engelmann, S., Becher, D., Fuchs, S., Hecker, M., 2016. Costs of life - Dynamics of the protein inventory of *Staphylococcus aureus* during anaerobiosis. Scientific Reports. 6, 28172.

2.2 Equal first author publication-Comparative proteome analysis in an *Escherichia coli* CyDisCo strain identifies stress responses related to protein production oxidative stress and accumulation of misfolded protein.

Isabel Guerrero Montero, Katarzyna Magdalena Dolata, Rabea Schlüter, Gilles Malherbe, Susanne Sievers, Daniela Zühlke, Thomas Sura, Emma Dave, Katharina Riedel, Colin Robinson, Far-reaching cellular consequences of tat deletion in Escherichia coli revealed by comprehensive proteome analyses. Microbiological Research, 218, 97–107.

2.2.1 Contribution

For this project, I heavily contributed in concept and experimental design, performed the experiments, analysed the proteomic data, reviewed the literature available on proteins that showed important shifts in their regulation, made the figures and wrote the manuscript in collaboration with Katarzyna Dolata. The manuscript was revised by all the authors, especially Colin Robinson, Katharina Riedel and Susanne Sievers.

For this study, we looked into the effects a misfolded protein not exportable by Tat would have on the *E. coli* cells. To do so I constructed a vector via molecular cloning that only contained the CyDisCo components to compare it with a vector containing both CyDisCo and a TorA-scFv and a second vector with CyDisCo and a TorA-scFv-26tail. The TorA-scFv-26tail is derived from the TorA-scFv of the first vector but has an additional 26 random amino acids added to the C-terminus. Both vectors have been used before in (Jones et al. 2016), which had shown that the TorA-scFv-26tail could not be exported to the periplasm due to its lack of correct folding. I transformed the three vectors and tested for expression of the scFv and misfolded scFv by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting (Figure 3), and

performed comparative densitometry on the band-intensities before taking samples 3h post induction for proteomic analysis. The samples were sent to the University of Greifswald and processed between Katarzyna Dolata and myself before mass spec analysis.

I also made growth curves with the three transformants by measuring the OD_{600} every 30 minutes of 3 replicates. Samples were taken of induced and un-induced culture and an additional control with no plasmid was also added to test the metabolic burden being suffered by the cells producing the protein. I graphed the results (Figure 1) and I calculated the specific growth rate (Table 2).

The proteomic results were analyzed by both myself and Katarzyna Dolata. A Venn diagram (Figure 4a) comparing the proteins found in the three samples and Voronoi maps (Figure 4b-d) showing the categorization of these proteins according to TIGRfam and manual curation. These figures were extremely useful to help pinpoint which routes were being affected by folded and misfolded protein production.

The proteomic data suggested changes in the concentration of tat proteins available under different conditions. To further study this phenomenon, I prepared samples for RT-qPCR analysis pre-induction and at various points post induction. The samples were sent to UCB, where Gilles Malherbe had designed the primers, and the RT-qPCR experiment and the analysis of the results was done by Gilles Malherbe and myself. The relative transcription levels resulting from this experiment were graphed in Figure 5. The proteomic data also suggested that the misfolded scFv induced the expression of Ag43, an outer membrane protein that confers bacterial cell to cell aggregation (Woude &

Henderson 1980), so a sedimentation assay accompanied by phase-contrast microscopy was also performed to determine if aggregation was happening in that transformant.

Finally, electron micrographs were taken by Rabea Schlüter and interpreted by myself and Katarzyna Dolata.

2.2.2 Publication

The paper as was sent for publication is as follows:

Comparative proteome analysis in an *Escherichia coli* CyDisCo strain identifies stress responses related to protein production, oxidative stress and accumulation of misfolded protein

Isabel Guerrero Montero[1]*, Katarzyna Magdalena Dolata[2]*, Rabea Schlüter[3], Gilles Malherbe[1,4], Susanne Sievers[2], Daniela Zühlke[2], Thomas Sura[2], Emma Dave[4], Katharina Riedel[2], Colin Robinson[1]

[1] *School of Biosciences, University of Kent, Canterbury CT2 7NJ, U.K.*

[2] *Institute of Microbiology, University of Greifswald, Felix-Hausdorff-Straße 8, 17487 Greifswald, Germany*

[3] *Imaging Center of the Department of Biology, University of Greifswald, Friedrich-Ludwig-Jahn-Str. 15, 17487 Greifswald, Germany*

[4] *UCB Celltech, 216 Bath road, Slough, SL1 3WE U.K.*

[*] Isabel Guerrero Montero and Katarzyna Magdalena Dolata contributed equally to this work

Corresponding author:

Colin Robinson, c.robinson-504@kent.ac.uk, tel. +44 1227 823443

2.2 Comparative proteome analysis in an Escherichia coli CyDisCo strain identifies stress responses related to protein production oxidative stress and accumulation of misfolded protein

Abstract

Background

The Twin-arginine translocation (Tat) pathway of *Escherichia coli* has great potential for the export of biopharmaceuticals to the periplasm due to its ability to transport folded proteins, and its proofreading mechanism that allows correctly folded proteins to translocate. Coupling the Tat-dependent protein secretion with the formation of disulfide bonds in the cytoplasm of *E. coli* CyDisCo provides a powerful platform for the production of industrially challenging proteins. In this study, we investigated the effects on the *E. coli* cells of exporting a folded substrate (scFv) to the periplasm using a Tat signal peptide, and the effects of expressing an export-incompetent misfolded variant.

Results

Cell growth is decreased when either the correctly folded or misfolded scFv is expressed with a Tat signal peptide. However, only the production of misfolded scFv leads to cell aggregation and formation of inclusion bodies. The comprehensive proteomic analysis revealed that both conditions, recombinant protein overexpression and misfolded protein accumulation, lead to downregulation of membrane transporters responsible for protein folding and insertion into the membrane while upregulating the production of chaperones and proteases involved in removing aggregates. These conditions also differentially affect the production of transcription factors and proteins involved in DNA replication. The most distinct stress response observed was the cell aggregation caused by elevated levels of antigen 43. Finally, Tat-dependent secretion causes an increase in *tatA* expression only after induction of protein expression, while the subsequent post-induction analysis revealed lower *tatA* and *tatB* expression levels, which correlate with lowered TatA and TatB protein abundance.

Conclusions

The study identified characteristic changes occurring as a result of the production of both a folded and a misfolded protein, but also highlights an exclusive unfolded stress response. Countering and compensating for these changes may result in higher yields of pharmaceutically relevant proteins exported to the periplasm.

Keywords

Escherichia coli CyDisCo, Twin-arginine transport, recombinant protein, disulfide bond formation, misfolding, proteome, stress response

2.2 Comparative proteome analysis in an Escherichia coli CyDisCo strain identifies stress responses related to protein production oxidative stress and accumulation of misfolded protein

Background

Escherichia coli is a popular expression platform for the production of recombinant proteins of therapeutic interest due to the ease and relative low cost with which it can be rapidly grown and modified in controlled laboratory and industrial settings. Among its advantages over other cell factories we highlight the simplification of downstream processing if the protein of interest is exported to the periplasmic space. Since *E. coli* does not naturally export many proteins into the periplasm the recovery of the protein of interest will occur with minimal contamination of unwanted host cytosolic proteins, DNA or endotoxins [1].

The export of the recombinant protein to the periplasmic space most typically occurs via the Sec transport pathway 'naturally' and in most industrial applications. In this pathway, an unfolded substrate with a cleavable N-terminal signal peptide is recognized and transported through the membrane-bound Sec translocase to the periplasm where it is then folded to a functional state. This is also the primary means of producing disulfide-bonded proteins due to the oxidizing environment of the periplasm [2, 3]. One of the major disadvantages of this system is the inability of the Sec pathway to successfully export highly complex proteins that rapidly fold in the cytoplasm before they reach the translocase or that have post-translational modifications such as cofactor insertions, which is the case of many biotechnologically relevant proteins. To circumvent this problem another transport pathway that transports fully folded proteins exists: the Tat system. This system, named for the twin arginine motif highly conserved in the signal peptides that localize the proteins to the transporter [4], has been shown not only to transport folded proteins from the cytoplasm to the periplasm [5, 6], but also to have a useful although poorly understood proofreading mechanism that blocks the export of proteins unless they are tightly folded and have the appropriate co-factors in place. By exploiting this system, heterologous proteins harvested from the periplasm will potentially contain a very low degree of protein heterogeneity, making it easier and more cost effective for downstream processing after harvest. Moreover, this export system is naturally capable of transporting proteins up to 150 kDa [7] which potentially allows the transport of relatively complex proteins [8]. The Tat system has also an important role in maintaining cellular homeostasis in *E. coli* [9].

Nevertheless, many proteins of biopharmaceutical interest are too complex to fully fold in the cytoplasm because disulfide bonds cannot properly form in the non-oxidizing

cytoplasm of *E. coli*. There are several routes available to avoid this problem, either by expressing proteins that fold tightly without the need of disulfide bonds [10] or by using strains with an oxidizing cytoplasm [11, 12]. One of these strains is the CyDisCo (Cytoplasmic Disulfide bond formation in *E. coli*) strain which expresses a yeast mitochondrial thiol oxidase, Erv1p and the human protein disulfide isomerase, PDI both lacking secretion signals [13]. The expression of these two proteins allows the formation of disulfide bonds in the cytoplasm, which leads to the correct folding of the protein and thus to its export to the periplasm via the Tat pathway [14].

In this study we sought to understand the physiological changes that result from the expression of a recombinant protein in both its correctly folded state and in a deliberately misfolded state using an *E. coli* CyDisCo platform. The expression of heterologous proteins in *E. coli* often results in protein misfolding and aggregation because the cell is not able to sustain non-physiological rates of expression. For its industrial application, the production process must be optimized to achieve high protein yields. Thus, under protein production process conditions it is important to identify stress parameters to circumvent and improve the process efficiency. For this reason, we were interested in characterizing the changes that occur during the expression of a misfolded Tat substrate, in order to understand whether additional responses occur. The substrate used was an scFv antibody construct that is efficiently exported by Tat while the 'misfolded' variant contains a 26-residue C-terminal extension that blocks export by Tat [14]. Both proteins were synthesised with an N-terminal TorA signal peptide that directs export by Tat. For our study, we used a comprehensive proteomic approach to monitor the changes in cell physiology and proteome caused by recombinant protein secretion, misfolding and aggregation.

Methods

Bacterial strains and plasmids

The *E. coli* strain utilized for this study was W3110. All plasmids (Table 1) are derived from pET23 and have a Ptac promoter before the signal peptide or CyDisCo components and a 6xHis-tag after the scFv and mutated scFv. pIGM01 was constructed by PCR amplification using pHAK13 as a template and primers Empty_CyD_F (5'-GAAGGAGATATTATGAAAGCCATCG-3') and Empty_CyD_R (5'-CACCTAAACGGACTGGCTGTTT-3') as forward and reverse primers. PCR solutions were made using 1 µL template DNA (100 ng/µL), 1 µL dNTPs (200 µM final), 10 µL 5

× GC buffer, 1.25 µL of each primer (0.5 µM final), 0.5 µL Phusion High-fidelity DNA polymerase (2 U/µL stock, New England Biolabs, UK) and made up to a final volume of 50 µL with milliQ H_2O. PCR was then carried out in Biometra T3 Thermocycler (Biometra Anachem, UK) as per Phusion polymerase instructions with an annealing temperature of primer T_m minus 5 °C and annealing time extended to 10 min. PCR product was then treated with 1 µL DpnI restriction enzyme (New England Biolabs, UK) for 1 hour at 37 °C to digest template DNA and then heat inactivated for 20 min at 80 °C. 2 µL of the product were mixed with 1 µL of ligase buffer, 1 µL of T4 DNA ligase (Roche) and made up to a final volume of 10 µL with milliQ H_2O before incubation overnight at 4 °C. The next day 5 µL ligation product was used to transform 100 µL *E. coli* W3110 competent cells. Final plasmid sequence was confirmed by GATC-Biotech before use in this study.

Table 1. Plasmids used in this work

Plasmid	Function	Reference/source
pHAK13	TorA-scFvM wild type with CyDisCo	**[14]**
pAJ22	TorA-scFvM-C-term26 (additional - SNAIIIIITNKDPNSSSVDKLAAALE′ 3′ to C-terminus) with CyDisCo	**[14]**
pIGM01	CyDisCo	**This study**

Cell culturing, sampling and fractionation

5 mL Luria-Bertani (LB) medium (10 g/L sodium chloride, 10 g/L tryptone, 5 g/L yeast extract) pre-cultures were inoculated from glycerol stocks and grown aerobically overnight at 30 °C, 200 rpm with 1:1000 antibiotic (5 µL, 1 M ampicillin). The next day cultures were diluted to OD_{600} of 0.05 in 50 mL fresh LB and antibiotic and set in triplicate. Cultures were then grown at 30 °C, 200 rpm in 250 mL Erlenmeyer flask and samples were taken every 30 min for an OD_{600} reading. At OD_{600} of 0.4-0.6 plasmids were induced with 0.5 mM IPTG (25 µL, 1 M IPTG). After 3h induction, cells equivalent to a density of OD_{600} of 10 (~8 mL) were taken and fractioned into cytoplasmic (C), insoluble/membrane (M), periplasmic (P) and inclusion bodies (IBs) samples. Periplasmic (P) fractions were collected using an EDTA/lysozyme/cold osmotic shock method previously described [15], with modifications [16]. Spheroplasts were further fractionated into inclusion bodies by centrifugation at 5.000 rpm for 15 min at 4 °C before

separating the cytoplasmic (C) and insoluble/membrane (M) fractions as described earlier [16]. Fractions of three biological replicates from each strain were prepared.

The specific growth rate μ was calculated from three consecutive OD_{600} measurements and using the time points of 1.5 hours as the initial point and 4.5 hours as the final point of the exponential phase. The growth rate was calculated as previously described by McKay *et al*. [17]:

$$\mu = \frac{(ln\, x_2 - ln\, x_1)}{\Delta t}$$

where x_1 and x_2 are the absorbance values measured at 600 nm at 1.5 hours and 4.5 hours, respectively.

Doubling time for the exponentially growing cultures was calculated as:

$$t_d = \frac{ln\, 2}{\mu}.$$

Protein expression analysis by Western blot

Protein was transferred to PVDF-membrane (GE Healthcare, Buckinghamshire, UK) by wet-western-blotting, with electrophoresis for 1 h at 80 V, 300 A. The PVDF membrane was then immersed in blocking solution (2.5% (w/v) skimmed milk powder in 50 mL 1 × PBS-Tween20 (0.1%)) and incubated at 4 °C overnight. The following day the membrane was washed 3 × for 5 min with 1 × PBS-Tween20 (0.1%) before incubation with primary antibody (3.5 μL anti-6-His (Life Technologies, CA, USA) in 20 mL 1 × PBS-Tween20 (0.1%), 3% (w/v) BSA) for 1 h at room temperature. Membranes were then washed 3 × for 5 min with 1 × PBS-Tween20 (0.1%) before incubation with secondary antibody (4 μL anti-Mouse HRP conjugate (Promega, WI, USA) in 20 mL 1 × PBS-Tween20 (0.1%)) for 1 h at room temperature. Finally, membranes were washed for 3 × for 5 min with 1 × PBS-Tween20 (0.1%). Immunoreactive bands were detected using an enhanced chemiluminescence (ECL) kit (BioRad, Herts, UK) following manufacturer's instructions. Bands were visualised using BioRad Gel-doc chemiluminescence imager and associated software. Comparative densitometry of band-intensities was also carried out on a BioRad Gel-doc imager with an Image Lab™ Software 4.1.

Transmission electron microscopy analysis

The cells were fixed (1 % glutaraldehyde, 4 % paraformaldehyde, 0.2 % picric acid, 50 mM sodium azide in 20 mM HEPES buffer pH 7.4) for 5 min at 40 °C by using a microwave processor for laboratory use (H2500 Microwave Processor, Energy Beam Sciences Inc. East Granby, Connecticut, USA), and then for 30 min at room temperature. Finally, samples were stored overnight at 4 °C until further processing. Subsequent to embedding in low gelling agarose, cells were postfixed in 2 % osmium tetroxide in washing buffer (20 mM cacodylate buffer pH 7, 10 mM calcium chloride) for 1 h at room temperature, followed by *en bloc* staining with 2 % uranyl acetate in 0.05 % sodium chloride for 30 min at room temperature and washing steps in between. After dehydration in graded series of ethanol (20 %, 30 %, 50 %, 70 %, 90 % for 10 min each step, 96 % two times for 10 min, 100 % ethanol three times for 10 min) the material was embedded in AGAR 100 resin. Sections were cut on an ultramicrotome (Reichert Ultracut, Leica UK Ltd, Milton Keynes, UK), stained with 4 % aqueous uranyl acetate for 3 min followed by lead citrate for 30 sec and analyzed with a transmission electron microscope LEO 906 (Zeiss, Oberkochen, Germany). Afterwards, the micrographs were edited by using Adobe Photoshop CS6.

Protein digestion and LC-MS analysis

The protein concentration was measured by the Bicinchoninic Acid (BCA) Protein Assay (Thermo Fisher Scientific). Cytoplasmic proteins (100 µg) were reduced with TCEP, alkylated with iodoacetamide and digested in-solution using trypsin as described by Muntel *et al*. [18]. Desalting of peptides prior to mass spectrometry analysis using Stage tips, C18 material (Thermo Fisher Scientific) was performed according to the protocol by Rappsilber *et al*. [19]. For the absolute quantification of proteins from cytoplasmic fraction, the peptide mix of a tryptic digest of yeast alcohol dehydrogenase (Waters, USA) was spiked to a final concentration of 50 fmol/µL. The nanoACQUITY^{TM} UPLC^{TM} system (Waters) was used to separate peptides and introduce the sample into the Synapt G2 (Waters) mass spectrometer. Parameters for liquid chromatography and IMS^{E} were used as described previously by Zühlke *et al*. [20].

Proteins from periplasmic and membrane fractions (30 µg) were separated via 1D SDS-PAGE and the protein staining was followed by Bradford method [21]. The entire gel

lanes were cut into ten pieces each and proteins digested with trypsin (Promega, USA) overnight. Peptides were purified using ZipTip C18 tips (Millipore). The eluted peptides were subjected to LC-MS/MS analysis performed on a Proxeon nLC 1200 coupled online to an Orbitrap Elite (Thermo Fisher Scientific) mass spectrometer. Peptides were separated on in-house self-packed columns (id 100 µm, od 360 µm, length 200 mm; packed with 3.6 µm Aeris XB-C18 reversed-phase material (Phenomenex)) in an 80 min nonlinear gradient from 1% acetonitrile and 0.1% acetic acid to 95% acetonitrile, 0.1% acetic acid. A full MS scan (resolution of 60,000) was acquired using the automatic data-dependent mode of the instrument. After acquisition of the full MS spectra, up to 20 dependent scans (MS/MS) were performed according to precursor intensity by collision-induced dissociation fragmentation (CID) in the linear ion trap. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD010078 (username: reviewer91495@ebi.ac.uk, password: gZ9aEwDr).

Protein identification and quantification

The MS^E spectra of cytoplasmic samples were processed using PLGS v 3.0 and searched against a randomized *E. coli* K12 W3110 UniProt/Swissprot database (Proteome ID: UP000000318, 4 257 entries, version December 2017) with added amino acid sequences of scFv, mutated scFv, yeast mitochondrial thiol oxidase, Erv1p and human protein disulfide isomerase, PDI, laboratory contaminants and yeast ADH1 sequence. Processing and search parameters were described earlier [20]. MS/MS spectra of periplasmic and membrane samples were searched against the above-mentioned database, excluding yeast ADH1, using MaxQuant software (version 1.5.8.3) [22]. Peptide search was performed with the Andromeda search algorithms [23]. For positive protein identification the following criteria had to be met: a minimal peptide length of six amino acids, up to two missed cleavages, carbamidomethylation of cysteine specified as a fixed modification, N-terminal protein acetylation and methionine oxidation were set as variable modifications. The false discovery rate (FDR) was estimated and protein identifications with FDR< 1% were considered acceptable. A protein had to be identified in at least two out of 3 biological replicates, and a minimum of two unique peptides per protein was required for relative quantification using the label free quantification (LFQ) algorithm

provided by MaxQuant. The Student's t test requiring a p-value of < 0.05 was subsequently conducted to determine proteins with significant alterations in protein abundance. Changes in protein abundance of the *E. coli* CyDisCo strains expressing the protein of interest *vs*. wild type (WT) were presented with a log2 fold change (FC). If proteins were quantified in only one of the strains, and were identified in three biological replicates, they were added to the list of identified ON/OFF proteins.

Functional analysis and data visualization

The protein biological functions were assigned based on their TIGRfam roles according to The SEED (http://pubseed.theseed.org/) [24] complemented with manual curation. If the protein was identified in more than one fraction, the subcellular localization was assigned based on the *E. coli* specific STEP database (STEPdb) [25]. Voronoi treemaps were created using the Paver software (DECODON GmbH) [26].

Transcriptional analysis

The cells were harvested right before induction, 30 min and 3 hours after induction, and then normalized using the absorbance measurement. The total RNAs were extracted using RNeasy Plus Mini Kit (Qiagen) with RNAprotect Bacteria Reagent (Qiagen) according to the manufacturer's instructions. The reverse transcription was performed using SuperScript IV Reverse Transcriptase (Invitrogen). Finally, the real-time PCR was run in a QuantStudio 7 Flex (Applied Biosystems) using TaqMan Gene Expression Master Mix in a 20 µL reaction volume with the protocol: 10 min at 95 °C and 40 cycles of 15 sec at 95 °C, 1 min at 60 °C. The gene *rrsH* encoding 16S ribosomal RNA was used as a reference and the $\Delta\Delta C_t$ was averaged over triplicate measurements.

Phase contrast microscopy

Cells were grown in LB medium with ampicillin up to OD_{600} of 0.5 and after that induced with 0.5 mM IPTG. Next, cells were incubated for 3 hours at 30 °C and 180 rpm. For microscopy 10 µL bacterial culture were pipetted onto a thin layer of 1.5 % agar (Sigma-Aldrich). Samples were photographed with an AxioCam MRm (Zeiss) camera mounted on an Axio Imager 2 (Zeiss) fluorescence microscope through an EC Plan-Neofluar 100x/1.3 oil objective. Images were acquired with ZEN 2011 software and exported as TIFF file format.

Aggregation assay

Cells were assessed for cell aggregation by monitoring of sedimentation. Overnight cultures of the strains were adjusted to the same OD_{600} and grown to an OD_{600} of 0.5. Subsequently, bacterial cultures were induced with 0.5 mM IPTG and grown for another 3 hours. Next, 5 mL samples were placed in sterile 10 mL tubes and 1 mL samples were transferred to a cuvette and covered with Parafilm. Samples were always prepared in 3 replicates each and incubated for 24 hours at 30 °C without shaking. Images were acquired using a digital camera. Percentage aggregation was measured as previously described [27].

Results and discussion

Physiological characterization of scFv-secreting CyDisCo strains

A first indication of stress in scFv expressing cells was the metabolic burden. Metabolic burden occurs when the expression of a recombinant protein causes a decrease in generation time by diverting the resources of the cells towards protein production and plasmid maintenance [28]. Under the same growth conditions, we observed that the cells reach stationary plateau at different ODs when the plasmid is present and induced (Fig. 1). These different conditions tax the cells in different ways: when induction occurs, the cells reached the stationary plateau at a lower OD while expressing the protein of interest in both folded and misfolded state in comparison to those expressing only the CyDisCo components. In case of the cells producing the correctly folded scFv, this additional growth stress is most likely due to the recombinant protein synthesis and its export to the periplasm of the protein: although Tat transport does not consume ATP it uses the proton motive force to export the proteins to the periplasm [29], causing an energetic and resource deficit in comparison to the control. In case of the misfolded scFv, the resources are diverted to the synthesis of a recombinant protein in a large quantity, which unlike the correctly folded scFv, is not secreted to the periplasm but deposited in inclusion bodies (visible in Fig. 2C). When embedded in IBs, protein is devoid of any biological activity and primarily cannot be used again in the cell. Cells lacking a plasmid and hence not producing a heterologous protein, reached a maximum OD significantly higher than that when expressing recombinant protein. In summary, the highest metabolic burden is experienced by the cells producing both the CyDisCo components and the protein of

interest (POI), however there is no clear difference between cells producing the correctly folded scFv and those producing the misfolded protein. Although the doubling times for all the strains were very similar (Table 2) we could observe that the lowest metabolic burden occurs when the cells do not have a plasmid followed by cells containing a non-induced plasmid since these reached much higher ODs.

For the proteomic study the time point of 3 hours post induction was chosen since this coincided with the complete growth arrest for the cells producing the correctly folded scFv (Fig. 1) and, at the same time, provided satisfactory yields of recombinant protein (Fig. 3).

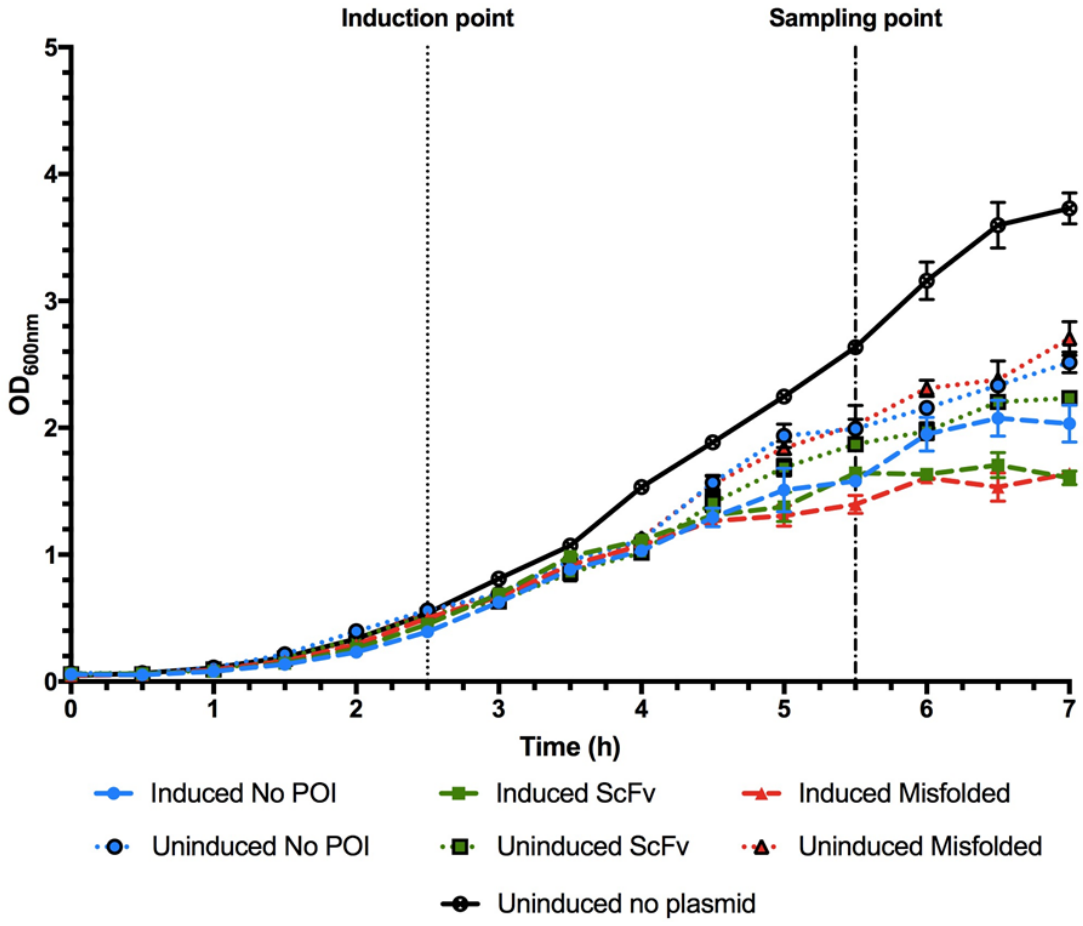


Fig. 1

Growth comparison of *E. coli* CyDisCo induced and uninduced strains expressing folded, misfolded and no protein of interest (POI) as well as a control without plasmid. Linear (A) and semilogarithmic (B) display of *E. coli* growth. Cultivations were performed at 30 °C and 200 rpm for all conditions. Average growth and standard deviation of at least three independent experiments are shown. The dashed and dotted

lines show the time of IPTG induction and sampling point (3 h post-induction), respectively.

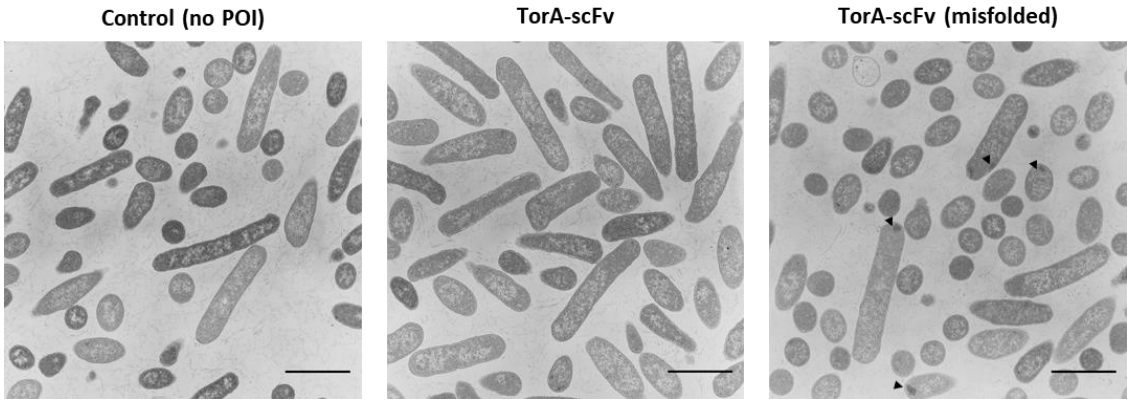


Fig. 2

Transmission electron micrographs of *E. coli* CyDisCo: control (A), scFv (B) and misfolded scFv (C) expressing strains. The cells were grown aerobically in LB medium for 3 hours post-induction with IPTG. The strain expressing misfolded TorA-scFv shows the formation of inclusion bodies (marked with arrowheads). Scale bar = 2 µm.

Table 2 Comparison of the growth rate and doubling time of the exponential phase and maximum OD reached by the cultures after 7 hours

Strains	Specific growth rate* (h^{-1})	Doubling time* (h)	Maximum OD (at 600 nm)
Induced No POI	0.75 ± 0.018	0.92 ± 0.022	2.08
Induced ScFv	0.74 ± 0.005	0.94 ± 0.006	1.71
Induced Misfolded	0.69 ± 0.013	1.01 ± 0.019	1.64
Uninduced No POI	0.66 ± 0.022	1.05 ± 0.035	2.52
Uninduced ScFv	0.69 ± 0.019	1.01 ± 0.028	2.23
Uninduced Misfolded	0.72 ± 0.016	0.96 ± 0.022	2.70
Empty	0.76 ± 0.018	0.91 ± 0.021	3.73

Cultivations were performed at 30 °C and 200 rpm for all conditions. The cells were grown aerobically in LB medium for 7 hours and induced with IPTG after 2.5 hours. The data for the calculation of the growth rate and generation time during the exponential phase was taken from time points at 1.5 hours and 4.5 hours. * Average from three replicates, ± standard deviation.

Total expression of Tat-targeted scFv and misfolded scFv

The misfolded scFv (mf_scFv), consisting of an additional 26-residue unstructured tail at the C-terminus of the folded scFv, is not transported to the periplasm since the Tat system's proofreading mechanism detects it as unfolded [14]. Fig. 2C shows that IBs are formed in the cells expressing the unfolded protein, while Fig. 3 shows that the unfolded scFv is also unprocessed: the signal peptide has not been cleaved and hence the protein runs higher on the SDS gel. A similar phenomenon occurs with the correctly folded scFv since a fraction of the uncleaved protein is also found in inclusion bodies, whereas the bulk of the protein is correctly exported to the periplasm and a small proportion remains in the cytoplasm. In both cases we can observe that there is less accumulation in inclusion bodies for the folded scFv and a densitometry analysis shows that the protein present in inclusion bodies in the unfolded scFv is 6.86 times more abundant than that of the folded. This accumulation in IBs in the cytoplasm may be due to the low abundance of Tat machinery available to transport the protein across the inner membrane, since over-expression of the Tat components leads to more efficient export [30]. Finally, it has been previously shown and discussed that the presence of the mature-sized scFv in the cytoplasm may be due to early cleavage taking place by cytoplasmic peptidases [10].

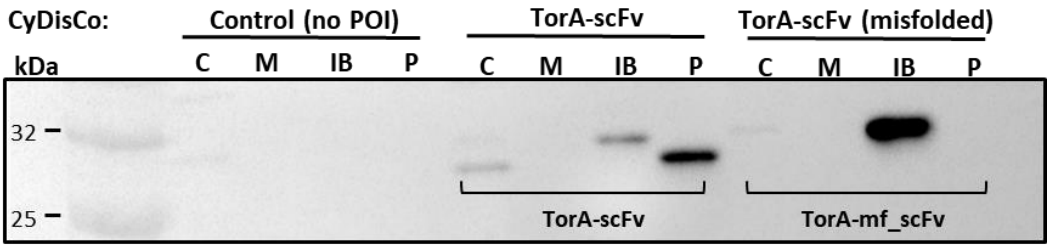


Fig. 3

Accumulation of the scFv construct and its misfolded variant in CyDisCo cells.

TorA-scFv and the misfolded TorA-scFv constructs were expressed in *E. coli* cells together with Erv1p and PDI, CyDisCo components. After expression, cells were fractionated into cytoplasmic (C), membrane (M), periplasmic (P) and inclusion body (IB) samples and extracted proteins separated via SDS-PAGE.

Impact of elevated scFv secretion and misfolded scFv accumulation on the E. coli proteome

In the previous section, we showed that the expression of TorA-scFv in CyDisCo cells results in a high secretion of the protein to the periplasm and minor accumulation in IBs,

whereas the misfolded version aggregates in the IB fraction. Both processes are likely to cause some similar proteome stress responses associated with recombinant protein production, but we were most interested to determine whether diverse responses occurred due to the intensified secretion of scFv and its accumulation in the periplasm, or an aggregation of oxidized misfolded scFv in the cytoplasm.

To study the stress responses, we performed a subcellular fractionation and applied LC-IMS^{E} and LC-MS/MS approaches for a comprehensive proteome analysis. Quantification of cytoplasmic, periplasmic and membrane/insoluble proteins was performed for samples harvested 3 hours post-induction (Supplementary Table 1). In our study we identified a total of 1629 proteins shared between three studied *E. coli* strains (Fig. 4 A). Among these, 441 proteins of the strain secreting scFv and 393 proteins of the strain expressing mf_scFv exhibited an at least two-fold change in expression when compared to the control (Supplementary Table 2). Changes in the amount of proteins are visualized by Voronoi treemaps (Fig. 4 B-D). The proteins are clustered by their functional category (TIGRfam complemented with manual curation) so that functionally related gene products are localized in one cluster. We distinguished several patterns of changes in the abundance of functionally related proteins and described them in the following section, providing log2 FC ratios for statistically significant results ($p < 0.05$). The first conclusion is that the expression of misfolded protein results in downregulation of many proteins involved in general metabolism and membrane transport. In addition, strains expressing either misfolded or correctly folded scFv, showed a high number of significant changes mostly in proteins associated with protein folding and degradation, oxidative stress, membrane transport and integrity.

2.2 Comparative proteome analysis in an Escherichia coli CyDisCo strain identifies stress responses related to protein production oxidative stress and accumulation of misfolded protein

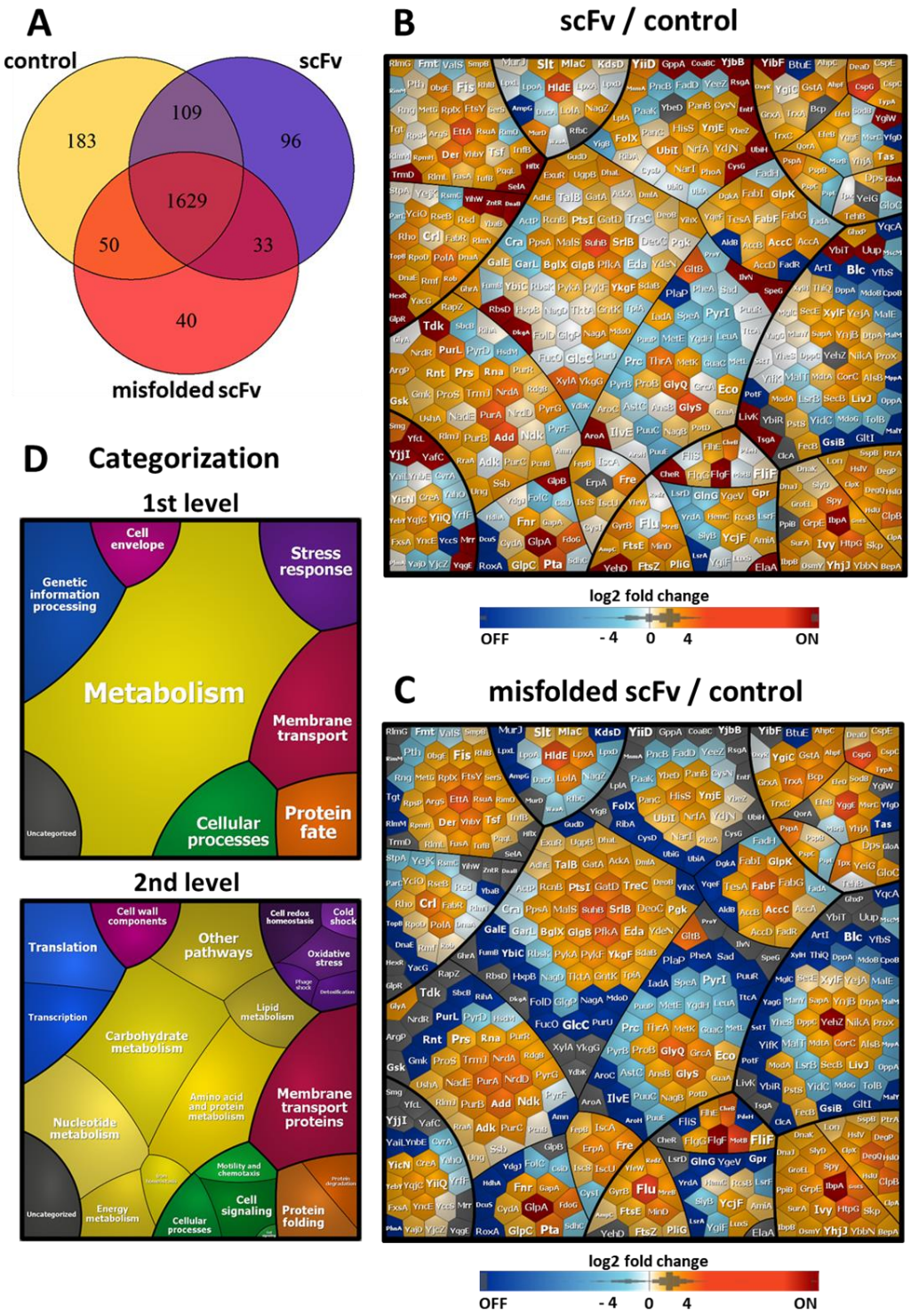


Fig. 4

Analysis of *E. coli* CyDisCo proteome changes in the strains expressing scFv and misfolded scFv *vs.* an empty plasmid control. Venn diagram showing the number of unique and shared proteins identified in *E. coli* CyDisCo strains (A). Proteome changes are shown as the change in protein abundance (Δlog2 ratio) of strains expressing: TorA-scFv *vs*. control (B) and TorA-mf_scFv *vs*. control (C). Changes in protein abundance

were color coded: orange indicates increase in the scFv and scFv (misfolded) proteome, gray indicates unidentified, and blue indicates decrease in the scFv and scFv (misfolded) proteome. Additionally, proteins which were identified in scFv expressing strains but not in the control strain were colored red (ON proteins) and proteins identified only in control strain but not in scFv expressing strains were colored navy blue (OFF proteins). The treemap legend shows the classification of the *E. coli* proteome according to TIGRfam annotations with manual adjustments (D).

Folded vs. misfolded scFv stress response affecting transcription factors and DNA replication

The adaptation to changes in growth conditions is often regulated by alternative sigma factors and anti-sigma factors [31]. Sigma factors are multi-domain subunits of the bacterial RNA polymerase and essential for the initiation of transcription. They are involved in the recognition of promoters and participate in the initial steps of RNA synthesis [32]. The proteomic analysis revealed that there was an increase in the levels of different sigma factors in the cells producing misfolded and folded scFv. This observation confirms that these cells are compensating for the additional production of the protein of interest by increasing the capacity of the cells to handle the metabolic burden. However, the results for the upregulation of transcription factors identified in our study failed to reach the desired level of statistical significance, thus they will be not discussed in detail.

Additionally, we observed an upregulation of many proteins involved in DNA replication occurring only in the cells producing the correctly folded TorA-scFv. For example, a single-stranded DNA-binding protein Ssb (1.46 log2 FC scFv *vs.* control) or even a protein as essential in DNA replication as PolA (DNA polymerase I) (0.85 log2 FC scFv *vs.* control in the cytoplasmic fraction) were found to be upregulated in scFv producing strain whereas the cells expressing mf_scFv showed no change. The upregulation of proteins involved in DNA replication only in cells producing the folded protein suggests that, even though according to the growth curve all three conditions seem to either have reached (scFv and mf_scFv) or be entering (control) the stationary phase, the strain producing folded scFv is maintaining expression of proteins involved in DNA replication, repair and recombination [33].

Finally, we would like to highlight the upregulation of RlmJ (previously known as YhiR) in both strains producing folded and misfolded protein (1.40 and 0.76 log2 FC,

respectively). RlmJ is involved in the catabolism of external DNA [34] and has been identified to repress cell-to-cell plasmid transfer [35]. Notably, the abundance of RlmJ is lower in the cells producing mf_scFv, meaning that even though the misfolded protein stress seems to be partially responsible for the observed upregulation, this effect is greater with the production and transport of a folded protein to the periplasm.

Protein folding and degradation processes

To assure the proteome stability and functionality during high-rate protein expression, *E. coli* activates the proteome quality control system termed the proteostasis network [36, 37]. This system consists of several complex protein machines that under stress conditions maintain proteome homeostasis by a proper regulation of protein folding and degradation processes. Here, we identified main chaperones and proteases upregulated upon scFv and mf_scFv expression in CyDisCo and describe their functional implication in proteome stability.

DnaK-DnaJ-GrpE, and GroEL-GroES are the best characterized molecular chaperones in the cytoplasm of *E. coli*. Both systems can recognize unfolded proteins, refold or direct them to the cellular protein degradation machinery [38]. We observed the upregulation of chaperones from above mentioned systems in the cytoplasmic fraction of *E. coli* expressing the recombinant protein. However, the highest abundance of these proteins was observed in the membrane fraction containing IBs, and the upregulation of GroES-GroEL and DnaK was notably increased in the strain expressing mf_scFv (GroES, 2.33; GroEL, 1.43; DnaK, 1.10 log2 FC mf_scFv *vs*. control). This observation correlates with our previous results and confirms the aggregation of moderate amounts of scFv in the cytoplasm and the high-level accumulation of mf_scFv in IBs. Moreover, the abundance of ClpB, a chaperone cooperating with DnaK to reactivate strongly aggregated proteins [39], was significantly increased in cells producing scFv and mf_scFv (2.39 and 2.84 log2 FC, respectively). The inclusion body formation in the CyDisCo scFv and mf_scFv strains was also confirmed by the presence of IbpA, a heat shock protein with a strong ability to associate to protein aggregates, that was not identified in the empty plasmid control strain, in which we did not find any traces of IBs. Additionally, several other chaperones such as HtpG involved in *de-novo* protein folding in the cytoplasm [40]; Spy (2.37 and 2.25 log2 FC) inhibiting protein aggregation and promoting refolding in the periplasm [41]; and Skp (1.44; 1.48), SurA(1.36; 1.24), BepA (1.46; 1.14) accelerating

folding and improving insertion of outer membrane proteins [42, 43], were found to be highly abundant in the scFv and mf_scFv expressing strains, respectively. The accumulation of protein aggregates in the cytoplasm of *E. coli* CyDisCo is followed by activation of protein degradation machineries. We observed the upregulation of the ATPase components ClpA and ClpX of the Clp protease, Lon and HslV/U, together with periplasmic proteases DegP/Q. All these proteins are responsible for recognition, resolubilization and degradation of unfolded proteins. The highest difference between the abundance of scFv and mf_scFv expressing strains and the control was identified for periplasmic DegP and DegQ (1.53, 2.56; 1.85, 3.10 log2 FC for scFv and mf_scFv, respectively) proteases. This suggests that in both conditions, scFv secretion to the periplasm and mf_scFv aggregation in the cytoplasm, the cell envelope is exposed to unfavourable and harmful changes.

The high upregulation of periplasmic proteases in the strain expressing misfolded scFv is particularly interesting. This could indicate that some of the protein is in fact successfully exported to the periplasm, rather than being blocked by the Tat proofreading machinery, and subsequently degraded. However, a previous study showed no evidence whatsoever for transport of this misfolded substrate to the periplasm and in our view it is more likely that the cell either senses the production of precursor protein and upregulates expression, or the proteases are upregulated as part of a general stress response.

Oxidative stress and redox regulation in the cytoplasm

The principle behind disulfide bond formation in the CyDisCo strain, as described before, is the co-expression of Erv1p and PDI in the cytoplasm. This system allows the oxidation of a dithiol to a disulfide in the cytoplasm, and at the same time does not require the disruption of any naturally occurring reduction pathways. The expression of two exogenous oxidases changes the cytoplasm into an oxidized environment, where the disulfide bonds can form efficiently, which is proven by the effective secretion of disulfide bond-containing scFv to the periplasm through the Tat system (Fig. 3). However, since the oxidative state of the cytoplasm does not occur naturally, we expected that this will provoke an oxidative stress response and changes in the abundance of redox proteins.

Indeed, the proteomic analysis revealed the upregulation of several proteins related to oxidative stress response. Among these, major thiol-disulfide oxidoreductases: thioredoxin TrxA, glutaredoxin GrxA, together with thiol-specific peroxidase Bcp, were

upregulated in strains expressing scFv (with statistically significant upregulation only in mf_scFv). The upregulation of oxidoreductases was significantly higher in the strain producing misfolded scFv, probably due to the increased protein accumulation in IBs (e.g. GrxA, 1.19; TrxA, 2.02, Bcp, 2.08 log2 FC mf_scFv *vs*. control). In the cytoplasm of *E. coli*, thioredoxin and glutathione/glutaredoxin systems are required for the reduction of proteins that become oxidized, and thus they maintain the redox homeostasis [11]. However, during severe oxidative stress both systems can be overwhelmed, or their protein function can shift from reductase to oxidase, leaving cytosolic proteins susceptible to oxidation. It has been shown that thioredoxins and glutaredoxins can act as oxidases when exported to the oxidizing periplasm [44, 45] and both, TrxA and TrxC, shift their function to oxidases in a *trxB* mutant [46]. A related point to consider is the fact that in the periplasm of *E. coli*, which is an oxidizing environment, other thioredoxin-related proteins such as DsbA, B, C, D, G catalyze the disulfide-bond formation and rearrangement. Thus, it appears that the function of thioredoxins and glutaredoxins is influenced by the redox potential of the cellular environment and in an oxidized cytoplasm of *E. coli* CyDisCo they will promote disulfide bond formation, rather than their reduction.

In conclusion, the production of disulfide bond-containing scFv in CyDisCo increases the abundance of cellular redox proteins in the cytoplasm. We hypothesize that the abundance of those proteins is regulated by the amount of oxidized protein aggregates in the cytoplasm, thus it is larger in the mf_scFv strain. Since the oxidizing potential of the cytoplasm in CyDisCo strains is not compromised, we assume that the redox processes are either not efficient, or in the oxidizing environment the oxidoreductases function is shifted to thiol oxidases.

Membrane transporters

It was interesting to note that the components of three major secretion machineries such as the Sec pathway, Tat system and YidC showed similar changes in their abundance in CyDisCo strains expressing scFv and its misfolded variant. Essentially, we might expect to find Tat components more abundant in the strain secreting high levels of scFv to the periplasm. However, we observed moderately lower levels of TatA and TatB, as well as

a higher level of TatE in CyDisCo strains expressing scFv and mf_scFv. We elaborate on this point in the next section.

The amounts of Sec system components showed an increased abundance in strains producing recombinant protein, with a log2 FC higher than 1 for SecB and SecE. The Sec protein translocation pathway is involved in transport of most of the membrane-bound and secreted proteins (reviewed in [47]). The upregulation of Sec machinery is likely to be an effect of elevated oxidative stress in the periplasm and changes in the cell envelope, to which the cell is responding by secretion of cytoprotective stress proteins. The Sec system is responsible for insertion of a large number of proteins into the *E. coli* membrane, however some of the membrane proteins require YidC activity for their insertion [48]. Interestingly, in our samples YidC was found in lowered amounts in the *E. coli* secreting scFv (-2.29 log2 FC) and mf_scFv (-2.57 log2 FC). Additionally, we have also observed a drastic decrease in the amount of ABC transporter proteins responsible for the utilization of di- and tripeptides. In *E. coli*, the oligopeptide transport systems play a main role in nutrition, but the transported peptides also serve as substrates for integral membrane assembly and take part in the functioning of quorum-sensing pathways [49]. In this study, particularly the substrate-binding proteins (SBPs) of the main *E. coli* peptide transporters were present in much-decreased amounts (DppA, -5.74, -5.60; OppA, -3.68, - 2.97 and MppA, OFF, -3.04 log2 FC in scFv *vs.* control and mf_scFv *vs.* control, respectively). These SBPs recognize the substrate and subsequently deliver it to the translocator, thus the decrease in the SBPs amount reduces the peptide transport rate. Moreover, some essential components of other transport systems such as maltose (MalE/M/T/Y), glutathione (GsiB), Tol-Pal (TolB, CpoB) and lipid (Blc), or osmoregulated- periplasmic glucan (OPG) biosynthesis pathway (MdoB) were downregulated. One feature that these downregulated proteins have in common is their role in membrane protein synthesis and membrane maintenance under stress conditions. This suggests that the scFv expression in the *E. coli* CyDisCo causes the membrane impairment through a decrease in number of membrane transporters. We hypothesize that this observed decrease in membrane transporters abundance can be an effect of YidC downregulation, since it has an important role in membrane protein folding and insertion into the membrane [50, 51]. Generally, it is plausible that downregulation of so many transporter proteins involved in metabolite uptake is a consequence of previously mentioned metabolic burden that occurs in cells expressing scFv and mf_scFv. The non-

stressed cells (empty plasmid control) are still in exponential phase, showing higher metabolic and biosynthetic activity than cells expressing POI, already in stationary phase at the harvesting time.

Decrease of tat genes transcription when expressing a Tat substrate

The proteomic analysis showed a moderate increase in the abundance of TatE (0.37; 0.89 log2 FC for the scFv and mf_scFv *vs.* control, respectively) and a decrease of TatA and TatB proteins (however not statistically significant) in cells expressing scFv and misfolded scFv in comparison to the control. These observations raised the question of whether the expression of *tatA* and *tatB* genes is downregulated or the proteins have a shorter half-life upon overexpression of a Tat substrate. To address this question, an RT-qPCR experiment was conducted on the cells grown under the same conditions as for the proteomic study and reported relatively to the control (Fig. 5). A time course experiment was run including points just before and after induction, to assess the impact of IPTG on the cells. Only one significant change was observed due to induction; an increase of 69% was detected between pre- and post-induction samples for *tatA* expression with the folded scFv. Therefore, the induction has an immediate impact on expression level of *tatA* only when expressing a Tat-suitable substrate. Since the *tatABCD* genes are part of a single operon, this result means that the genes experience a differential transcription. Moreover, the higher transcription of *tatA* compared to *tatB* matches the previously established fact that *E. coli* contains more TatA than TatB proteins [52].

After induction, a significant decrease of transcription levels of *tatA* and *tatB* genes is observed when expressing TorA-scFv, whether correctly or incorrectly folded, by over 30 %. This decrease matches the lower abundance of TatA and TatB proteins detected 3 hours after induction. Therefore, less TatA and TatB proteins are expressed by the cell upon overexpression of a Tat substrate no matter whether it can fold correctly or not. These observations are surprising, given that the cells are expressing such high amounts of Tat substrate. The decrease in the amount of Tat operon being expressed seems to be partly compensated by the increase of TatE protein available in the cell. TatA and TatE share 53% homology and have overlapping functions. It has been shown that individual deletions of TatA and TatE are still functional, however the double deletion blocks export to the periplasm via the Tat pathway [53]. This preference for the use of either TatA or TatE is understudied and may be due to the additional stress conditions of the study or

for a determined preference for the TorA signal peptide employed. Another explanation for this behaviour may be the fact that the Tat pathway exports few and complex proteins to the periplasm, many of them involved in cell division, biofilm formation and pathogenicity [7]. These proteins are in most cases not needed in high concentrations in the periplasm, and this negative feedback loop may be responsible for assuring that the export remains low to avoid toxicity in the periplasm.

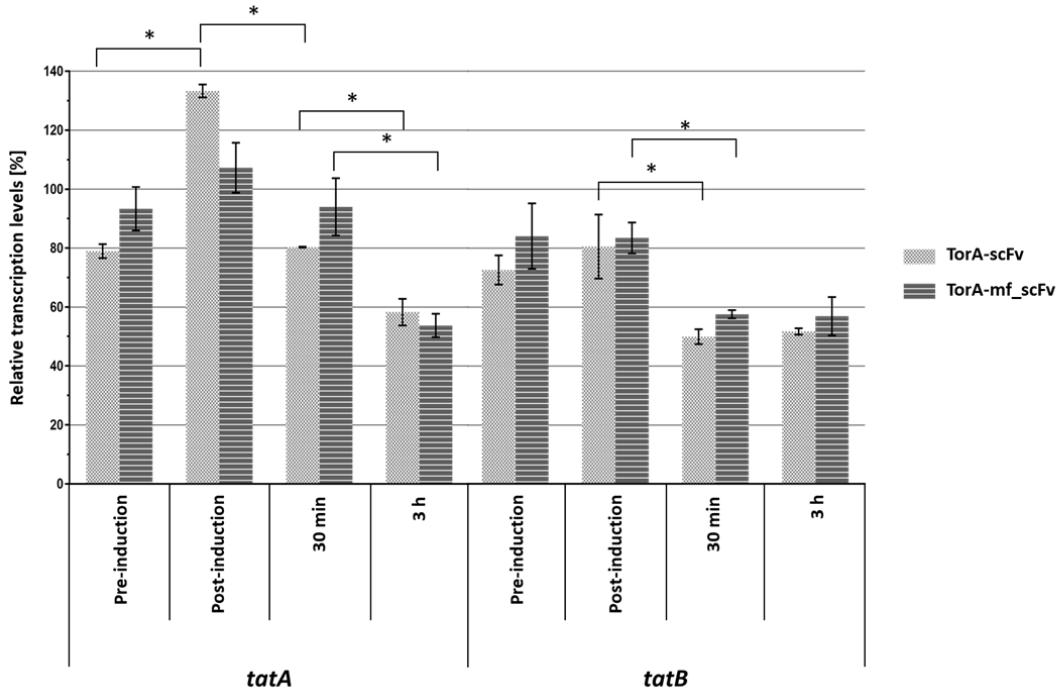


Fig. 5

Transcription levels of *tatA* and *tatB* genes calculated from RT-qPCR measurements when expressing folded scFv (dark grey) or misfolded scFv (light grey) at four different time points after normalization to the control vector. The mean and standard deviations of three replicates are shown. * $p < 0.02$.

Misfolded scFv expression induces Ag43-mediated aggregation in E. coli CyDisCo cells

Our proteome studies revealed significantly higher (5.71 log2 FC) abundance of antigen 43 (Ag43) in the strain expressing a misfolded scFv. Ag43 is an outer membrane protein that confers bacterial cell-to-cell aggregation [54]. To our knowledge, the formation of cell aggregates connected with Ag43 expression has not yet been reported as a consequence of a cellular stress caused by the accumulation of misfolded protein in the cytoplasm. To verify if Ag43 upregulation promotes the cell aggregation in our strain, we

first performed phase-contrast microscopy (Fig. 6 A1-3). Additionally, since the cell aggregation leads to enhanced sedimentation, we applied a simple sedimentation assay (Fig. 6 B) and calculated the aggregation rate in all CyDisCo strains (Fig. 6 C). Indeed, a comparison of the aggregation and sedimentation rate of these strains confirmed that the *E. coli* expressing misfolded scFv shows enhanced sedimentation and aggregation. We propose that the increase in the abundance of Ag43 is related to the lowered abundance of the transcription factor OxyR, which is a suppressor of antigen 43 expression [55]. OxyR transcription factor is activated through the formation of a disulfide bond and is deactivated by enzymatic reduction with glutaredoxin 1, GrxA [56]. We observed a decrease in the abundance of OxyR in mf_scFv (-0.67 log2 FC) and its upregulation in scFv expressing strain. These results could only be observed after the induction of the scFv expression, thus we assume that the accumulation of misfolded scFv leads to the Ag43-mediated cell clustering. We believe that the upregulation of Ag43 synthesis is related to the accumulation of oxidized misfolded scFv which perturbs the thiol-disulfide balance in the cytoplasm and causes a situation referred as 'disulfide stress'. Interestingly, the downregulation of OxyR was only observed in the mf_scFv expressing strain, in which we observed higher upregulation of redox proteins. This suggests that the activity of OxyR is responsive to the thiol-disulfide redox status of the cell, however its abundance in the cell is regulated by another, yet not known, mechanism. Additionally, the absence of CpxR, an antagonistic regulator of adhesion in *E. coli* [57], could contribute to the enhanced cell aggregation under misfolded protein expression stress. The aggregation phenomenon was not accompanied by the increase in bacterial motility in the strain expressing misfolded scFv (Supplementary Figure 1).

2.2 Comparative proteome analysis in an Escherichia coli CyDisCo strain identifies stress responses related to protein production oxidative stress and accumulation of misfolded protein

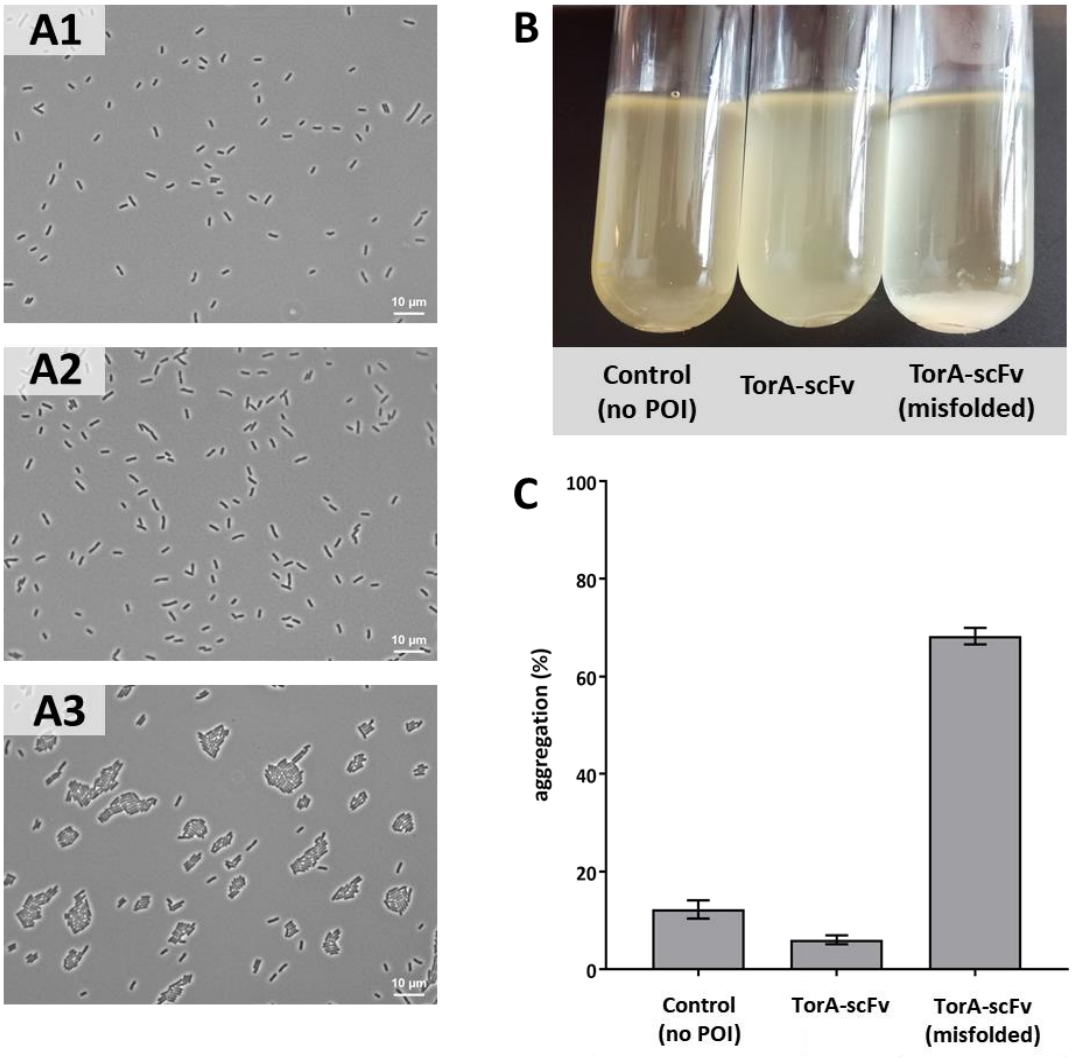


Fig. 6

The *E. coli* CyDisCo strain expressing misfolded scFv shows Ag43-mediated cell aggregation. Phase-contrast microscopy images of CyDisCo: control (A1), scFv (A2) and misfolded scFv (A3) expressing strains. Samples were harvested 3 h post-induction with IPTG. (B) Sedimentation profile of *E. coli* CyDisCo strains. Cultures were grown for 3 h post-induction, transferred into tubes and incubated statically for 24 h. (C) Percentage of aggregation as quantified from the change in OD_{600} over 24 hours from the experiment performed as in (B).

Conclusions

The *E. coli* CyDisCo strain enables folding of disulfide bond-containing scFv in the cytoplasm and its secretion in high amounts to the periplasm via the Tat system. The expression of an unfolded scFv led to protein aggregation and inclusion body formation. Decreased cell growth, low abundance of proteins involved in DNA replication and sigma 70 factor in *E. coli* expressing a misfolded scFv when compared to a folded scFv suggest that the unfolded protein stress accelerates cell transition to a stationary phase. We confirmed that the recombinant protein overexpression triggers the upregulation of proteins involved in removing the aggregates, such as chaperones and proteases. The high abundance of GroEL-GroES, DnaK and ClpB in the insoluble fraction of *E. coli* expressing misfolded protein suggests their association with protein aggregates. We also observed that the formation of disulfide bonds in the cytoplasm of *E. coli* CyDisCo initiates the oxidative stress response and increases the abundance of redox regulation system proteins. This upregulation of thioredoxins and glutaredoxins was higher in the strain accumulating the misfolded oxidized scFv in the cytoplasm. Nevertheless, the redox protein ability to reduce oxidized proteins seems to be compromised by the oxidative environment of the cytoplasm.

Interestingly, we observed an extreme downregulation of membrane transporters which could be related to a major decrease in the abundance of YidC transport system, responsible for protein folding and insertion into the membrane. Surprisingly, a high rate of Tat-dependent secretion caused an increase in *tatA* expression only just after induction of synthesis. The following post induction analysis revealed lower *tatA* and *tatB* expression levels, which correlate with lower TatA and TatB protein abundance. This could suggest that the presence of the Tat signal peptide actively regulates the amount of Tat components being produced in a feedback loop to prevent an unnecessarily high concentration of Tat substrates in the periplasm. Finally, the most distinct unfolded protein stress response we observed was the cell aggregation caused by the elevated levels of antigen 43. Interestingly, to the best of our knowledge, this phenomenon has not been reported yet as a consequence of misfolded protein expression, which makes it an interesting topic for further investigation.

References

1. Sandkvist M, Bagdasarian M. Secretion of recombinant proteins by Gram-negative bacteria. Curr Opin Biotechnol. 1996; doi.org/10.1016/S0958-1669(96)80053-X.
2. Osborne AR, Rapoport TA, van den Berg B. Protein Translocation by the Sec61/SecY channel. Annu Rev Cell Dev Biol. 2005; doi.org/10.1146/annurev.cellbio.21.012704.133214.
3. Robson A, Collinson I. The structure of the Sec complex and the problem of protein translocation. EMBO Rep. 2006; doi.org/10.1038/sj.embor.7400832.
4. Palmer T, Berks BC. The twin-arginine translocation (Tat) protein export pathway. Nat Rev Microbiol. 2012; doi.org/10.1038/nrmicro2814.
5. Müller M, Klösgen RB. The Tat pathway in bacteria and chloroplasts (Review). Mol Membr Biol. 2005; doi.org/10.1080/09687860500041809.
6. Berks BC. The Twin-Arginine Protein Translocation Pathway. Annu Rev Biochem. 2015; doi.org/10.1146/annurev-biochem-060614-034251.
7. Böck A, Curtiss RI, Kaper JB, Karp PD, Neidhardt FC, Nyström T, Slauch JM, Squires CL, UseryD. EcoSal— *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. ASM Press, Washington, DC; 2011.
8. Rodrigue A, Chanal A, Beck K, Müller M, Wu LF. Co-translocation of a Periplasmic Enzyme Complex by a Hitchhiker Mechanism through the Bacterial Tat Pathway. J Biol Chem. 1999; doi.org/10.1074/jbc.274.19.13223.
9. Dolata KM, Montero IG, Miller W, Sievers S, Sura T, Wolff C, Schuter R, Riedel K, Robinson C. Far-reaching cellular consequences of tat deletion in *Escherichia coli* revealed by comprehensive proteome analyses. Microbiol Res. 2018; doi.org/10.1016/j.micres.2018.10.008.
10. Alanen HI, Walker KL, Lourdes Velez Suberbie M, Matos CF, Bönisch S, Freedman RB, Keshavarz-Moore E, Ruddock LW, Robinson C. Efficient export of human growth hormone, interferon α2b and antibody fragments to the periplasm by the *Escherichia coli* Tat pathway in the absence of prior disulfide bond formation. Biochim Biophys Acta Mol Cell Res. 2015; doi.org/10.1016/j.bbamcr.2014.12.027.
11. Prinz WA, Åslund F, Holmgren A, Beckwith J. The Role of the Thioredoxin and Glutaredoxin Pathways in Reducing Protein Disulfide Bonds in the *Escherichia coli* Cytoplasm. J Biol Chem. 1997; doi.org/10.1074/jbc.272.25.15661.
12. DeLisa MP, Tullman D, Georgiou G. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. Proc Natl Acad Sci USA. 2003; doi.org/10.1073/pnas.0937838100.
13. Matos CF, Robinson C, Alanen HI, Prus P, Uchida Y, Ruddock LW, Ruddock LW, Freedman RB. Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains. Biotechnol Prog. 2014; doi.org/10.1002/btpr.1858.
14. Jones AS, Austerberry JI, Dajani R, Warwicker J, Curtis R, Derrick JP, Robinson C. Proofreading of substrate structure by the Twin-Arginine Translocase is highly

dependent on substrate conformational flexibility but surprisingly tolerant of surface charge and hydrophobicity changes. Biochim Biophys Acta Mol Cell Res. 2016; doi.org/10.1016/j.bbamcr.2016.09.006.

15. Randall LL, Hardy JS, Correlation of competence for export with lack of tertiary structure of the mature species: A study in vivo of maltose-binding protein in *E. coli*. Cell. 1986; doi.org/10.1016/0092-8674(86)90074-7.
16. Pierce JJ, Turner C, Keshavarz-Moore E, Dunnill P. Factors determining more efficient large-scale release of a periplasmic enzyme from *E. coli* using lysozyme. J Biotechnol. 1997; doi.org/10.1016/S0168-1656(97)00116-8.
17. McKay AL, Peters AC, Wimpenny JWT. Determining specific growth rates in different regions of *Salmonella typhimurium* colonies. 1997; Lett Appl Microbiol doi.org/10.1046/j.1472-765X.1997.00354.x.
18. Muntel J, Hecker M, Becher D. An exclusion list based label-free proteome quantification approach using an LTQ Orbitrap. Rapid Commun. Mass Spectrom. 2012; doi.org/10.1002/rcm.6147.
19. Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc. 2007; doi.org/10.1038/nprot.2007.261.
20. Zühlke D, Dörries K, Bernhardt J, Maaß S, Muntel J, Liebscher V, Pané-Farré J, Riedel K, Lalk M, Volker U, Engelmamn S, Becher D, Fuchs S. Costs of life - Dynamics of the protein inventory of *Staphylococcus aureus* during anaerobiosis. Sci Rep. 2016; doi.org/10.1038/srep28172.
21. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem. 1976; doi.org/10.1016/0003-2697(76)90527-3.
22. Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol. 2006; doi.org/10.1038/nbt.1511.
23. Cox J, Neuhauser N, Michalski A, Scheltema RA, Olsen JV, Mann M. Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res. 2011; doi.org/10.1021/pr101065j.
24. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 2014; doi.org/10.1093/nar/gkt1226.
25. Orfanoudaki G, Economou A. Proteome-wide subcellular topologies of E. coli polypeptides database (STEPdb). Mol. Cell Proteomics. 2014; doi.org/10.1074/mcp.O114.041137.
26. Bernhardt J, Michalik S, Wollscheid B, Völker U, Schmidt F. Proteomics approaches for the analysis of enriched microbial subpopulations and visualization of complex functional information. Curr Opin Biotechnol. 2013; doi.org/10.1016/j.copbio.2012.10.009.

27. Dorken G, Ferguson GP, French CE, Poon WCK. Aggregation by depletion attraction in cultures of bacteria producing exopolysaccharide. J R Soc Interface. 2012; doi.org/10.1098/rsif.2012.0498.
28. Bentley WE, Mirjalili N, Andersen DC, Davis RH, Kompala DS. Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. Biotechnol Bioeng. 1990; doi/abs/10.1002/bit.260350704.
29. Natale P, Brüser T, Driessen AJM. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane- Distinct translocases and mechanisms. Biochim Biophys Acta - Biomembranes. 2008; doi.org/10.1016/j.bbamem.2007.07.015.
30. Browning DF, Richards KL, Peswani AR, Roobol J, Busby SJW, Robinson C. *Escherichia coli* "TatExpress" strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway. Biotechnol Bioeng. 2017; doi.org/10.1002/bit.26434.
31. Saecker RM, Record MT, Dehaseth PL. Mechanism of bacterial transcription initiation: RNA polymerase-Promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. J Mol Biol. 2011; doi.org/10.1016/j.jmb.2011.01.018.
32. Paget M. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. Biomolecules. 2015; doi.org/10.3390/biom5031245.
33. Meyer RR, Laine PS. The single-stranded DNA-binding protein of *Escherichia coli*. Microbiol Rev. 1990;54(4):342-380
34. Finkel SE, Kolter R. DNA as a Nutrient: Novel Role for Bacterial Competence Gene Homologs. J Bacteriol, 2001; doi.org/10.1128/jb.183.21.6288-6293.2001.
35. Kurono N, Matsuda A, Etchuya R, Sobue R, Sasaki Y, Ito M, Ando T, Maeda S. Genome-wide screen for *Escherichia coli* genes involved in repressing cell-to-cell transfer of non-conjugative plasmids. Biochem Biophys Res Commun. 2012; doi.org/10.1016/j.bbrc.2012.10.098.
36. Anfinsen CB. Principles that Govern the Folding of Protein Chains. Science. 1973; doi.org/10.1126/science.181.4096.223.
37. Y. Powers ET, William EB. Diversity in the origins of proteostasis networks- a driver for protein function in evolution. Nat Rev Mol Cell Biol. 2013; doi.org/10.1038/nrm3542.
38. Ziemienowicz A, Skowyra D, Zeilstra-Ryalls J, Fayet O, Georgopoulos C, Zylicz M. Both the *Escherichia coli* chaperone systems, GroEL/GroES and DnaK/DnaJ/GrpE, can reactivate heat-treated RNA polymerase. Different mechanisms for the same activity. J Biol Chem. 1993;34:25425-31.
39. Kedzierska S, Chesnokova LS, Witt SN, Zolkiewski M. Interactions within the ClpB/DnaK bi-chaperone system from *Escherichia coli*. Arch Biochem Biophys. 2015; doi.org/10.1016/j.abb.2005.10.005.
40. Thomas JG, Baneyx F. ClpB and HtpG facilitate de novo protein folding in stressed *Escherichia coli* cells. Mol Microbiol. 2000; doi.org/10.1046/j.1365-2958.2000.01951.x.

41. Quan S, Koldewey P, Tapley T, Kirsch N, Ruane KM, Pfizenmaier J, Shi R, Hofmann S, Foit L, Ren G, Jakob U, Xu Z, Cygler M, Bardwell JC. Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. Nat Struct Mol Biol. 2011; doi.org/10.1038/nsmb.2016.
42. Sklar JG, Wu T, Kahne D, Silhavy TJ. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. Genes Dev. 2007; doi.org/10.1101/gad.1581007.
43. Narita S, Masui C, Suzuki T, Dohmae N, Akiyama Y. Protease homolog BepA (YfgC) promotes assembly and degradation of β-barrel membrane proteins in *Escherichia coli*. Proc Natl Acad Sci U S A. 2013; doi.org/10.1073/pnas.1312012110.
44. Debarbieux L, Beckwith J. The reductive enzyme thioredoxin 1 acts as an oxidant when it is exported to the *Escherichia coli* periplasm. Proc Natl Acad Sci USA. 1998;95(18):10751-6.
45. Eser M, Masip L, Kadokura H, Georgiou G, Beckwith J. Disulfide bond formation by exported glutaredoxin indicates glutathione's presence in the *E. coli* periplasm. PNAS. 2009; doi.org/10.1073/pnas.0812596106.
46. Stewart EJ, Åslund F, Beckwith J. Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins. The EMBO Journal. 1998; doi.org/ 10.1093/emboj/17.19.5543.
47. Danese PN, Silhavy TJ. Targeting and assembly of periplasmic and outer-membrane proteins in *Escherichia coli*. Annu Rev Genet. 1998; doi.org/10.1146/annurev.genet.32.1.59.
48. Petriman NA, Jauß B, Hufnagel A, Franz L, Sachelaru I, Drepper F, Warscheid B, Koch HG. The interaction network of the YidC insertase with the SecYEG translocon, SRP and the SRP receptor FtsY. Sci Rep. 2018; doi.org/10.1038/s41598-017-19019-w.
49. Gomolplitinant KM, Saier MH. Evolution of the Oligopeptide Transporter Family. J Membr Biol. 2011; doi.org/10.1007/s00232-011-9347-9.
50. Wagner S, Pop OI, Haan GJ, Baars L, Koningstein G, Klepsch MM. Biogenesis of MalF and the MalFGK(2) maltose transport complex in *Escherichia coli* requires YidC. J Biol Chem. 2008; doi.org/10.1016/j.bbabio.2011.12.006.
51. van Bloois E, Nagamori S, Koningstein G, Ullers RS, Preuss M, Oudega B, Harms N, Kaback HR, Hermann JM, Luirink J. The Sec-independent function of *Escherichia coli* YidC is evolutionary-conserved and essential. J Biol Chem. 2005; doi.org/10.1074/jbc.M414094200.
52. Hauer RS, Schlesier R, Heilmann K, Dittmar J, Jakob M, Klosgen RB. Enough is enough: TatA demand during Tat-dependent protein transport. Biochim Biophys Acta. 2013; doi.org/10.1016/j.bbamcr.2013.01.030.
53. Sargent, F. (1998). Overlapping functions of components of a bacterial Sec-independent protein export pathway. The EMBO Journal, 17(13), 3640–3650. https://doi.org/10.1093/emboj/17.13.3640
54. Diderichsen B. flu, a metastable gene controlling surface properties of *Escherichia coli*. J Bacteriol. 1980; 141:858–867.

55. Ulett GC, Valle J, Beloin C, Sherlock O, Ghigo J, Schembri MA. Functional Analysis of Antigen 43 in Uropathogenic *Escherichia coli* Reveals a Role in Long-Term Persistence in the Urinary Tract. Infect Immun. 2007; doi.org/10.1128/IAI.01952-06.
56. Zheng M, Åslund F, Storz G. Activation of the OxyR transcription factor by reversible disulfide bond formation. Science. 1998; doi.org/10.1126/science.279.5357.1718.
57. Dudina O, Geiselmanna J, Ogasawarad H, Ishihamab A, Lacoura S. Repression of Flagellar Genes in Exponential Phase by CsgD and CpxR, Two Crucial Modulators of *Escherichia coli* Biofilm Formation. J Bacteriol. 2014; doi.org/10.1128/JB.00938-13.

Additional files

Supplementary Fig. 1 (.pptx)

The swarming motilities of *E. coli* CyDisCo strains and *P. aeruginosa*

A swarming motility assay was performed using E. coli CyDisCo with an empty plasmid; expressing scFv; expressing misfolded scFv (mf_scFv) and *P. aeruginosa* PA01 as a positive control. Swarming plates (0.3 % agar in LB medium) were spotted with 10 µL of overnight cultures and incubated for 24 h at 37 °C. After 24 h the *E. coli* CyDisCo strains did not show signs of cell swarming, while P. aeruginosa showed a positive swarming phenotype. Assays were performed in triplicate. Figure shows representative results from 24 h incubation.

Supplementary Table 1 A-C (.xlsx)

Average protein abundances and fold changes for proteins from cytoplasmic (A), periplasmic (B) and insoluble/membrane (C) fractions

Supplementary Table 2 (.xlsx)

Voronoi Treemap of log2 fold changes in protein abundance (*E. coli* CyDisCo *vs.* scFv and *E. coli* CyDisCo *vs.* mf_scFv)

Declarations

Authors' contributions

KD and IG designed and performed the experiments, analyzed the data and wrote the manuscript. CR devised the project. RS performed the electron microscopy. TS and DZ performed mass spectrometric measurements and KD analyzed the proteomic results. IG and GM prepared and analyzed qPCR samples. ED helped with the design of the primers.

CR, KR and SS were involved in planning and supervision of the work. All authors provided critical feedback and helped shape the research, analysis and manuscript.

Acknowledgements

We thank Annette Meuche for excellent technical assistance regarding electron microscopy and David Humphreys for constructive criticism of the manuscript.

Ethics approval and consent to participate

Not applicable

Consent for publication

The authors are consent for publication.

Availability of data and material

Please contact corresponding author for data requests.

Competing interests

The authors declare that they have no competing interests.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642836. This work was generously supported by Industrial Biotechnology Catalyst (Innovate UK, BBSRC, EPSRC) grant BB/M018288/1 to support the translation, development and commercialisation of innovative Industrial Biotechnology processes.

2.3 Equal first author publication under review-*Escherichia coli* 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway.

Isabel Guerrero Montero, Kirsty L. Richards, Douglas F. Browning, Amber R. Peswani, Mickael Labrit, Matthew Allen, Cedric Aubry, Emma Davé, David P. Humphreys, Stephen J. W. Busby and Colin Robinson, Escherichia coli 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway. Biotechnology and Bioengineering, under review.

2.3.1 Contribution

For this project, I heavily contributed in concept and experimental design, performed the experiments, analysed the data, reviewed the literature available, made the figures and assisted with the writing of the manuscript. The manuscript was revised by all the authors, especially Colin Robinson.

The initial premise of this project was to compare scaled up protein production and export to periplasm in wild type and TatExpress *E. coli*. To do so, I tested the strains of interest at shake flask with the defined fermentation medium modified from (Smales & James 2005) to ensure that the cells grew to acceptable OD_{600} and continued to produce protein, in this case human growth hormone (hGH), as has been previously demonstrated (Browning et al. 2017). Afterwards, Dr Kirsty Richards and I were able to perform 5 fermentation runs using 2L Sartorious fermenters at UCB under the guidance of their technical team. However, due to timing constraints it was not possible to continue to use the facility at UCB, so 3 Infors 1.5L benchtop bioreactors were purchased and installed. I was responsible for the optimisation of both the bioreactor setup, including modifying the height of the impellers, creating foam traps and the installing a more powerful cooling

system, and the fermentation protocol used for the rest of the runs necessary to acquire the remaining data.

Samples were taken at different time-points and processed to visually measure the amount of protein being produced by by SDS-PAGE gel electrophoresis and immunoblotting. I was also responsible for optimizing the fractionation steps for purification to obtain high volumes of periplasmic material without the use of lysozyme which is regularly used in the standard fractionation protocol developed in the lab (Pierce et al. 1997). Once lysozyme free samples were obtained, I purified the extract using a His-column on the AKTA and then tested both a crude periplasmic sample against the purified to measure protein expression by SDS-PAGE gel electrophoresis and immunoblotting. These samples were also subjected to quantification via a specific hGH bioassay.

Finally, the purified samples alongside a positive control consisting of hGH from a commercial house were taken for mass spectrometry, which was performed by Dr Kevin Howland at the University of Kent, who also aided in the interpretation of the results.

2.3.2 Preliminary data

Due to the specifications of the journal it was decided that not all of the preliminary data generated for the paper was to be added to the final draft. Said data can be found in the following section since it provides the initial proof of concept for better protein export in TatExpress at shake flask level that is later worked upon in the fermentation publication.

Characterisation of TatExpress: TE shows no additional metabolic burden on the cells in comparison with WT.

The aim of this chapter was to characterize the *E. coli* TatExpress W3110 strain and determine its advantages for protein export in comparison with the WT W3110 strain. To do so we first compared the growth profiles of TatExpress producing and exporting hGH, a truncated version of hGH and an empty plasmid with no expressible protein of interest, as well as the profiles for when they were not induced.

The truncated hGH was created from the TorA-hGH-H6 plasmid and has a 30 amino acid deletion between S80 and S109. This deletion is able to change the structure enough so that it is not able to be exported via Tat, as we can see in Figure 6.B. We observe that although the truncation consists of a deletion, the truncated hGH has a higher molecular band than the hGH. This can be explained by the fact that there is no export to the periplasm, meaning that cleavage of the TorA signal peptide would not have taken place. The TorA signal peptide and the deleted fragment have an approximate weight of 4.56 kDa and 3.34 kDa respectively. In the end the combination of TorA and truncated hGH with His-tag has a final weight of 24.31 kDa and the cleaved hGH with His-tag has a final weight of 23.09 kDa, coinciding with the sizes shown on Figure 6.B.

A

MFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLCF

SESIPTPSNREETQQKSNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNV

YDLLKDLEEGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLL

YCFRKDMDKVETFLRIVQCRSVEGSCGF

B

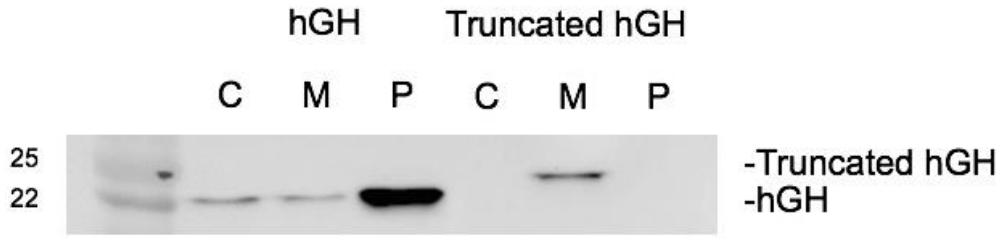


Figure 6: Comparison of hGH and truncated hGH: A: Sequence of original hGH and underline and bold the part of the sequence that was deleted. **B:** Samples 3 hours post-induction and fractionated in C/M/P fractions of the TorA-hGH and TorA-truncated_hGH are shown. Protein was expressed in W3110 TE cells. Molecular weight markers are shown on the left (in kDa) and mature and truncated hGH heights are indicated on the right.

Figure 7 shows the growth curve for *E. coli* TatExpress W3110 expressing the TorA-hGH-H6 construct (red), TorA-truncated hGH-H6 construct (green) and an identical plasmid as the ones mentioned but with no protein of interest expressed both in TatExpress (blue) and WT (black). Data for the strains bearing the plasmid but not induced were also obtained (dashed line). We observe that the strains producing protein show an important arrest in growth in comparison with the un induced samples due to resources being diverted to making protein, however, the strain with a plasmid that did not produce a protein of interest had an identical growth in comparison with the same but un induced. This is striking since even though the protein of interest is not being made the TatExpress strain is overproducing the tat components from the TatABCD operon. It appears that this production of additional tat components does not have an impact on the

growth and general wellbeing of the cells, as it not only does it show no growth difference in comparison to the un induced strain, it also has an identical growth pattern to un induced *E. coli* W3110 WT cells.

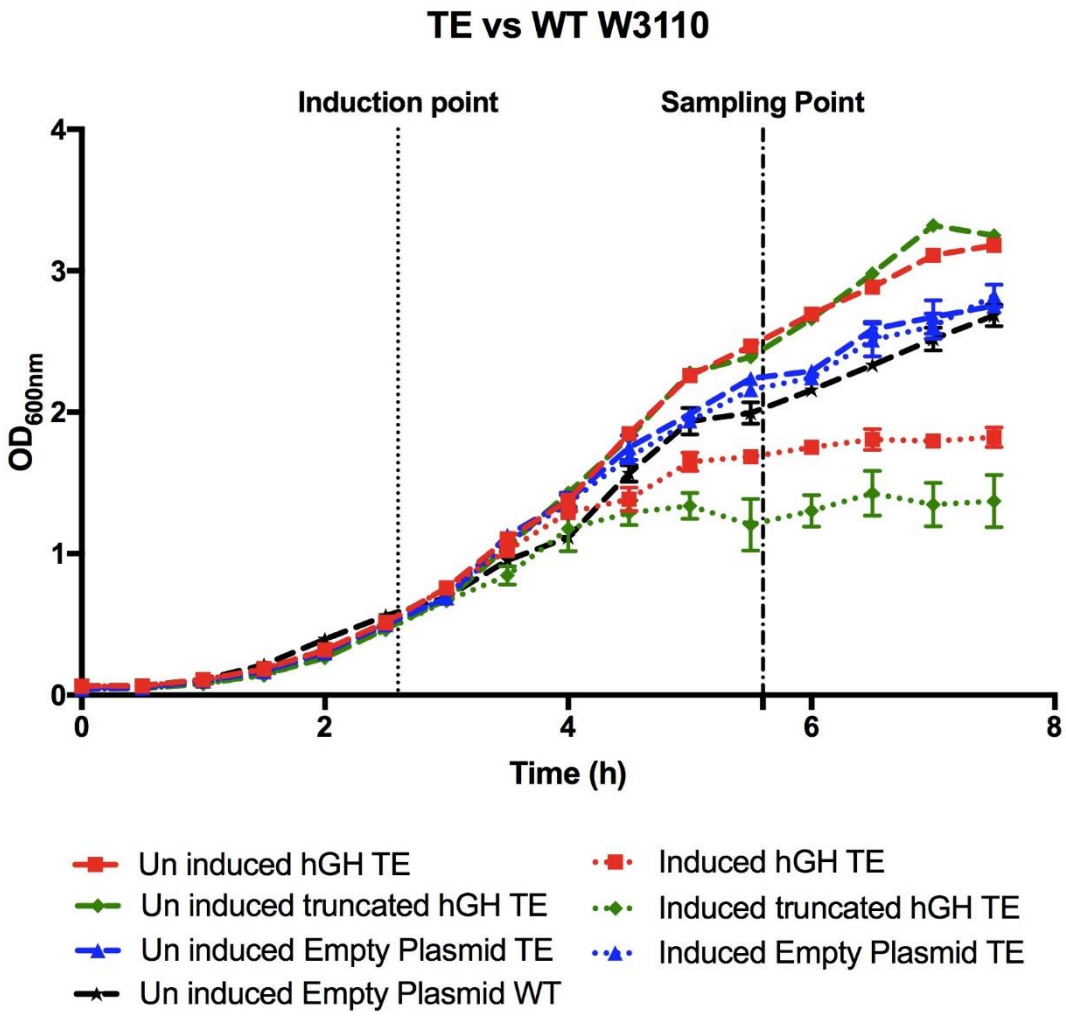


Figure 7. **Growth comparison of *E. coli* TatExpress W3110 induced and un induced strains expressing folded and truncated hGH and no protein of interest (POI) as well as a un induced *E. coli* W3110 as a control.** Cultivations were performed in LB at 30 °C and 200 rpm for all conditions. Average growth and standard deviation of at least three independent experiments is shown. The perpendicular dashed and dotted lines show the time of IPTG induction and sampling point (3 h post-induction), respectively.

Pre-scale up evaluation of TatExpress in SM6 media.

As has already been reported the TatExpress strain manages to export a significantly larger amount of protein to the periplasm than WT at shake flask level under laboratory conditions and does not seem to have any detrimental effects to cell viability and growth. However, to scale up the cultures it was necessary to assess what the growth and expression would be like in the minimal defined media used in fermentation. The protocol and recipe for shake flask expression and the media SM6c used for scale up was provided by Dr Emma Davé at UCB. SM6c uses glycerol as its sole carbon source and allows for slower growth over a longer period of time without affecting cell viability, letting us obtain higher ODs. During this period, a new TatExpress BL21 strain was developed and comparative trials in SM6c media took place.

To prevent growth arrest due to lack of oxygen and high culture concentration (maximum obtained was OD of 8.57 after 40 hours) the growing conditions were modified from the usual protocol: cells were grown at 30°C at 250 rpm until OD of 2 and then induced. Temperature was then lowered to 25°C and samples taken for analysis. Figure 8 shows a comparison of the levels of expression during different time points between BL21 and W3110 TatExpress. We observe that while both strains are producing and exporting protein to the periplasm, and a high percentage of the protein is being accumulated in the cytoplasmic and insoluble fraction, in the case of the BL21 strain we also observe accumulation of the precursor hGH: the TorA signal peptide has not been cleaved during the processing. This does not occur with the W3110 strain, where we see that all the protein is immediately processed for export.

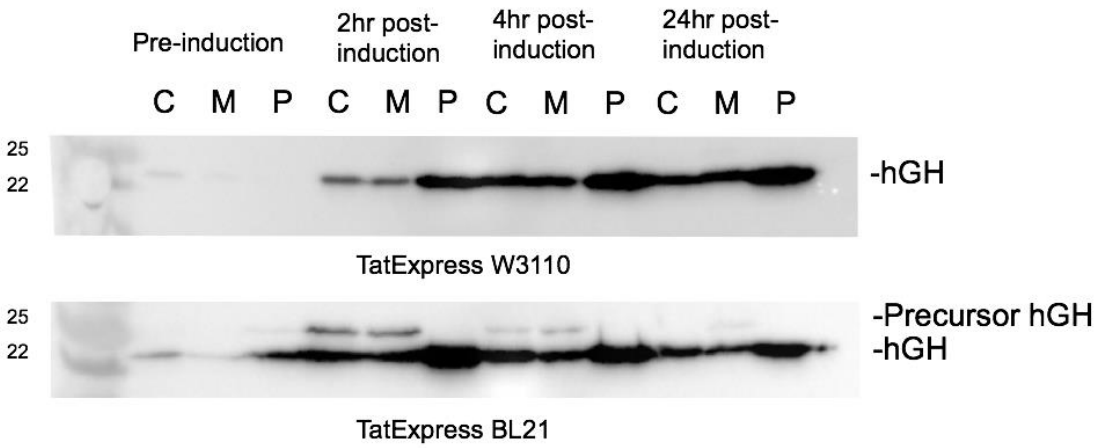


Figure 8. **Comparison of hGH produced in TatExpress W3110 and BL21 in SM6c media.** Samples taken at pre-induction and 2, 4 and 24 hours post-induction and fractionated in C/M/P fractions of the TorA-hGH are shown. Molecular weight markers are shown on the left (in kDa) and mature and precursor hGH heights are indicated on the right.

To accurately compare how well the TatExpress strain were exporting protein in comparison to the wild type the periplasmic bands on the blots for multiple repeats and serial dilutions were subject to densitometry. Figure 9 shows that for the same dilution (1:8) there is a band with a higher intensity at both 2 and 4 hours post induction for the TatExpress BL21 strain when compared to the WT. We determined that at 4 hours post induction the BL21 TatExpress strain has produced approximately 4.77 times more protein in the periplasm than BL21 WT at the same time point.

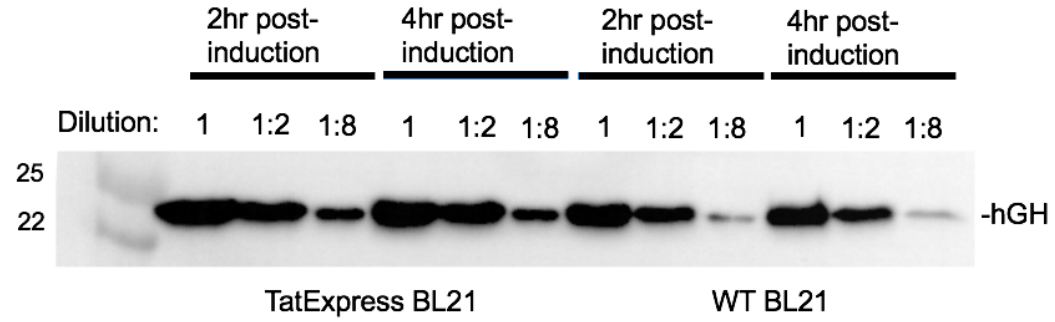


Figure 9. **BL21 TatExpress exports more protein to the periplasm than BL21 WT.** Periplasmic samples of TatExpress BL21 and WT BL21 expressing TorA-hGH 2 and 4 hours post induction grown in SM6c media. Lanes show no dilution, 1:2 and 1:8 dilution for each sample. Molecular weight markers are shown on the left (in kDa) and mature hGH height is indicated on the right.

2.3.3 Publication

The paper was published after this work was submitted and the published version can be found in annex 1.

The paper as was sent for publication in Biotechnology & Bioengineering is as follows:

Escherichia coli 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway

Isabel Guerrero Montero[1†], Kirsty L. Richards[1†], Douglas F. Browning[2], Amber R. Peswani[1], Mickael Labrit[3], Matthew Allen[3], Cedric Aubry[3], Emma Davé[3], David P. Humphreys[3], Stephen J. W. Busby[2] and Colin Robinson[1*]

[1] School of Biosciences, University of Kent, Ingram Building, Canterbury CT2 7NJ, UK

[2] Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK

[3] Discovery Biology, UCB Celltech, 216 Bath Road, Slough SL1 4EN, UK

[†] These authors contributed equally to this work.

* For correspondence:

Colin Robinson : Tel: +44 (0)1227- 823443. Email: C.Robinson-504@kent.ac.uk

2.3 *Escherichia coli* 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway

Abstract

Escherichia coli is a heavily used platform for the production of biotherapeutic and other high-value proteins, and a favoured strategy is to export the protein of interest to the periplasm in order to simplify downstream processing and facilitate disulphide bond formation. The Sec pathway is the standard means of transporting the target protein but it is unable to transport complex or rapidly folding proteins because the Sec system can only transport proteins in an unfolded state. The Tat system also operates to transport proteins to the periplasm, and it has significant potential as an alternative means of recombinant protein production because it transports fully folded proteins. Here, we have tested the Tat system's full potential for the production of biotherapeutics for the first time using fed-batch fermentation. We expressed human growth hormone (hGH) with a Tat signal peptide in *E. coli* W3110 'TatExpress' strains that contain elevated levels of the Tat apparatus. We show that the protein is efficiently exported to the periplasm during extended fed-batch fermentation, to the extent that it is by far the most abundant protein in the periplasm. The protein was shown to be homogeneous, disulphide bonded and active. The bioassay showed that the yields of purified periplasmic hGH are 5.4 g/L culture.

1.0 Introduction

Many high-value proteins are produced in *E. coli,* and a favoured strategy is to export the protein to the periplasm, usually by the well-characterised 'Sec' pathway (Walsh, 2014). Several strategies have been used, including expression of soluble proteins in the cytoplasm, expression in the form of insoluble inclusion bodies or export to the periplasm.

The latter approach is particularly favoured for the production of proteins that contain disulphide bonds, since the periplasm is the only oxidising compartment in wild type cells (Pooley *et al.*, 1996). Downstream processing is also simplified due to the lowered levels of cytoplasmic contaminants, including DNA (Balasundaram *et al.* (2009). Export of proteins to the periplasm is usually achieved by targeting via the well-characterised Sec pathway, whereby a cleavable Sec-specific signal peptide is present at the N-terminus of the target protein. However, the Sec pathway cannot handle many proteins and attention has focused on the alternative Tat pathway, which transports fully folded proteins by a completely different pathway (reviewed by Natale *et al.*, 2008). Initial studies showed that Tat could export a model protein, GFP, at high levels with no large-scale release of cytoplasmic contents (Matos *et al.*, 2012). This work was followed by additional studies that showed Tat to be capable of exporting a range of biotherapeutics including human growth hormone, single chain antibody fragments and interferons (DeLisa *et al.*, 2003; Alanen *et al.*, 2015; Tullman-Ercek *et al.*, 2007; Browning *et al.*, 2017; Matos *et al.*, 2014).

Human Growth Hormone (hGH) is a 22 kDa protein consisting of 191 amino acids (Pooley *et al.*, 1996) used to treat burn injuries, wound healing, hipopituitarism and obesity (Isaksson *et al.*, 1985) since it is safe to use even in high doses (Van, 1998). Since hGH only has 2 disulphide bonds (Cys53-Cys165 and Cys182-Cys189) (Ultsch *et al.*, 1994) and no glycosylation, most of its large-scale production has used *E. coli* platforms, although a wide range of eukaryotic host systems have been tested (e.g. Hahm and Chung, 2001; Ecamilla-Trevino *et al.*, 2000). The highest recorded yield is 5 g of recombinant hGH per litre of milk in transgenic cows (Salamone *et al.*, 2006).

One of the major difficulties in producing hGH in *E. coli* is its aggregation into insoluble bodies that complicate the production process. Until recently, most commercial hGH has been obtained from insoluble bodies that have been solubilized, with the hGH refolded into its active form (Olsen *et al.*, 1981; Patra *et al.*, 2000). Recent, novel approaches have managed to produce high quantities of soluble hGH in fed-batch fermentation, with reported yields of up to 678 mg/L cytosolic (Song *et al.*, 2017).

Although studies exporting hGH to the periplasm have reported relatively low yields (Sockolosky and Szoka, 2013), this is an attractive option due to the ease of downstream processing. In recent studies, we have shown that the Tat system can export hGH with high efficiency and that it is disulphide-bonded (Alanen *et al.*, 2015). Their study concluded that, although the protein was presumably synthesised and exported in the reduced state, the periplasmic protein was fully disulphide-bonded by the normal Dsb machinery. We have also reported that a new series of engineered *E. coli* strains, which overexpress the Tat system components (termed TatExpress), demonstrate even higher export efficiencies (Browning *et al.* 2017). These strains bear a modified chromosomal *tatABC* operon, the expression of which is induced by IPTG, so that the Tat system is over-expressed at the same time as the IPTG-induced target protein. However, the activity of the hGH was not assayed and our studies used laboratory shake flask systems that bear little resemblance to industrial production processes. Here, we have assessed the feasibility of this system for industrial production using TatExpress *E. coli* strains under fed-batch fermentation conditions. We report that the system is able to produce 5 g/L of purified, disulphide bonded, active hGH.

2.0 Materials and Methods

2.3 *Escherichia coli* 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway

Bacterial strains and plasmids

The *E. coli* strain utilized for this study was W3110 WT and W3110 TatExpress which was previously described in Browning *et al.,* 2017 alongside the plasmid pKRK7 (pEXT22 expressing TorA-hGH-6His). pKRK38 was derived from pKRK7 to eliminate the TorA signal peptide that precedes the hGH. pKRK38 was constructed by PCR amplification using pKRK7 as a template and primers Mature_hGH_F (5'-ATGTTCCCAACCATTCCCTTATCCA-3') and Mature_hGH_R (5'-CATACATGTTCCTCTGTGGTAGGGT-3') as forward and reverse primers. PCR solutions were made using 1 µL template DNA (100 ng/µL), 1 µL dNTPs (200 µM final), 10 µL 5 × GC buffer, 1.25 µL of each primer (0.5 µM final), 0.5 µL Phusion High-fidelity DNA polymerase (2 U/µL stock, New England Biolabs, UK) and made up to a final volume of 50 µL with milliQ H_2O. PCR was then carried out in Biometra T3 Thermocycler (Biometra Anachem, UK) as per Phusion polymerase instructions with an annealing temperature of primer T_m minus 5 °C and annealing time extended to 10 min. PCR product was then treated with 1 µL DpnI restriction enzyme (New England Biolabs, UK) for 1 hour at 37 °C to digest template DNA and then heat inactivated for 20 min at 80 °C. 2 µL of the product were mixed with 1 µL of ligase buffer, 1 µL of T4 DNA ligase (Roche) and made up to a final volume of 10 µL with milliQ H_2O before incubation overnight at 4 °C. The next day 5 µL ligation product was used to transform 100 µL *E. coli* W3110 competent cells. Final plasmid sequence was confirmed by GATC-Biotech before use in this study.

Fed-batch Fermentation

Starter cultures were grown in 50 mL of 2xP media in 250 mL shake flasks for 6 h at 37°C, 200 rpm with 1:1000 antibiotic (5 µL, 1 M kanamycin). 1 mL of culture was transferred to 200 mL of SM6 defined media and grown aerobically overnight at 30 °C,

200 rpm in shake flasks with 1:1000 antibiotic (5 µL, 1 M kanamycin). The next day an equivalent of 300 OD_{600} was used to inoculate fresh defined SM6 media to a final volume of 500 mL in Infors Multifors 1.5L fermenters (Infors UK Ltd., Reigate, UK). The pH was maintained at 7 using 25% (v/v) ammonia solution and 25% (v/v) sulphuric acid. Dissolved oxygen tension (DOT) was maintained at 30% using gas blending with 100% oxygen where necessary and the culture was maintained at 30°C until both stirrer and airflow was maximal and then dropped to 25°C. Cell density was measured by optical density in the range of 0.1-0.5 at 600 nm after dilution. Supplementation of $MgSO_4$ occurred when the OD_{600} reached 38-42 (8 mL/L of 1M $MgSO_4$) and of Na_2HPO_4 when the OD_{600} reached 54-58 (5 mL/L of 232.8 g/L Na_2HPO_4) and when the OD_{600} reached 66-77 (7 mL/L of 232.8 g/L Na_2HPO_4). A glycerol feed containing 80% w/w glycerol at a rate of 0.35% pump capacity was started at OD_{600} 70. Induction with 9 ml/L of IPTG at a concentration of 4.31 g/L occurred at OD_{600} 75.

Fractionation

Cells equivalent to a density of OD_{600} of 10 were taken and fractioned into Cytoplasm (C), Membrane (M) and Periplasm (P) fractions by the EDTA/lysozyme/ cold osmotic-shock method. In brief, cells were pelleted by spinning at 3000 rpm, 4 °C, 10 min. The supernatant was discarded and the cell pellet resuspended in 500 µL of Buffer 1 (100 mM Tris-acetate pH 8.2, 500 mM sucrose, 5 mM EDTA pH 8.0) and 500 µL dH_2O before addition of 40 µL hen egg white lysozyme (1 mg/mL) and incubation on ice for 5 min to digest the outer membrane. 20 µL $MgSO_4$ (1M) was added to stabilize the inner membrane before centrifugation at 14,000 rpm, 4 ° C, 2 min. 500 µL of the supernatant was taken and frozen at − 20 °C as the Periplasmic fraction, P. The remaining supernatant was discarded and the pellet washed by resuspension in 750 µL Buffer 2 (50 mM Tris-acetate pH 8.2, 250 mM sucrose, 10 mM $MgSO_4$) and subsequently spun at 14,000 rpm,

4 °C, 5 min. The cell pellet was resuspended in 750 µL Buffer 3 (50 mM Tris-acetate pH 8.2, 2.5 mM EDTA pH 8.0) and sonicated for 6 × 10 s, amplitude 8 µm to disrupt membranes (Soniprep 150plus, Sanyo Gallenkamp, Loughborough, UK). The resulting solution was then centrifuged at 70,000 rpm, 4 °C, 30 min to sediment the insoluble fraction. 500 µL of the supernatant was then taken and frozen at − 20 °C and designated as the Cytoplasmic fraction, C. The remainder of the supernatant was discarded and the pellet resuspended in 500 µL Buffer 3 before freezing at − 20 °C, and designated as the membrane or insoluble fraction, M.

Protein purification

For purification of 6x Histidine-tagged (C-term) proteins by Nickel IMAC, 10 mL of the culture post induction was taken centrifuged at 3000 rpm, 45 min, 4°C. Cell pellet was resuspended in 10 mL/g of chilled Buffer 1 without EDTA and 10 mL/g of milliQ H_2O and incubated on ice for 30 min before centrifuging at 14 000 rpm, 20 min, 4°C (Beckman Avanti J- 25, JA 25.5 rotor). Supernatant was taken as the periplasmic fraction and placed into SnakeSkin® dialysis tubing (Thermo scientific) and dialysed at 4°C overnight into 50 mM sodium phosphate, 150 mM sodium chloride. Using an ÄKTA™ pure protein purification system and a HisTrap HP histidine-tagged protein column (GE Healthcare, Buckinghamshire, UK) the protein was purified: storage solution (20% ethanol) was washed off with 10 column volumes (CV) of milliQ H_2O before adding 5 mL 0.2 M NiCl, followed by another 2CV milliQ H_2O wash. Columns were equilibrated with 3CV equilibration buffer (50 mM sodium phosphate pH 7.2, 0.3 M NaCl) before loading Periplasmic sample and collecting Flow Through (FT). Unbound matter was removed with 6CV Wash buffer (50 mM sodium phosphate pH 7.2, 50 mM Imidazole, 0.3 M sodium chloride) and sample collected as Wash (W). Finally, the 6x Histidine-tagged

protein was eluted with elution buffer with imidazole (50 mM sodium phosphate pH 7.2, 0.3 M NaCl with 250 mM Imidazole) and the peaks collected for analysis.

Protein expression analysis by Coomassie and Western blot

Protein samples were resolved by reducing SDS–PAGE and analyzed using Coomassie blue staining and Western blotting. Protein was transferred to a PVDF-membrane (GE Healthcare, Buckinghamshire, UK) by wet-western-blotting, with electrophoresis for 1 h at 80 V, 300 A. The PVDF membrane was then immersed in blocking solution (2.5% (w/v) skimmed milk powder in 50 mL 1 × PBS-Tween20 (0.1%)) and incubated at 4 °C overnight. The following day, the membrane was washed 3 × for 5 min with 1 × PBS-Tween20 (0.1%) before incubation with primary antibody (3.5 µL anti-6-His (Life Technologies, CA, USA) in 20 mL 1 × PBS-Tween20 (0.1%), 3% (w/v) BSA) for 1 h at room temperature. Membranes were then washed 3 × for 5 min with 1 × PBS-Tween20 (0.1%) before incubation with secondary antibody (4 µL anti-Mouse HRP conjugate (Promega, WI, USA) in 20 mL 1 × PBS-Tween20 (0.1%)) for 1 h at room temperature. Finally, membranes were washed for 3 × for 5 min with 1 × PBS-Tween20 (0.1%). Immunoreactive bands were detected using an enhanced chemiluminescence (ECL) kit (BioRad, Herts, UK) following the manufacturer's instructions. Bands were visualised using BioRad Gel-doc chemiluminescence imager and associated software. Comparative densitometry of band-intensities was also carried out on a BioRad Gel-doc imager with an Image Lab™ Software 4.1. Periplasmic hGH was assayed using an hGH ELISA kit (Roche Diagnostics, Charles Ave, Burgess Hill, West Sussex RH15 9RY) and an hGH bioassay (PathHunter® Human Growth Hormone Bioassay Kit, Sigma) as per the manufacturer's protocol. Standardized OD10 periplasmic fractions were diluted 1:500,000 using PBS and absorbance was read using a BMG Labtech Spectrostar

microplate reader at 405 nm, with a reference wavelength at 490 nm. Concentrations were calculated from two independent experiments and all samples measured in triplicate, and used to calculate average periplasmic yield (in mg/L) by factoring in culture OD readings at 600 nm.

Intact Protein Electrospray LC-MS

The electrospray mass spectrum was recorded on a Bruker micrOTOF-Q II mass spectrometer. An aliquot of protein in solution, corresponding to approximately 20 picomoles of protein, was desalted on-line by reverse-phase HPLC on a Phenomenex Jupiter C4 column (5 μm, 300Å, 2.0 mm x 50 mm) running on an Agilent 1100 HPLC system at a flow rate of 0.2 ml/min using a short water, acetonitrile, 0.05% trifluoroacetic acid gradient. The eluent was monitored at 280 nm and then directed into the electrospray source, operating in positive ion mode, at 4.5 kV and mass spectra recorded from 500-3000 m/z. Data was analysed and deconvoluted to give uncharged protein masses with Bruker's Compass Data Analysis software.

3.0 Results

3.1 Fed-batch fermentation of WT and TatExpress cells expressing TorA-hGH

We expressed a construct comprising the TorA signal peptide linked to hGH (TorA-hGH) in W3110 *E. coli* TatExpress cells under fed-batch fermentation conditions as detailed in Materials and Methods. The construct was previously used in shake flask studies (Browning *et al.,* 2017) and comprises the signal peptide of *E. coli* TMAO reductase (TorA) linked to hGH via a 5-residue linker (the first 5 residues of mature TorA). The

construct also contain a C-terminal 6-His tag. Parallel cultures were carried out with the construct expressed in wild type (WT) cells. Figure 1 shows the growth curves from the cultures; duplicate WT and TatExpress cultures reached ODs of about 100 and the Figure shows the point at which induction of TorA-hGH and TatABC synthesis (in TatExpress) was induced using IPTG.

Samples were removed 13, 17 and 21h after induction and the cells were fractionated to generate cytoplasmic, membrane and periplasmic samples. The periplasmic samples were analysed by immunoblotting to detect the hGH and the same samples were analysed using Coomassie stained SDS PAGE gels to analyse the total proteome of the periplasmic samples. Figure 2 shows the 13h and 21h samples from the two WT cultures, which indicate the presence of a clear hGH signal in the periplasm. The Coomassie-stained gel shows the presence of a 22 kDa protein, the relative abundance of which matches the strength of the blot signals. This was confirmed to be hGH (see below). 13h, 17h and 21h samples from the two TatExpress cultures were also analysed and the data show that the hGH blot signals are significantly higher than those of the WT cultures, confirming that TatExpress cells export this construct with much higher efficiency as observed by Browning *et al.* in shake flask culture (2017). The abundance of the 22 Da protein is correspondingly greater in the Coomassie stained gels, providing further evidence that this is indeed hGH.

Figure 3 shows a time course analysis of the abundance of the periplasmic hGH after induction of synthesis in TatExpress cells. Samples were taken from 2 - 51 h after induction and the blot shows a steady increase in hGH level over this time period. Equal numbers of cells were used in each sample, so this reflects an increase in the amount of

hGH per cell. The Coomassie gel confirms that the abundant 22 kDa protein is indeed hGH, since its abundance increases in parallel and the protein is virtually absent in the absence of induction. Analysis of the later time points demonstrates that hGH is by far the most abundant periplasmic protein, which shows that Tat is exporting high amounts of protein.

We have carried out controls for fractionation artefacts in previous studies, and the periplasmic proteome is clearly distinct from that of the cytoplasm, suggesting that contamination of cytoplasmic proteins is minimal. However, to confirm that hGH does not reach the periplasmic fraction 'spontaneously' we expressed hGH lacking any form of signal peptide. Mature hGH was expressed in parallel with Tor-hGH and samples were fractionated after 1h and 18 h induction (Figure 4). The immunoblot shows that the bulk of hGH is in the periplasm in the TorA-hGH culture, as expected, with a minimal level of protein present in the cytoplasm at the 18h point. In contrast, hGH is found exclusively in the cytoplasm in the sample expressing mature hGH, with none detected in the periplasmic fraction. These data confirm that the presence of hGH in the periplasm of the TorA-hGH cultures is due to Tat-dependent export and not contamination by cytoplasmic fraction or spontaneous transfer across the plasma membrane.

It is notable that the total level of hGH in the culture producing mature hGH is far lower than that in the TorA-hGH culture. This reflects a phenomenon observed previously in our shake flask studies: that hGH is subjected to rapid turnover in the cytoplasm (Alanen *et al.,* 2015)..

3.2 Tat-exported hGH is homogeneous and cleaved at the correct signal peptidase site

To assess the homogeneity of the exported hGH we performed mass spectrometry analysis of the purified protein, using purified commercial hGH as a standard. The hGH construct used in this study should contain additional amino acids when compared to commercial hGH, specifically the 5 amino acid linker from the mature TorA protein (TMAO reductase) at the N-terminus, the initiation methionine (cleaved from hGH *in vivo*) and the 6-His tag at the C-terminus (Figure 5). The combined molecular weight of the additional amino acids is 609.7 for the N-terminal residues and 822.86 for the C-terminal His tag (1432.56 in total). The predicted molecular weights are 22129.05 for the commercial hGH, and 23561.5 for the exported hGH. As shown in Figure 5, mass spectrometry analysis of commercial hGH gives a single prominent peak corresponding to 22124.3 Da, which matches almost exactly with the predicted molecular mass, having only a difference of 4.75 Da, 4 of which are a consequence of the formation of 2 disulphide bonds. We purified the Tat-exported periplasmic hGH using affinity chromatography and an analysis of this protein again shows a single prominent peak, and the mass of 23555 Da is again very close to, but slightly smaller than the predicted mass of 23561 Da. The presence of 2 disulphide bonds again accounts for the difference. In both cases, the 1-2 Da deviation from the predicted protein masses is within the error range for the mass spectrometer.

Other smaller peaks are observed in both samples, which appear to be an artefact of the hGH mass spectrometry since they are identical in the two samples. This would imply that the hGH produced in this study is (i) as homogeneous as the commercial protein (ii)

processed at the correct site, and (iii) fully disulphide bonded; there is no indication of a peak corresponding to the reduced form.

3.3 High yields of active protein are obtained from fed-batch fermentation using TorA-hGH

To quantify the yield of periplasmic hGH and to assess its activity, we used a commercially available hGH bioassay (PathHunter® Human Growth Hormone Bioassay Kit, Sigma). The bioassay uses engineered cells in which one fragment of β-galactosidase is present on the hGH receptor and the complementary fragment is present on a phosphor-tyrosine SH2-domain-containing protein that is only able to bind the hGH receptor once it is activated. Binding of hGH to its receptor results in receptor phosphorylation by a cytosolic tyrosine kinase such as JAK1, enabling the SH2-EA fusion protein to bind the phosphorylated receptor and generate active β-galactosidase. Enzymatic activity is quantitatively measured using a chemiluminescent substrate, and the expected result is a dose-response sigmoidal function where the phosphorylation of the receptor will be dependent on the amount of hGH in the sample. A standard curve was produced with the supplied commercial hGH and all samples were done in triplicate and with dilution factors in the range of the standard curve (see Supplementary Figure 2). This assay gave a figure of 5.4 g purified hGH per litre of fed-batch culture (see Table 2).

4.0 Discussion

2.3 *Escherichia coli* 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway

The Tat system has been proposed to offer a viable alternative to the Sec pathway for the export of high-value proteins to the bacterial periplasm, but its potential for the production of biotherapeutic proteins has not been fully explored in previous studies. This is primarily because the vast majority of those studies were carried out using laboratory shake flask culture systems, whereas industrial processes almost invariably use fed-batch fermentation. Here, we have used fed-batch fermentation systems to test the robustness of the TatExpress cells and the yields of hGH that are obtained after export to the periplasm.

It was important to assess the TatExpress strains in fermentation systems because they rely on the simultaneous expression of the target protein (in this case TorA-hGH) and 3 different membrane proteins (TatABC). While the TatABC proteins are expressed to a much lower extent than TorA-hGH, the increased expression of any membrane protein can potentially lead to cell stress and lowered productivity in extended fed-batch fermentation systems. Here, we directly compared the growth characteristics of the TatExpress strain with its parental W3110 strain and we observed no significant differences. Clearly, TatExpress strains are viable production hosts.

In terms of target protein yield, we consistently observe that the cells export TorA-hGH throughout an extended induction period and it is notable that the exported hGH is by far the most abundant periplasmic protein by the end of the induction period. In control tests, mature-size hGH is not exported at all and we can conclude that TorA-hGH is being exported by the TorA signal peptide system as reported in shake-flask studies (Alanen *et al.,* 2015). The yields of protein are high: the bioassay indicates a yield of over 5 g/L active purified protein.

In summary, we have shown that a model biotherapeutic protein can be exported by the Tat system in high amounts, and that it is homogeneous, disulphide-bonded and active. It will be of interest to conduct further studies to assess the full capability of the Tat system, and in particular to explore its potential for the export of more complex proteins.

ACKNOWLEDGEMENTS

This work was generously supported by Industrial Biotechnology Catalyst (Innovate UK, BBSRC, EPSRC) grant BB/M018288/1 to support the translation, development and commercialisation of innovative Industrial Biotechnology processes. We are grateful to Richard Davies and Geoff Brown for fermentation expertise (USP development 1, Biotech Sciences, UCB Celltech).

REFERENCES

Alanen HI, Walker KL, Lourdes Velez Suberbie M, Matos CF, Bonisch S, Freedman RB, Keshavarz-Moore E, Ruddock LW, Robinson C. 2015. Efficient export of human growth hormone, interferon alpha2b and antibody fragments to the periplasm by the *Escherichia coli* Tat pathway in the absence of prior disulfide bond formation. Biochim Biophys Acta 1853(3):756-63.

Balasundaram B, Harrison S, Bracewell DG. 2009. Advances in product release strategies and impact on bioprocess design. Trends Biotechnol 27(8):477-85.

Browning DF, Richards KL, Peswani AR, Roobol J, Busby SJW, Robinson C. 2017. *Escherichia coli* "TatExpress" strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway. Biotechnol Bioeng. 114, 2828-2836.

DeLisa MP, Tullman D, Georgiou G. 2003. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. Proc Natl Acad Sci U S A 100(10):6115-20.

Ecamilla-Trevino LL, Viader-Salvado JM, Barrera-Saldana HA, Guerrero- Olazaran M. 2000. Biosynthesis and secretion of recombinant human growth hormone in Pichia pastoris,

Biotechnol. Lett. 22; 109–114.

Hahm MS, Chung BH. 2001. Secretory expression of human growth hormone in Saccharomyces cerevisiae using three different leader sequences, Biotechnol. Bioprocess Eng. 6; 306–309.

Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, Ohtsubo E, Baba T, Wanner BL, Mori H and others. 2006. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Mol Syst Biol 2:2006.0007.

Matos CF, Branston SD, Albiniak A, Dhanoya A, Freedman RB, Keshavarz-Moore E, Robinson C. 2012. High-yield export of a native heterologous protein to the periplasm by the tat translocation pathway in *Escherichia coli*. Biotechnol Bioeng 109(10):2533-42.

Matos CF, Robinson C, Alanen HI, Prus P, Uchida Y, Ruddock LW, Freedman RB, Keshavarz-Moore E. 2014. Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains. Biotechnol Prog 30(2):281-90.

Natale P, Bruser T, Driessen AJ. 2008. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms. Biochim Biophys Acta 1778(9):1735-56.

Olson KC, Fenno J, Lin N, Harkins RN, Snider C, Kohr WH, Ross MJ, Fodge D, Prender G, St ebbing N. 1981. Purified human growth hormone from *E. coli* is biologically active. Nature 293:408-11.

Patra AK, Mukhopadhyay R, Mukhija R, Krishnan A, Garg LC, Panda AK. 2000. Optimization of inclusion body solubilization and renaturation of recombinant human growth hormone from *Escherichia coli*. Protein Exp Purification 18:182-92;

Pooley H. M., Merchante R., Karamata, D. 1996. Overall Protein Content and Induced Enzyme Components of the Periplasm of Bacillus subtilis. Microbial Drug Resistance, 2 (1): 9-15.

Salamone D, Baranao L, Santos C, Bussmann L, Artuso J, Werning C, Prync A, Carbonetto C, Dabsys S, Munar C, Salaberry R, Berra G, Berra I, Fernandez N, Papouchado M, Foti M, Judewicz N, Mujica I, Munoz, L, Alvarez, SF. Gonzalez E, Zimmermann J, Criscuolo M, Melo C. 2006. High level expression of bioactive recombinant human growth hormone in the milk of a cloned transgenic cow, J. Biotechnol. 124; 469–472.

Sambrook J, Russell DW. 2001. Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Sockolosky JT, Szoka FC. 2013. Periplasmic production via the pET expression system of soluble, bioactive human growth hormone. Protein Expression and Purification, 87(2), 129–135

Song H, Jiang J, Wang X, Zhang J. 2017. High purity recombinant human growth hormone (rhGH) expression in *Escherichia coli* under *phoA* promoter, Bioengineered, 8:2, 147-153

Tullman-Ercek D, DeLisa MP, Kawarasaki Y, Iranpour P, Ribnicky B, Palmer T, Georgiou G. 2007. Export pathway selectivity of *Escherichia coli* twin arginine translocation signal peptides. J Biol Chem 282(11):8309-16.

Ultsch MH. Somers AA, Kossiakoff AM, Devos AM. 1994. The crystal-structure of affinity-matured human growth-hormone at 2-Angstrom resolution, J. Mol. Biol. 236; 286–299.

Walsh G. 2014. Biopharmaceutical benchmarks 2014. Nat Biotechnol 32(10):992-1000.

2.3 *Escherichia coli* 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway

Table Legends

Table I. Strains and plasmids used in this work.

Strains/Plasmids	Description	Source/Reference
W3110	*E. coli* K-12 strain. F^- λ^- IN(*rrnD-rrnE*)1 *rph*-1	(Hayashi et al., 2006)
TatExpress W3110 Km	W3110 carrying a kanamycin resistance cassette and the *ptac* promoter upstream of *tatABCD*	This work
pKRK7	pEXT22 expressing TorA-hGH-6His	This work
pKRK36	pEXT22 expressing hGH-6His	This work

Table 2

Assay	Purified (g/L)	Periplasmic extract (g/L)
hGH Bioassay	**5.4**	**10.01**

Table 2: Concentration of hGH in the periplasmic sample and after purification. Table showing concentration of hGH present in the fermenter calculated from the purified fraction and from the periplasmic fraction for the TorA-hGH culture, using an hGH bioassay.

2.3 *Escherichia coli* 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway

Figure Legends

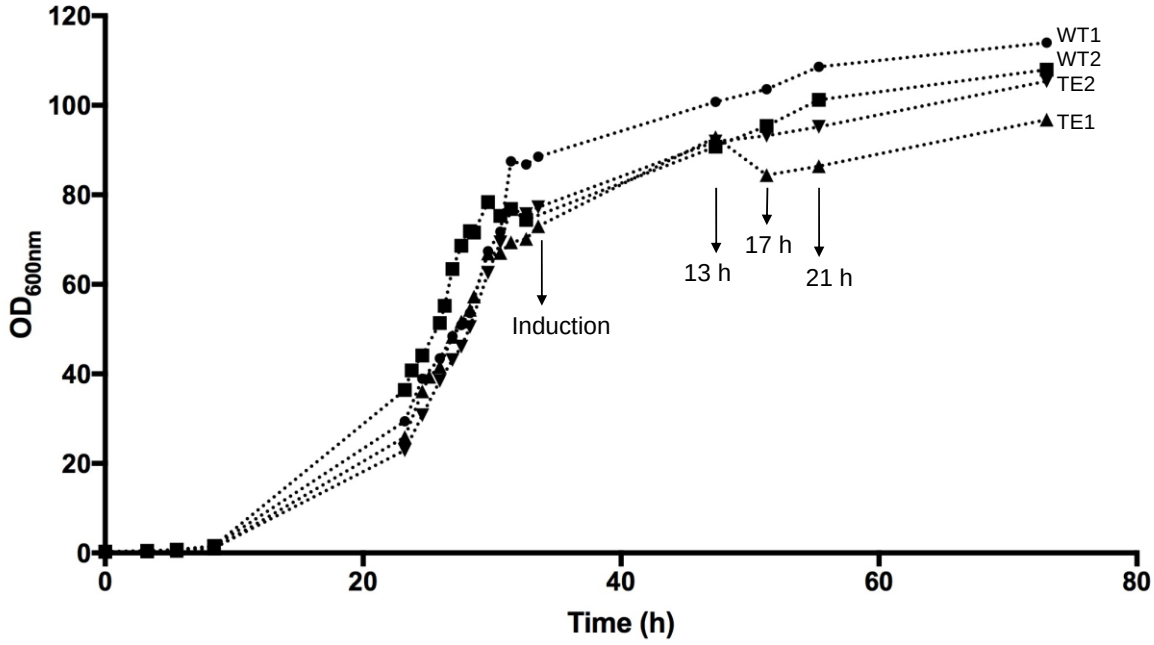


Figure 1 Growth data during fed-batch fermentation of *E. coli* W3110 WT and TatExpress cells expressing TorA-hGH. Duplicate cultures of WT cells (WT1, WT2) and TatExpress (TE1, TE2) were analysed, with $OD_{600\ nm}$ values shown. Fermentation was carried out at 30°C as detailed in Materials and Methods. At the indicated times, the cultures were induced with IPTG (0.1 mM) and samples were removed for analysis.

2.3 *Escherichia coli* 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway

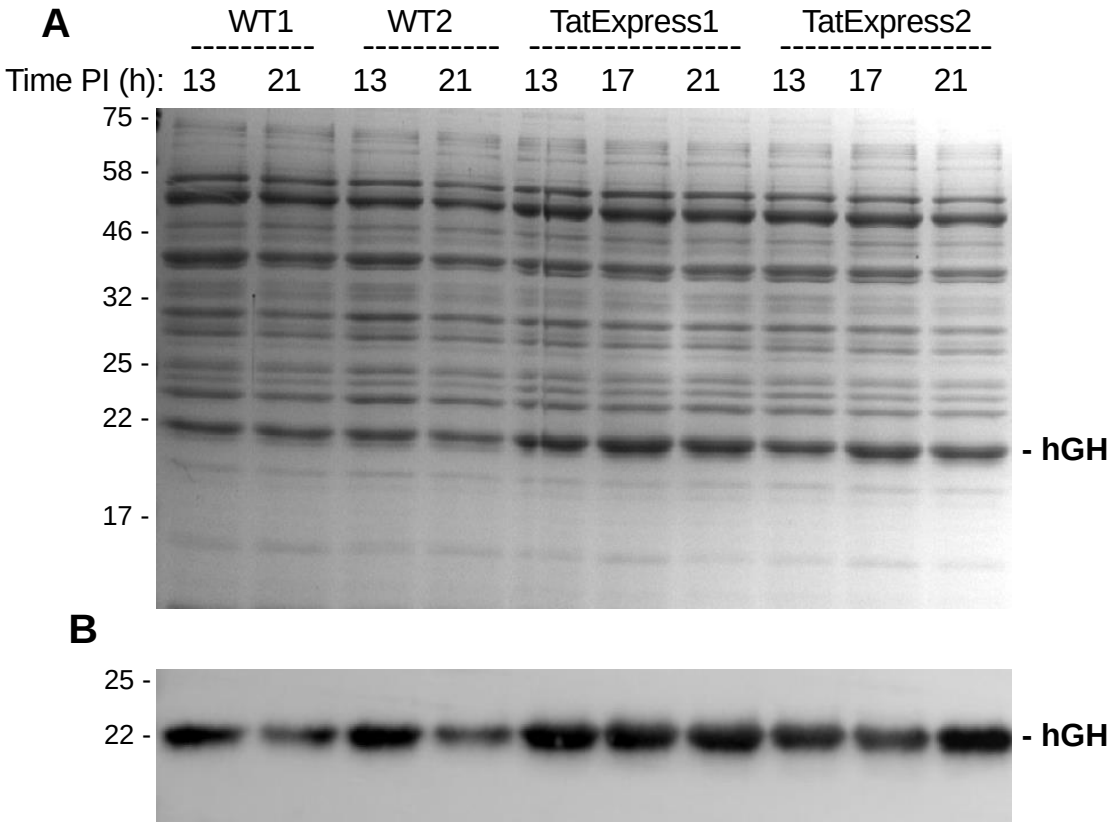


Figure 2. Tat-dependent export of TorA-hGH in W3110 wild type and TatExpress W3110 cells. Samples from duplicate fed-batch fermentation cultures of WT and TatExpress expressing TorA-hGH were removed at the indicated times post induction and fractionated to yield periplasmic samples. The periplasmic samples were analysed by Coomassie blue stained gels (A) and immunoblotting using antibodies to the His tag on hGH (B). For each lane a normalized amount of protein has been loaded, equivalent to OD_{600} 0.08 AU total cells (Coomassie blue stained gels), and OD_{600} 0.008 AU (immunoblots). Mobilites of molecular weight markers (in kDa) are shown on the left.

2.3 *Escherichia coli* 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway

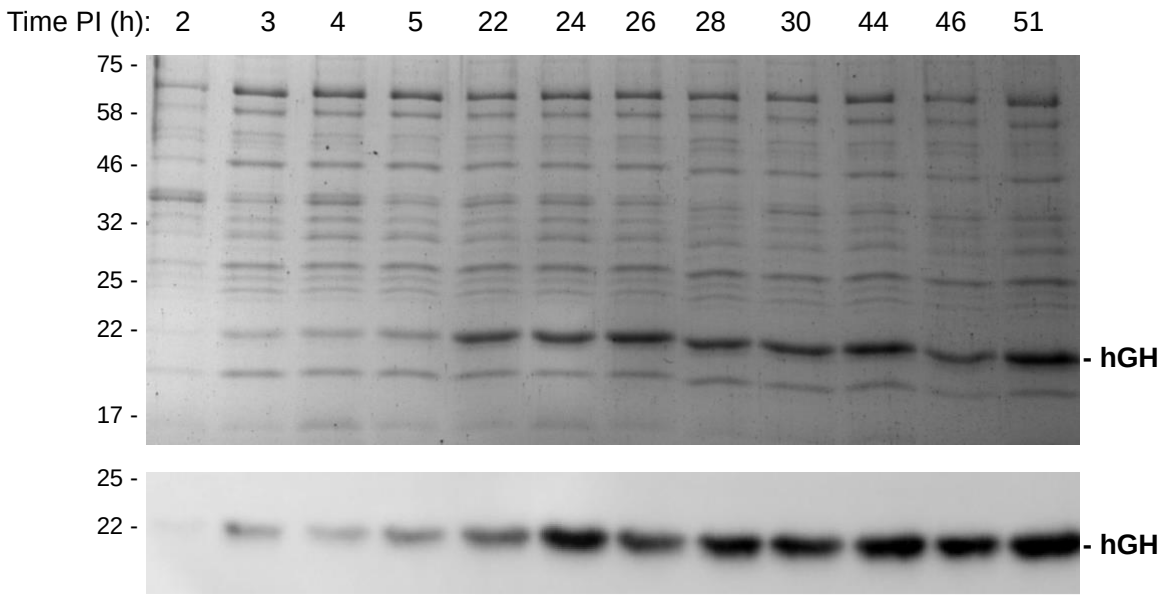


Figure 3. High-level export of TorA-hGH during extended fed-batch fermentation of TatExpress cells. TorA-hGH was expressed in TatExpress cells in a fed-batch fermentation system and samples were removed at the indicated times post induction. The samples were fractionated to generate periplasm samples which were then analysed on Coomassie blue stained gels (A) and by immunoblotting using antibodies to the His tag on hGH (B).

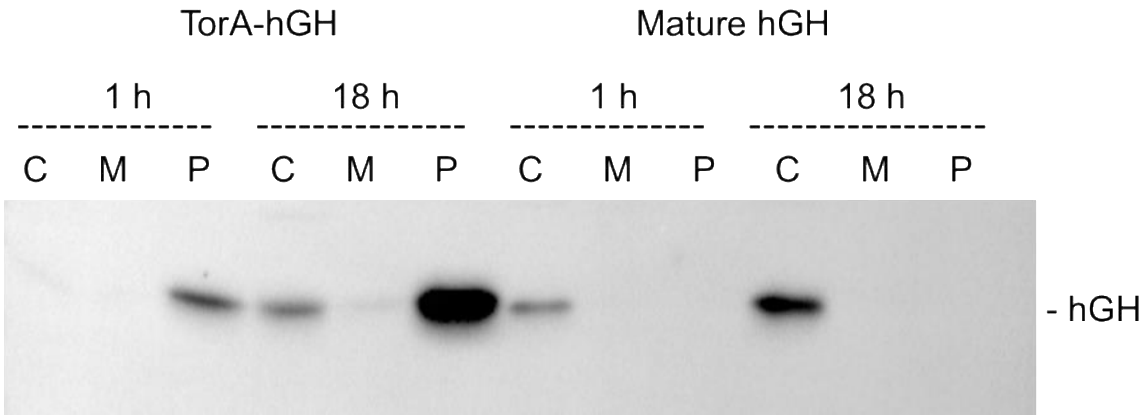
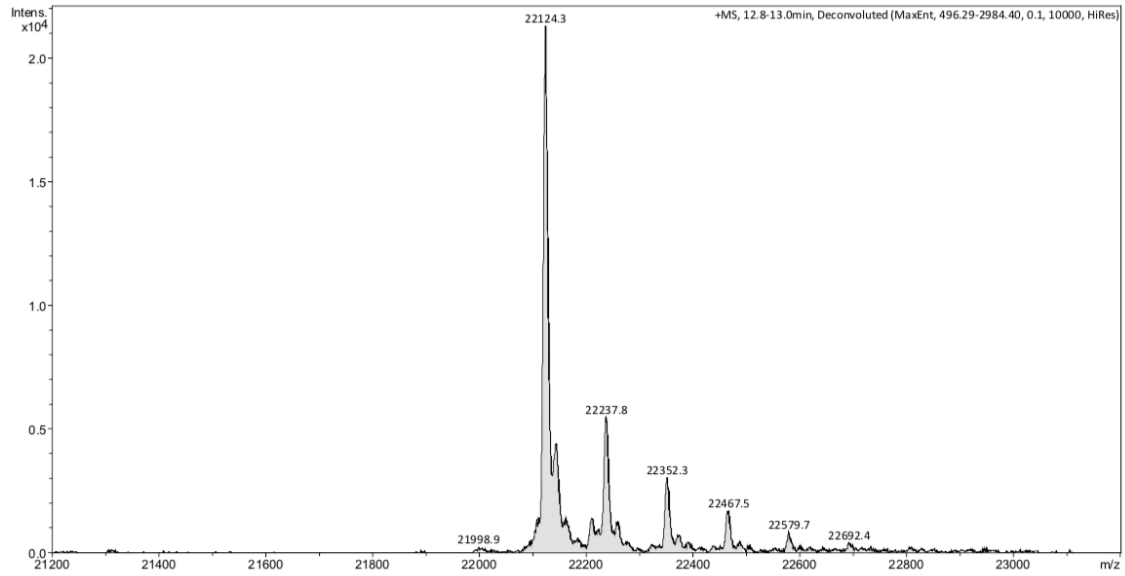


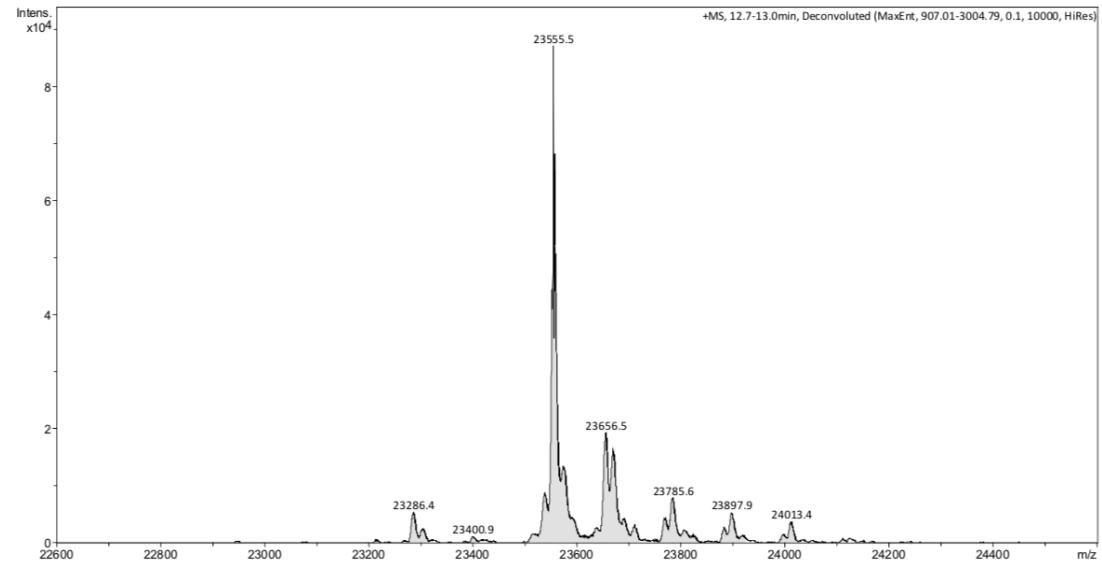
Figure 4. Mature hGH is not exported during fed-batch fermentation of TatExpress. Mature-size hGH and TorA-hGH were expressed in TatExpress cells using fed-batch fermentation systems. Samples were removed 1 h and 18 h post induction and fractionated

to generate cytoplasm, membrane and periplasm samples (C, M, P) which were then analysed by immunoblotting using antibodies to the His tag on hGH.

A



B



C

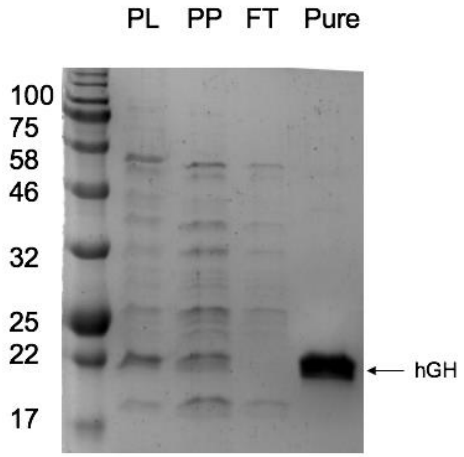
NNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATA♦AQAAHMFPTIPLSR LFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLCFSESIPTPSN REETQQKSNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKDLE

2.3 *Escherichia coli* 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway

EGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKDMD

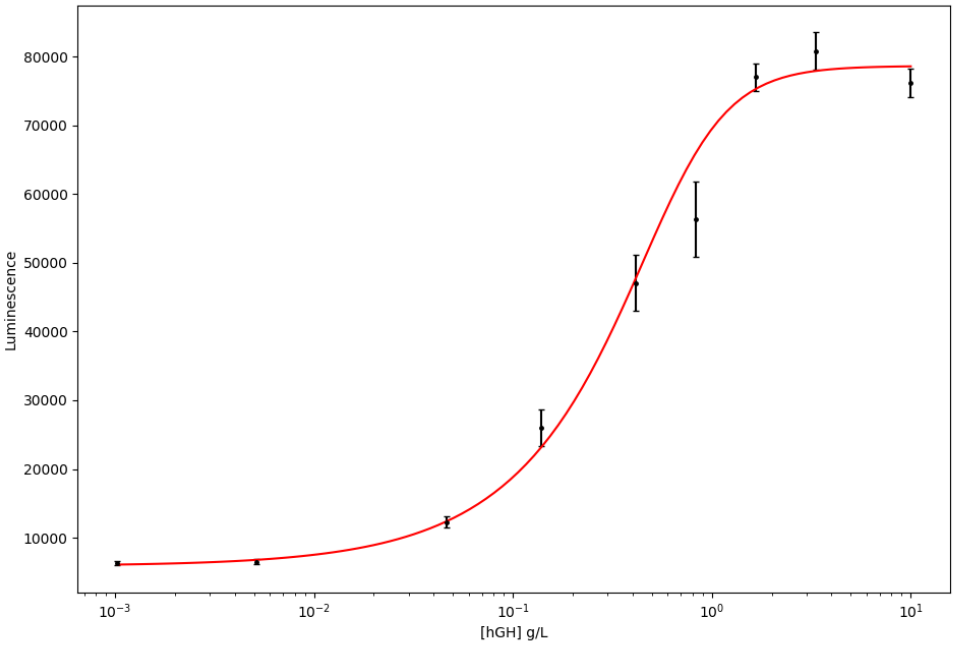
KVETFLRIVQCRSVEGSCGF**HHHHHH**

Figure 5: Exported periplasmic hGH is homogeneous and cleaved at the correct site. Periplasmic hGH was purified by affinity chromatography (see Supplementary Figure 1) and subjected to mass spectrometry analysis as detailed in Materials and Methods. Commercial hGH was analysed in an identical manner. The mass spectra for commercial hGH is shown in panel A and for periplasmic hGH in B. C: amino acid sequence of TorA-hGH-H6. The sequence for the TorA signal peptide (TMAO reductase) can be seen in blue, along with the first five amino acids of the mature TMAO protein and the initial methionine for the hGH protein (yellow). The sequence for the commercial hGH is underlined and the 6 histidine for the His-tag are in bold. The cleavage site between ATA and AQA is indicated with a ♦. The sequence for the commercial hGH is underlined, and this is followed by the 6-histidine tag.



Supplementary Figure 1. Purification of periplasmic hGH. Purification of hGH from periplasmic sample of fed-batch fermentation after 51 hours post induction via Nickel affinity column. Coomassie showing periplasmic released with lysozyme (PL),

periplasmic release with lysozyme free method (PP), flow through after purification (FT) and purified hGH (Pure). Molecular weight markers are shown on the left (in kDa) and mature hGH height is indicated on the right.



Supplementary Figure 2. Quantification of hGH concentration using the PathFinder bioassay. An hGH concentration standard curve was constructed using the supplied commercial hGH and all samples were analysed in triplicate using dilution factors in the range of the standard curve. The standard curve was adjusted using numpy, matplotlib.pyplot and scipy in Python.

3. General Discussion

3.1 Far-reaching consequences of *tat* deletion in *Escherichia coli* revealed by comprehensive proteome analyses.

3.1.1 Context

The applications of the Tat pathway in *E. coli* as a tool to facilitate downstream processes for complex proteins in an industrial setting has gained interest in recent years since the discovery of this pathway in the 1990s (Mould & Robinson 1991; Sargent et al. 1998), however our understanding of the importance of this pathway is still vague. The lack of a functioning pathway causes significant growth defects, particularly due to the mislocalization of Tat-dependant AmiA and AmiC, which are involved in the splitting the murein septa and the separation of cells (Bernhardt & de Boer 2003) (Ize et al. 2003). This mislocalization provokes the formation of long chains of cells that are unable to complete the separation from one another (Stanley et al. 2001). Other phenotypes associated with the loss of the Tat pathway function are problems with copper detoxification and iron acquisition (Ize et al. 2004), however this study was done only with a *ΔtatC* mutant, not a *ΔtatABCDE* mutant, meaning that different phenotypes could apply to the total elimination of the pathway. Finally, the role of the Tat pathway in bacterial pathogenesis is under investigation (Ball et al. 2016) (Pradel et al. 2003) (Ochsner et al. 2002), and learning more about it could help to modify and develop this export system for protein production.

3.1.2 Discussion of work presented

In this study, we have sought to ascertain at a comprehensive level how *E. coli* responds to a defect in the Tat protein export pathway. The deletion mutant *ΔtatABCDE* was used to determine the *E. coli* phenotype and proteome alterations under disruption of the Tat pathway. To do so the samples were fractionated into cytoplasmic, periplasmic and membrane fractions before analysis to determine changes not only in protein regulation but in localization as well.

The first problem we encountered was the abundance of impurities of cytoplasmic proteins in the periplasmic and membrane fractions due to both the limitation of the fractionation methods (Brown et al. 2010; Wolff et al. 2008) and the impairment of the *Δtat* membrane (Stanley et al. 2001). We calculated the relative the abundance of protein impurities in all fractions based on signal intensity. Due to the high percentage of cytoplasmic proteins in the periplasmic and membrane fractions of the *E. coli ΔtatABCED*, but not in the control strain we concluded that the complete deletion of the Tat system leads to the disruption of the inner membrane and cytoplasmic protein leakage during fractionation.

As stated before, the most notable phenotype of *Δtat* is the formation of filamentous, chain like structures where the cells have not been able to separate. During this study, we observed two novel phenotypes by examining transmission electron micrographs of the cells. The first was the formation of outer membrane vesicles (OMV), mostly predominant in cells that had failed to divide and likely due as a stress response to the cell envelope defect. The second phenotype was the moderate increase in inclusion body formation in comparison with the wild type. Although these features are interesting, since

the focus of the paper was the proteomic data this information was taken as a guideline to look more closely at the modifications in wall membrane regulation, which is why no statistical analysis of how common these features appeared in the sample was done. Although the low number of selected images do not truly reflect if the OMVs and IBs are a definitive feature, the fact that they are not present in the WT samples makes it likely that they are indirectly or directly caused by the deletion of tat.

By revising the proteomic data, we found that GroEL and DnaK, the main drivers of protein folding (Thomas & Francois Baneyx 1996), and IbpA and ClpB, identified as cytoplasmic components of E. coli IBs (Carrió & Villaverde 2002) were found to be more abundant in the membrane fraction of *Δtat*. We also found ZipA, an essential protein that forms the septal ring structure that mediates cell division (Hale & de Boer 1997), mislocalized to the membrane fraction in *Δtat* cells. Since none of these proteins are tat dependant their mislocalization and regulation is most likely a consequence of the membrane defect phenotype.

We then examined the tat substrates themselves. Of 27 reported proteins exported through the tat pathway and 8 additional hitchhiker proteins the data only showed 8 tat substrates and 3 hitchhikers. This low number of hits is due to the growth conditions used for this study, since the other Tat-dependant proteins have been found to be expressed under anaerobiosis (Hussain et al. 1994; Méjean et al. 1994; Jack et al. 2001), presence of cinnamaldehyde (Neumann et al. 2009) or hydroxyurea (Davies et al. 2009), or to be inhibited by other proteins present (Constantinidou et al. 2006). For the identified Tat substrates, we could distinguish 3 different localization and expression profiles. The first 4 proteins, HybA, EfeB, AmiA and AmiC, were found only in the periplasm of the WT

but not in the *Δtat* strain, strongly suggestion that they are exported exclusively by tat. Of these 4 proteins 2 of them, AmiA and AmiC were also found to be accumulated in the membrane fraction of both WT and the *Δtat* mutant, which may be due to association with the membrane before export. The next 2 proteins, CueO and NapA, were found in the periplasm of both WT and the Δtat mutant, however the relative quantity of protein was very low, suggesting that these traces of protein are artefacts of the fractionation process previously discussed. The next localization group consists of another 2 proteins, FdnG and DmsA, that are not present in the periplasmic fraction in either strain due to their tight association with inner membrane proteins. This association makes it likely that during the fractionation process these proteins were not separated with the rest of the periplasmic proteins. Finally, the last group are the before mentioned hitchhiker proteins: DmsB, HyaB and HybC. Unfortunately, none of them were found in the periplasm of either strain, so we rationalized that this was due to the lack of the proteins with which they form the multimeric complexes before export.

After analysing the destiny of the available Tat-substrates we turned our attention to other noticeable changes in the proteome linked to the loss of Tat export. A set of 98 proteins related to stress response were identified as having significant variation and were separate according to their functional group and represented in the heat map in Fig 3. We highlighted the activation of different chaperones and proteases in the periplasm involved in protein folding and degradation and other general cellular stress responses related to oxidative, detoxification, osmotic and heat stress. We also encountered changes in proteins involved in cell division and in cell wall, iron and molybdenum homeostasis. Overall, we obtained a thorough view of the different regulation patterns occurring in the cells when faced with the lack a functioning Tat system.

Finally, the last major change we observed was the presence of many of the proteins from the *cps* operon in *ΔtatABCDE* under aerobic conditions. The proteins from the *cps* genes found were those responsible for the production of colanic acid, an exopolysaccharide necessary for the formation of biofilms (Danese et al. 2000). To test if the presence of these proteins corresponded with production of colanic acid an assay was performed to determine the quantity of L-fucose, a specific component in colanic acid (Stevenson et al. 1996). In Figure 4C we can see that there is an increased amount in the *Δtat* strain in comparison with the WT. An additional assay to determine the ability of the cells to adhere to plastic surface to test for biofilm development was performed in both LB and minimal media since it has been reported that biofilm formation in *E. coli* can be observed mostly in minimal media (Naves et al. 2008). We found that that there was enhanced biofilm formation in *E. coli* Δtat, especially after 48 h in both media, although previous studies have reported a decrease in biofilm formation of the *ΔtatC* mutant (Ize et al. 2004).

In conclusion, we have identified novel phenotypes for this mutant such as the production of OMVs, increased inclusion body formation and colanic acid production which contributes to biofilm formation. We have also presented a global view of the consequences of a *tatABCDE* deletion through a comprehensive analysis of the proteome changes. This analysis has revealed the upregulation of proteins involved in protein folding, degradation and response to heat, oxidation, osmolarity and cold, as well as the induction of proteins responsible for cell wall biogenesis and the negative effect on the synthesis of iron and molybdenum transporters.

3.1.3 Future prospects

The major obstacle we encountered when analysing the data was distinguishing the effects of the *ΔtatABCDE* from the effects of the membrane defect, as the most notable stress responses seemed to be a consequence of the lack of cell separation. The cell chain phenotype has been successfully reversed by overexpression of AmiB (Ize et al. 2003). To repeat a similar study with this variation would provide insight as to which of the observed changes are related to the lack of Tat system and which are a stress response to not being able to correctly divide. It would also be of great interest to repeat this study under the specific conditions needed for the expression of the other Tat substrates, such as anaerobiosis or the presence of cinnamaldehyde or hydroxyurea, to examine what the effects of the mislocalization of those substrates has on the cells.

3.2 Comparative proteome analysis in an *Escherichia coli* CyDisCo strain identifies stress responses related to protein production oxidative stress and accumulation of misfolded protein.

3.2.1 Context

In this study, we sought to understand the physiological changes that result from the expression of a recombinant protein in both its correctly folded state and in a deliberately misfolded state using an *E. coli* CyDisCo platform. The expression of heterologous proteins in *E. coli* often results in protein misfolding and aggregation because the cell is not able to sustain non-physiological rates of expression. For its industrial application, the production process must be optimized to achieve high protein yields. Thus, under protein production process conditions it is important to identify stress parameters to circumvent and improve the process efficiency. For this reason, we were interested in characterizing the changes that occur during the expression of a misfolded Tat substrate, in order to understand whether additional responses occur. The substrate used was an scFv antibody construct that is efficiently exported by Tat while the 'misfolded' variant contains a 26-residue C-terminal extension that blocks export by Tat (Jones et al. 2016). Both proteins were synthesised with an N-terminal TorA signal peptide that directs export by Tat. For our study, we used a comprehensive proteomic approach to monitor the changes in cell physiology and proteome caused by recombinant protein secretion, misfolding and aggregation.

3.2.2 Discussion of work presented

This study examined the differences between cells that were exporting a correctly folded protein (scFv) through the Tat pathway and cells producing a mutant of that protein incapable of transport (Jones et al. 2016). After 3h post induction we see that the bulk of the correctly folded scFv is transported to the periplasm, but a fraction of the uncleaved protein is found in the insoluble body fraction, whereas the misfolded scFv does not export to the periplasm and is entirely found in the insoluble body fraction. This is also supported by the transmission electron micrographs showing that IBs are formed in the cells expressing the unfolded protein in Figure 2. This difference in protein localization must be putting additional strain on the system and the objective of this project was to identify this stress response.

The first indication of stress observable from the cells was the metabolic burden caused by the expression of recombinant protein and the maintenance of the plasmid (Bentley et al. 1990). Under the same growth conditions cells reached the stationary plateau at different ODs depending on if they were carrying a plasmid or not and if this plasmid was induced and was producing a protein of interest or not. Although the doubling times for all strains were similar we showed that cells with no plasmid had the reached the highest OD before the stationary phase, followed by those with uninduced plasmids. Of the induced cells, the one carrying a plasmid with no protein of interest reached a slightly higher OD than the other two carrying scFv and misfolded scFv. In the case of the cells producing the correctly folded scFv, this additional growth stress is most likely due to the recombinant protein synthesis and its export to the periplasm of the protein: although Tat transport does not consume ATP it uses the proton motive force to export the proteins to the periplasm (Natale et al. 2008), causing an energetic and resource deficit in comparison

to the control. In case of the misfolded scFv, the resources are diverted to the synthesis of a recombinant protein in a large quantity, which unlike the correctly folded scFv, is not secreted to the periplasm but deposited in inclusion bodies (visible in Fig. 2C).

To examine the stress responses of the cells we did a comprehensive proteome analysis on induced cells carrying a plasmid with the scFv, another with the misfolded scFv and a control plasmid that had no protein of interest. The changes to the amount of proteins were visualized using a Voronoi treemap where the proteins were clustered by their functional category so that the functionally related proteins were placed together for easier interpretation. The first conclusion is that expression of misfolded protein results in downregulation of many proteins involved in general metabolism and membrane transport. In addition, both strains expressing correctly folded scFv and misfolded showed a high number of significant changes mostly in proteins associated with protein folding and degradation, oxidative stress, membrane transport and integrity.

The first set of data we will look into is the stress response affecting transcription factors and DNA replication. The proteomic analysis revealed that there was an increase in the levels of different sigma factors in the cells producing misfolded and folded scFv. This observation confirms that these cells are compensating for the additional production of the protein of interest by increasing the capacity of the cells to handle the metabolic burden. We also observed an upregulation of many proteins involved in DNA replication occurring only in the cells producing the correctly folded TorA-scFv. For example, a protein as essential in DNA replication as PolA (DNA polymerase I) were found to be upregulated in scFv producing strain whereas the cells expressing mf_scFv showed no change. The upregulation of proteins involved in DNA replication only in cells producing

the folded protein suggests that, even though according to the growth curve all three conditions seem to either have reached (scFv and mf_scFv) or be entering (control) the stationary phase, the strain producing folded scFv is maintaining expression of proteins involved in DNA replication, repair and recombination (Meyer & Laine 1990). Finally, RlmJ, a protein involved in the catabolism of external DNA (Finkel & Kolter 2001) that has also been identified to repress cell-to-cell plasmid transfer (Matsuda et al. 2012), was upregulated in both strains producing folded and misfolded protein. Notably, the abundance of RlmJ is lower in the cells producing misfolded scFv, meaning that even though the misfolded protein stress seems to be partially responsible for the observed upregulation, this effect is greater with the production and transport of a folded protein to the periplasm.

The next data set involves changes in proteins involved in folding and the degradation process. We observed the upregulation of the chaperone systems DnaK-DnaJ-GrpE and GroEL-GroES, which can recognize unfolded proteins, refold or direct them to the cellular protein degradation machinery (Ziemienowicz et al. 1993). The highest abundance of these proteins was observed in the membrane fraction containing IBs, and the upregulation of GroES-GroEL and DnaK was notably increased in the strain expressing misfolded scFv. This observation correlates with our previous results and confirms the aggregation of moderate amounts of scFv in the cytoplasm and the high-level accumulation of misfolded scFv in IBs. The inclusion body formation in the CyDisCo scFv and misfolded strains was also confirmed by the presence of IbpA (Matuszewska et al. 2005), a heat shock protein with a strong ability to associate to protein aggregates, that was not identified in the empty plasmid control strain, in which we did not find any traces of IBs. Additionally, several other chaperones were found to

be highly abundant in the scFv and mf_scFv expressing strains. The accumulation of protein aggregates in the cytoplasm of *E. coli* CyDisCo is followed by activation of protein degradation machineries. We observed the upregulation of the ATPase components ClpA and ClpX of the Clp protease, Lon and HslV/U, together with periplasmic proteases DegP/Q. All these proteins are responsible for recognition, resolubilization and degradation of unfolded protein (Hoskins et al. 2000; Gottesman et al. 1998; Gao & O'Brian 2007). The highest difference between the abundance of scFv and mf_scFv expressing strains and the control was identified for periplasmic DegP and DegQ proteases. This suggests that in both conditions, scFv secretion to the periplasm and misfolded scFv aggregation in the cytoplasm, the cell envelope is exposed to unfavourable and harmful changes. The high upregulation of periplasmic proteases in the strain expressing misfolded scFv is particularly interesting and could indicate that some of the protein is in fact successfully exported to the periplasm, rather than being blocked by the Tat proofreading machinery. However, a previous study showed no evidence for transport of this misfolded substrate to the periplasm and it is more likely that the cell either senses the production of precursor protein and upregulates expression, or the proteases are upregulated as part of a general stress response.

Another point of interest we wanted to explore was the oxidative stress and redox regulation in the cytoplasm as a consequence of disulphide bonds being formed thanks to the CyDisCo components. Since the oxidative state of the cytoplasm does not occur naturally, we expected that this will provoke an oxidative stress response and changes in the abundance of redox proteins. Indeed, the proteomic analysis revealed the upregulation of several proteins related to oxidative stress response. Among these, major thiol-disulfide oxidoreductases: thioredoxin TrxA, glutaredoxin GrxA, together with thiol-

specific peroxidase Bcp, were upregulated in strains expressing protein. The upregulation of oxidoreductases was significantly higher in the strain producing misfolded scFv, probably due to the increased protein accumulation in IBs. In the cytoplasm of *E. coli*, thioredoxin and glutathione/glutaredoxin systems are required for the reduction of proteins that become oxidized, and thus they maintain the redox homeostasis (Prinz et al. 1997). However, during severe oxidative stress both systems can be overwhelmed, or their protein function can shift from reductase to oxidase, leaving cytosolic proteins susceptible to oxidation. It has been shown that thioredoxins and glutaredoxins can act as oxidases when exported to the oxidizing periplasm (Debarbieux & Beckwith 1998; Eser et al. 2009) and both, TrxA and TrxC, shift their function to oxidases in a *trxB* mutant (Stewart et al. 1998). A related point to consider is the fact that in the periplasm of *E. coli*, which is an oxidizing environment, other thioredoxin-related proteins such as DsbA, B, C, D, G catalyze the disulfide-bond formation and rearrangement. Thus, it appears that the function of thioredoxins and glutaredoxins is influenced by the redox potential of the cellular environment and in an oxidized cytoplasm of *E. coli* CyDisCo they will promote disulfide bond formation, rather than their reduction. In conclusion, the production of disulfide bond-containing scFv in CyDisCo increases the abundance of cellular redox proteins in the cytoplasm.

Finally, we also observed interesting changes occurring to membrane transporter proteins. The amount of Sec system components showed an increase abundance in strains producing recombinant protein. The upregulation of Sec machinery is likely an effect of elevated oxidative stress in the periplasm and changes in the cell envelope to which the cell is responding by secretion of cytoprotective stress proteins. Proteins involved in other transport systems were found in lowered amounts in both secreting strains: YidC,

involved in inserting proteins in the membrane, and ABC transporter proteins responsible for the utilization of di- and tripeptides as well as substrate-binding proteins (SBPs) of the main *E. coli* peptide transporters. Moreover, some essential components of other transport systems such as maltose (MalE/M/T/Y), glutathione (GsiB), Tol-Pal (TolB, CpoB) and lipid (Blc), or osmoregulated- periplasmic glucan (OPG) biosynthesis pathway (MdoB) were also downregulated. One feature that these downregulated proteins have in common is their role in membrane protein synthesis and membrane maintenance under stress conditions. This suggests that the scFv expression in the *E. coli* CyDisCo causes membrane impairment through a decrease in number of membrane transporters. Generally, it is plausible that downregulation of so many transporter proteins involved in metabolite uptake is a consequence of previously mentioned metabolic burden that occurs in cells expressing scFv and misfolded scFv. Lastly and most surprising we observed moderately lower levels of TatA and TatB as well as higher level of TatE in the strains expressing scFv and misfolded scFv. To follow up on this data we performed an RE-qPCR experiment on cells grown under the same conditions, taking additional samples both pre-induction and immediately, 30 min and 3h post-induction to assess the evolution of the changes in the cells. We found that induction has an immediate impact on expression level of *tatA* when expressing an exportable Tat substrate. After induction, a significant decrease of transcription levels of *tatA* and *tatB* genes, which matches with the proteomic data obtained. The decrease of expression of these proteins may be due to a negative feedback loop responsible for assuring that export remains low to avoid toxicity in the periplasm since many of the Tat substrates are involved in cell division, biofilm formation and pathogenicity (Lee et al. 2006).

Our proteome studies revealed significantly higher (5.71 log2 FC) abundance of antigen 43 (Ag43) in the strain expressing a misfolded scFv. Ag43 is an outer membrane protein that confers bacterial cell-to-cell aggregation (Diderichsen 1980). To verify if Ag43 upregulation promotes the cell aggregation in our strain, we first performed phase-contrast microscopy (Fig. 6 A1-3). Additionally, since the cell aggregation leads to enhanced sedimentation, we applied a simple sedimentation assay (Fig. 6 B) and calculated the aggregation rate in all CyDisCo strains (Fig. 6 C). Indeed, a comparison of the aggregation and sedimentation rate of these strains confirmed that the *E. coli* expressing misfolded scFv shows enhanced sedimentation and aggregation. We propose that the increase in the abundance of Ag43 is related to the lowered abundance of the transcription factor OxyR, which is a suppressor of antigen 43 expression (Ulett et al. 2007). OxyR transcription factor is activated through the formation of a disulfide bond and is deactivated by enzymatic reduction with glutaredoxin 1, GrxA (Zheng et al. 1998). We observed a decrease in the abundance of OxyR in the misfolded (-0.67 log2 FC) and its upregulation in scFv expressing strain. These results could only be observed after the induction of the scFv expression, thus we assume that the accumulation of misfolded scFv leads to the Ag43-mediated cell clustering.

In conclusion, *E. coli* CyDisCo strain enables folding of disulfide bond-containing scFv in the cytoplasm and its secretion in high amounts to the periplasm via the Tat system. The expression of an unfolded scFv led to protein aggregation and inclusion body formation. Decreased cell growth, low abundance of proteins involved in DNA replication and sigma 70 factor in *E. coli* expressing a misfolded scFv when compared to a folded scFv suggest that the unfolded protein stress accelerates cell transition to a stationary phase. We confirmed that the recombinant protein overexpression triggers the

upregulation of proteins involved in removing the aggregates, such as chaperones and proteases. The high abundance of GroEL-GroES, DnaK and ClpB in the insoluble fraction of *E. coli* expressing misfolded protein suggests their association with protein aggregates. We also observed that the formation of disulfide bonds in the cytoplasm of *E. coli* CyDisCo initiates the oxidative stress response and increases the abundance of redox regulation system proteins. This upregulation of thioredoxins and glutaredoxins was higher in the strain accumulating the misfolded oxidized scFv in the cytoplasm. Nevertheless, the redox protein ability to reduce oxidized proteins seems to be compromised by the oxidative environment of the cytoplasm. We also observed an extreme downregulation of membrane transporters which could be related to a major decrease in the abundance of YidC transport system, responsible for protein folding and insertion into the membrane. Surprisingly, a high rate of Tat-dependent secretion caused an increase in *tatA* expression only just after induction of synthesis. The following post induction analysis revealed lower *tatA* and *tatB* expression levels, which correlate with lower TatA and TatB protein abundance. Finally, the most distinct unfolded protein stress response we observed was the cell aggregation caused by the elevated levels of antigen 43. Interestingly, to the best of our knowledge, this phenomenon has not been reported yet as a consequence of misfolded protein expression, which makes it an interesting topic for further investigation.

3.2.3 Future prospects

The study identified characteristic changes occurring as a result of the production of both a folded and a misfolded protein, but also highlights an exclusive unfolded stress response. Countering and compensating for these changes may result in higher yields of pharmaceutically relevant proteins exported to the periplasm, however we acknowledge that to gain further insight into this topic further research must be done. It would be interesting to determine the structure of the misfolded scFv to see how much the additional 26 amino acid tail is disrupting the protein, however, many attempts have been made in the lab and due to its low stability, it has proved difficult to purify. It would also be interesting to observe how these changes occur over time by repeating these experiments with various time-points, both pre- and post-induction, as was done with the qRT-PCR experiment. A study of these characteristics was begun, using hGH and a misfolded variant of hGH as the protein of interest, but the analysis of the data generated was not complete by the time of writing. Finally, as mentioned before we see an upregulation of TatE and a downregulation of TatA when expressing proteins with a Tat signal peptide. TatA and TatE share 53% homology and have overlapping functions. It has been shown that individual deletions of TatA and TatE are still functional, however the double deletion blocks export to the periplasm via the Tat pathway (Sargent et al. 1998) (Baglieri et al. 2012). This preference for the use of either TatA or TatE is understudied and may be due to the additional stress conditions of the study or for a determined preference for the TorA signal peptide employed. To follow up on this observation we could more closely monitor the abundances of the Tat components at different time-points, under the expression of different proteins and with the use of different signal peptides to establish a pattern.

3.3 *Escherichia coli* 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway.

3.3.1 Context

In the previous segments, we have studied the response of the cells to a lack of Tat system (3.1) and to the use of said system for the export of a correctly folded and incorrectly folded protein (3.2) in order to better understand how the Tat system interacts with the rest of the cellular components. One of the findings from these studies was the downregulation both on a proteomic and transcriptomic level of TatA and TatB when a protein with a Tat specific signal peptide was induced. To compensate for this shift in the amount of TatA and TatB available and better exploit this pathway for commercial use we studied a new strain developed in collaboration with the university of Birmingham: TatExpress. The TatExpress strain has a pTac promoter introduced before the original promoter of the **TatABCD operon** (Browning et al. 2017), thus allowing for constitutive levels of Tat to be expressed until the moment of induction in which both the protein of interest and the Tat components will be overexpressed. In this chapter, we studied the robustness of TatExpress for the production and export of protein to the periplasm both at shake flask level as well as its up-scalability for use in a bioprocessing plant. The protein chosen to conduct this export study was hGH; a biopharmaceutical with great commercial worth and that has shown to give good export through the Tat pathway. Some of the experiments for the up-scale studies took place during a secondment to UCB's upstream and downstream processing department using 2L Sartorious bioreactors, while the rest of the study was carried out in house on 1.5L Minifors 2 bioreactors.

3.3.2 Discussion on work presented and future prospects

We have previously highlighted the advantages of using *E. coli* for protein production, from the ease of genetic manipulation to the affordable and fast growth profile that make this organism and ideal platform for making biopharmaceuticals. In this chapter, we have shown that the new TatExpress strain created in collaboration with the University of Birmingham is a versatile and robust strain that can greatly increase the amount of protein exported to the periplasm. Not only is this strain capable of surviving fermentative growth during up-scale to high ODs, it also continues producing and exporting protein to the periplasm for a prolonged time. This protein is easily purified from the periplasm and quantitative and activity assays suggest that the yields are comparable to industrial production in other platforms.

Initial steps towards the study of high yielding protein production in a non-industrial setting have been made. The optimization of the fermentation facility and of the protocols used are an important contribution towards the study of "up-scale" and will allow us to further contribute to the improvement of *E. coli* strains for the production of biopharmaceuticals. However, these are just the first stepping stones and there are several challenges that still need to be addressed before being able to move on to larger up-scale projects. Currently we were only able to optimize the fed-batch fermentation protocol for use of the W3110 TatExpress cells, however attempts to use BL21 TatExpress cells under the same conditions for up-scale have not been successful. Further modifications of the growth conditions need to be made before this strain may be used for fermentation.

Although there are studies about fermentative stress of different organisms using a proteomic approach (Helmel et al. 2014), little is recorded about fermentation stress in *E.*

coli (Kim et al. 2005). These studies pose a special difficulty due to the heterogeneity of the cells inside the bioreactor: though much effort is put into homogeneous stirring occurring the general shape and size of the vessel create isolated zones with specific conditions, from low oxygen spaces where cells enter micro-anaerobiosis to film formation around the probes or accessible tubing. These difference in cell clusters would make difficult any proteomic study since it would be easy for any cross contamination to occur within the own sample. However, a proteomic study similar to what we have seen in sections 3.1 and 3.2 would provide great insight into the fermentation process and may signal new stress response routes different from those already studied. The integration of this knowledge to reduce the stress in the cells by creating knockouts or overproduce determined proteins that would allow faster growth or better export would be an interesting and important contribution for industry.

Finally, there are still many more Tat specific proteins both from E. coli and from other organisms, whose signal peptide may be a better choice for protein export. However, for high-throughput screening the 1.5 L bioreactors pose a serious limitation, both on time and resources since each run is long and labour intensive. For this reason, alternative bioreactors should be considered. Although the cost of each run would be higher than that of the Infors bioreactors, the Ambr250 system allows for many more experiments to take place in parallel under the same conditions and would prove a time efficient way of selecting the best signal peptides for further upscale. The need to study different combinations of signal peptide is caused by limitations of the Tat transport proofreading mechanism itself. Although we see high yields of hGH and ScFv in the earlier chapters of this work, other proteins of interest tested at shake flask level did not give us similar results. Grown in the same conditions as those used for generating Fig 7, we tried

expressing a list of POI obtained through different collaborations with other laboratories or industry. The list of proteins can be found in annex 2. Of all the TorA-POI-His Tag constructs tried, the only one that was able to successfully translocate to the periplasm was WxL1, a surface cell wall-binding domain. This poses a unique challenge since although TatExpress has proved to be an excellent platform for protein production, the selectivity with which proteins are transported makes it difficult to design a system in which one POI can easily be interchanged with another.

In summary, the new bioreactors provide multiple opportunities for the better study of *E. coli* fed-batch fermentation and allows us to simulate industrial conditions in a lab setting. Although initial breakthroughs have been made during this study there are still many angles that need to be explored before a comprehensive understanding of high yielding protein production is obtained.

4 Concluding remarks

The aim of this project was to investigate the Twin Arginine Translocation Pathway in depth and determine if it is a suitable platform for biopharmaceutical production. Currently most industrial processes for protein production in *E. coli* is harvested from inclusion bodies by breaking both the inner and outer membrane, and then solubilizing the inclusion bodies and submitting the proteins to refolding, which makes this method not ideal for complex biopharmaceuticals. Although alternative host systems for these complex biopharmaceuticals are already available, they are excessively time consuming and expensive, as well as easily contaminated. For this reason, a bacterial system capable of replicating the essential post-translational modifications necessary for activity and also able to ease downstream processing by exporting the product to the periplasm where it can be easily obtained would be a welcome contribution to the current industrial market.

The focus of this project was therefor to investigate the Tat Pathway and its relationship with protein export and with the stress response this generates. Two proteomic studies took place to observe the changes occurring in the proteome when the Tat pathway is lacking and when it is being used to produce a protein of interest, even if said protein of interest activates the proof-reading mechanism and is rejected. These studies provided insight into the different defense mechanisms and stress responses that were occurring in the cells under said conditions and with this knowledge new strains can be designed to compensate for the additional stress the cells are subject to. A scale-up study was performed and not only showed that obtaining correctly folded, disulfide bonded and active protein from the periplasm can be achieved, but it helped to set up a protocol for in lab fermentation at the University of Kent. Taking together the obtained results it will now be easier to design new strains capable of better export by reducing the stressful

environment of the cells and then test them in bioreactors to determine their efficiency. In conclusion, this study has provided the stepping stones towards a future bacterial production platform able to compete with others in terms of ease of downstream processing, cost and product yields, however it is not without its own limitations.

5 References

References not including those specific to the previously discussed published papers.

Abrams, J.Y. et al., 2011. Lower Risk of Creutzfeldt-Jakob Disease in Pituitary Growth Hormone Recipients Initiating Treatment after 1977. *The Journal of Clinical Endocrinology & Metabolism*, 96(10), pp.E1666–E1669. Available at: http://dx.doi.org/10.1210/jc.2011-1357.

Baglieri, J. et al., 2012. Structure of TatA Paralog, TatE, Suggests a Structurally Homogeneous Form of Tat Protein Translocase That Transports Folded Proteins of Differing Diameter. *Journal of Biological Chemistry* , 287(10), pp.7335–7344. Available at: http://www.jbc.org/content/287/10/7335.abstract.

Balamurugan, V., Biotechnology in the Production of Recombinant Vaccine or Antigen for Animal Health. *Journal of Animal and Veterinary Advances*, v. 5.

Baldi, L. et al., 2007. Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives. *Biotechnology Letters*, 29(5), pp.677–684. Available at: https://doi.org/10.1007/s10529-006-9297-y.

Ball, G. et al., 2016. Contribution of the Twin Arginine Translocation system to the exoproteome of Pseudomonas aeruginosa. *Nature Publishing Group*, (May), pp.1–14. Available at: http://dx.doi.org/10.1038/srep27675.

Bandaranayake, A.D. & Almo, S.C., 2014. Recent advances in mammalian protein production. *FEBS letters*, 588(2), pp.253–260. Available at: https://www.ncbi.nlm.nih.gov/pubmed/24316512.

Beckwith, J., 2013. Fifty Years Fused to Lac. *Annual Review of Microbiology*, 67(1), pp.1–19. Available at: https://doi.org/10.1146/annurev-micro-092412-155732.

Behrendt, J. et al., 2004. Topological studies on the twin-arginine translocase component TatC. *FEMS Microbiology Letters*, 234(2), pp.303–308. Available at: https://doi.org/10.1111/j.1574-6968.2004.tb09548.x.

Bentley, W.E. et al., 1990. Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnology and Bioengineering*, 35(7), pp.668–681. Available at: https://doi.org/10.1002/bit.260350704.

Berks, B.C., 2015. The Twin-Arginine Protein Translocation Pathway. *Annual Review*

of Biochemistry, 84(1), pp.843–864. Available at: https://doi.org/10.1146/annurev-biochem-060614-034251.

Berks, B.C., Palmer, T. & Sargent, F., 2005. Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Current Opinion in Microbiology*, 8(2), pp.174–181. Available at: http://www.sciencedirect.com/science/article/pii/S1369527405000214.

Berks, B. C., Lea, S. M. and Stansfeld, P. J. (2014) 'Structural biology of Tat protein transport', *Current Opinion in Structural Biology*, 27, pp. 32–37. doi: https://doi.org/10.1016/j.sbi.2014.03.003.

Berks, B. C., Sargent, F. and Palmer, T. (2000) 'The Tat protein export pathway', *Molecular Microbiology*. John Wiley & Sons, Ltd (10.1111), 35(2), pp. 260–274. doi: 10.1046/j.1365-2958.2000.01719.x.

Berlec, A. & Štrukelj, B., 2013. Current state and recent advances in biopharmaceutical production in Escherichia coli, yeasts and mammalian cells. *Journal of Industrial Microbiology and Biotechnology*, 40(3–4), pp.257–274.

Bernhardt, T.G. & de Boer, P.A.J., 2003. The Escherichia coli amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway. *Molecular microbiology*, 48(5), pp.1171–1182. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12787347.

Biedendieck, R., 2016. A Bacillus megaterium System for the Production of Recombinant Proteins and Protein Complexes BT - Advanced Technologies for Protein Complex Production and Characterization. In M. C. Vega, ed. Cham: Springer International Publishing, pp. 97–113. Available at: https://doi.org/10.1007/978-3-319-27216-0_7.

Blümmel, A.-S. *et al.* (2015) 'Initial assembly steps of a translocase for folded proteins', *Nature Communications*, 6(1), p. 7234. doi: 10.1038/ncomms8234.

de Boer, H.A., Comstock, L.J. & Vasser, M., 1983. The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proceedings of the National Academy of Sciences*, 80(1), p.21 LP-25. Available at: http://www.pnas.org/content/80/1/21.abstract.

Bolhuis, A. *et al.* (2001) 'TatB and TatC Form a Functional and Structural Unit of the Twin-arginine Translocase from Escherichia coli ', *Journal of Biological Chemistry* , 276(23), pp. 20213–20219. doi: 10.1074/jbc.M100682200 .

Brown, R.N. et al., 2010. Mapping the Subcellular Proteome of Shewanella oneidensis

MR-1 using Sarkosyl-Based Fractionation and LC−MS/MS Protein Identification. *Journal of Proteome Research*, 9(9), pp.4454–4463. Available at: https://doi.org/10.1021/pr100215h.

Browning, D.F. et al., 2017. Escherichia coli 'TatExpress' strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway. *Biotechnology and bioengineering*, 114(12), pp.2828–2836. Available at: https://www.ncbi.nlm.nih.gov/pubmed/28842980.

Campbell, N. A. *et al.* (2004) *Biology: Concepts & Connections*. Pearson Education, Inc. as Benjamin Cummings.

Carrió, M.M. & Villaverde, A., 2002. Construction and deconstruction of bacterial inclusion bodies. *Journal of Biotechnology*, 96(1), pp.3–12. Available at: http://www.sciencedirect.com/science/article/pii/S0168165602000329.

Chien, A., Edgar, D.B. & Trela, J.M., 1976. Deoxyribonucleic acid polymerase from the extreme thermophile Thermus aquaticus. *Journal of bacteriology*, 127(3), pp.1550–1557. Available at: https://www.ncbi.nlm.nih.gov/pubmed/8432.

Constantinidou, C. et al., 2006. A Reassessment of the FNR Regulon and Transcriptomic Analysis of the Effects of Nitrate, Nitrite, NarXL, and NarQP as Escherichia coli K12 Adapts from Aerobic to Anaerobic Growth. *Journal of Biological Chemistry* , 281(8), pp.4802–4815. Available at: http://www.jbc.org/content/281/8/4802.abstract.

Cregg, J.M. et al., 2000. Recombinant protein expression in Pichia pastoris. *Molecular Biotechnology*, 16(1), pp.23–52. Available at: https://doi.org/10.1385/MB:16:1:23.

Danese, P.N., Pratt, L.A. & Kolter, R., 2000. Exopolysaccharide production is required for development of Escherichia coli K-12 biofilm architecture. *Journal of bacteriology*, 182(12), pp.3593–3596. Available at: https://www.ncbi.nlm.nih.gov/pubmed/10852895.

Davies, B.W. et al., 2009. Hydroxyurea Induces Hydroxyl Radical-Mediated Cell Death in Escherichia coli. *Molecular Cell*, 36(5), pp.845–860. Available at: http://dx.doi.org/10.1016/j.molcel.2009.11.024.

Debarbieux, L. & Beckwith, J., 1998. The reductive enzyme thioredoxin 1 acts as an oxidant when it is exported to the Escherichia coli periplasm. *Proceedings of the National Academy of Sciences of the United States of America*, 95(18), pp.10751–10756. Available at: https://www.ncbi.nlm.nih.gov/pubmed/9724776.

Deken, De, R.H., 1966. The Crabtree Effect: A Regulatory System in Yeast. *Microbiology*, 44(2), pp.149–156. Available at: https://mic.microbiologyresearch.org/content/journal/micro/10.1099/00221287-44-2-149.

DeLisa, M.P., Tullman, D. & Georgiou, G., 2003. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 100(10), pp.6115–6120. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12721369.

Diderichsen, B., 1980. flu, a metastable gene controlling surface properties of Escherichia coli. *Journal of bacteriology*, 141(2), pp.858–867. Available at: https://www.ncbi.nlm.nih.gov/pubmed/6102553.

Ding, N. et al., 2017. Increased glycosylation efficiency of recombinant proteins in Escherichia coli by auto-induction. *Biochemical and Biophysical Research Communications*, 485(1), pp.138–143. Available at: http://www.sciencedirect.com/science/article/pii/S0006291X17303054.

Driessen, A.J.M., 2001. SecB, a molecular chaperone with two faces. *Trends in Microbiology*, 9(5), pp.193–196. Available at: http://www.sciencedirect.com/science/article/pii/S0966842X01019801.

Eser, M. et al., 2009. Disulfide bond formation by exported glutaredoxin indicates glutathione's presence in the E. coli periplasm. *Proceedings of the National Academy of Sciences*, 106(5), p.1572 LP-1577. Available at: http://www.pnas.org/content/106/5/1572.abstract.

Finkel, S.E. & Kolter, R., 2001. DNA as a Nutrient: Novel Role for Bacterial Competence Gene Homologs. *Journal of Bacteriology*, 183(21), p.6288 LP-6293. Available at: http://jb.asm.org/content/183/21/6288.abstract.

Fröbel, J., Rose, P. & Müller, M., 2012. Twin-arginine-dependent translocation of folded proteins. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1592), pp.1029–1046. Available at: https://www.ncbi.nlm.nih.gov/pubmed/22411976.

Gao, T. & O'Brian, M.R., 2007. Control of DegP-dependent degradation of c-type cytochromes by heme and the cytochrome c maturation system in Escherichia coli. *Journal of bacteriology*, 189(17), pp.6253–6259. Available at: https://www.ncbi.nlm.nih.gov/pubmed/17616605.

Gibson, B. et al., 2018. The distribution of bacterial doubling times in the wild.

Proceedings. Biological sciences, 285(1880), p.20180789. Available at: https://www.ncbi.nlm.nih.gov/pubmed/29899074.

Gohlke, U. *et al.* (2005) 'The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter', *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), pp. 10482 LP – 10486. doi: 10.1073/pnas.0503558102

Gottesman, S. et al., 1998. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes & development*, 12(9), pp.1338–1347. Available at: https://www.ncbi.nlm.nih.gov/pubmed/9573050.

Gottesman, S. & Stout, V., 1991. Regulation of capsular polysaccharide synthesis in Escherichia coli K12. *Molecular Microbiology*, 5(7), pp.1599–1606. Available at: https://doi.org/10.1111/j.1365-2958.1991.tb01906.x.

Gouffi, K. *et al.* (2004) 'Dual Topology of the Escherichia coli TatA Protein', *Journal of Biological Chemistry* , 279(12), pp. 11608–11615. doi: 10.1074/jbc.M313187200 .

Hale, C.A. & de Boer, P.A.J., 1997. Direct Binding of FtsZ to ZipA, an Essential Component of the Septal Ring Structure That Mediates Cell Division in E. coli. *Cell*, 88(2), pp.175–185. Available at: https://doi.org/10.1016/S0092-8674(00)81838-3.

Hamilton, S.R. & Gerngross, T.U., 2007. Glycosylation engineering in yeast: the advent of fully humanized yeast. *Current Opinion in Biotechnology*, 18(5), pp.387–392. Available at: http://www.sciencedirect.com/science/article/pii/S0958166907001139.

Hatahet, F. et al., 2010. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. *Microbial Cell Factories*, 9(1), p.67. Available at: https://doi.org/10.1186/1475-2859-9-67.

Helmel, M. et al., 2014. Proteome profiling illustrated by a large-scale fed-batch fermentation of Penicillium chrysogenum. *EuPA Open Proteomics*, 4, pp.113–120. Available at: http://www.sciencedirect.com/science/article/pii/S2212968514000415.

Higel, F. et al., 2016. N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *European Journal of Pharmaceutics and Biopharmaceutics*, 100, pp.94–100.

Available at: http://www.sciencedirect.com/science/article/pii/S0939641116000163.

Hoskins, J.R. et al., 2000. Protein binding and unfolding by the chaperone ClpA and degradation by the protease ClpAP. *Proceedings of the National Academy of Sciences of the United States of America*, 97(16), pp.8892–8897. Available at: https://www.ncbi.nlm.nih.gov/pubmed/10922051.

Huang, M., Bao, J. & Nielsen, J., 2014. Biopharmaceutical protein production bySaccharomyces cerevisiae: current state and future prospects. *Pharmaceutical Bioprocessing*, 2(2), pp.167–182.

Hussain, H. et al., 1994. A seven-gene operon essential for formate-dependent nitrite reduction to ammonia by enteric bacteria. *Molecular Microbiology*, 12(1), pp.153–163. Available at: https://doi.org/10.1111/j.1365-2958.1994.tb01004.x.

Isaksson, O.G.P., Eden, S. & Jansson, J., 1985. Mode of Action of Pituitary Growth Hormone on Target Cells. *Annual Review of Physiology*, 47(1), pp.483–499. Available at: https://doi.org/10.1146/annurev.ph.47.030185.002411.

Itakura, K. et al., 1977. Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. *Science*, 198(4321), p.1056 LP-1063. Available at: http://science.sciencemag.org/content/198/4321/1056.abstract.

Ize, B. et al., 2004. Novel Phenotypes of Escherichia coli tat Mutants Revealed by Global Gene Expression and Phenotypic Analysis. *Journal of Biological Chemistry* , 279(46), pp.47543–47554. Available at: http://www.jbc.org/content/279/46/47543.abstract.

Ize, B. et al., 2003. Role of the Escherichia coli Tat pathway in outer membrane integrity. *Molecular Microbiology*, 48(5), pp.1183–1193. Available at: https://doi.org/10.1046/j.1365-2958.2003.03504.x.

Jack, R.L. et al., 2001. Constitutive expression of Escherichia coli tat genes indicates an important role for the twin-arginine translocase during aerobic and anaerobic growth. *Journal of bacteriology*, 183(5), pp.1801–1804. Available at: https://www.ncbi.nlm.nih.gov/pubmed/11160116.

Jackson, D.A., Symons, R.H. & Berg, P., 1972. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 69(10), pp.2904–2909. Available at:

https://www.ncbi.nlm.nih.gov/pubmed/4342968.

Jones, A.S. et al., 2016. Proofreading of substrate structure by the Twin-Arginine Translocase is highly dependent on substrate conformational flexibility but surprisingly tolerant of surface charge and hydrophobicity changes. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1863(12), pp.3116–3124. Available at: http://www.sciencedirect.com/science/article/pii/S0167488916302269.

Jozala, A.F. et al., 2016. Biopharmaceuticals from microorganisms: from production to purification. *Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology]*, 47 Suppl 1(Suppl 1), pp.51–63. Available at: https://www.ncbi.nlm.nih.gov/pubmed/27838289.

Jung, E. & Williams, K.L., 1997. Theproduction of recombinant glycoproteins with special reference to simple eukaryotesincluding Dictyostelium discoideum. *Biotechnology and Applied Biochemistry*, 25(1), pp.3–8. Available at: https://doi.org/10.1111/j.1470-8744.1997.tb00407.x.

Kim, Y.-H. et al., 2005. Proteome response of Escherichia coli fed-batch culture to temperature downshift. *Applied Microbiology and Biotechnology*, 68(6), pp.786–793. Available at: https://doi.org/10.1007/s00253-005-0053-3.

Klint, J.K. et al., 2013. Production of Recombinant Disulfide-Rich Venom Peptides for Structural and Functional Analysis via Expression in the Periplasm of E. coli. *PLOS ONE*, 8(5), p.e63865. Available at: https://doi.org/10.1371/journal.pone.0063865.

Kost, T.A. & Condreay, J.P., 1999. Recombinant baculoviruses as expression vectors for insect and mammalian cells. *Current Opinion in Biotechnology*, 10(5), pp.428–433. Available at: http://www.sciencedirect.com/science/article/pii/S0958166999000051.

Kuczewski, M. et al., 2011. A single-use purification process for the production of a monoclonal antibody produced in a PER.C6 human cell line. *Biotechnology Journal*, 6(1), pp.56–65. Available at: https://doi.org/10.1002/biot.201000292.

Kurtzman, C.P., 2009. Biotechnological strains of Komagataella (Pichia) pastoris are Komagataellaphaffii as determined from multigene sequence analysis. *Journal of Industrial Microbiology & Biotechnology*, 36(11), p.1435. Available at: https://doi.org/10.1007/s10295-009-0638-4.

Landowski, C.P. et al., 2016. Enabling low cost biopharmaceuticals: high level interferon alpha-2b production in Trichoderma reesei. *Microbial cell factories*, 15(1), p.104. Available at: https://www.ncbi.nlm.nih.gov/pubmed/27287473.

Laurent, J.M. et al., 2016. Efforts to make and apply humanized yeast. *Briefings in functional genomics*, 15(2), pp.155–163. Available at: https://www.ncbi.nlm.nih.gov/pubmed/26462863.

Leake, M.C. et al., 2008. Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by in vivo single-molecule imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 105(40), pp.15376–15381. Available at: https://www.ncbi.nlm.nih.gov/pubmed/18832162.

Lee, H.S., Qi, Y. & Im, W., 2015. Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Scientific reports*, 5, p.8926. Available at: https://www.ncbi.nlm.nih.gov/pubmed/25748215.

Lee, P.A. et al., 2002. Truncation analysis of TatA and TatB defines the minimal functional units required for protein translocation. *Journal of bacteriology*, 184(21), pp.5871–5879. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12374820.

Lee, P.A., Tullman-Ercek, D. & Georgiou, G., 2006. The bacterial twin-arginine translocation pathway. *Annual review of microbiology*, 60, pp.373–395. Available at: https://www.ncbi.nlm.nih.gov/pubmed/16756481.

Lens, J. & Evertzen, A., 1952. The difference between insulin from cattle and from pigs. *Biochimica et Biophysica Acta*, 8, pp.332–338. Available at: http://www.sciencedirect.com/science/article/pii/0006300252900486.

Lewis, U.J., Sinha, Y.N. & Haro, L.S., 2018. Variant forms and fragments of human growth hormone in serum. *Acta Paediatrica*, 83(s399), pp.29–31. Available at: https://doi.org/10.1111/j.1651-2227.1994.tb13282.x.

Lodish H, Berk A, Zipursky SL, et al., 2000. DNA Cloning with Plasmid Vectors. *Molecular Cell Biology*, (Section 7.1).

van Loon, K., 1998. Safety of High Doses of Recombinant Human Growth Hormone. *Hormone Research in Paediatrics*, 49(suppl 2(Suppl. 2), pp.78–81. Available at: https://www.karger.com/DOI/10.1159/000053092.

Luirink, J. et al., 2012. Biogenesis of inner membrane proteins in Escherichia coli.

Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1817(6), pp.965–976. Available at: http://www.sciencedirect.com/science/article/pii/S0005272811002994.

Ma, X. and Cline, K. (2010) 'Multiple precursor proteins bind individual Tat receptor complexes and are collectively transported', *The EMBO Journal*. John Wiley & Sons, Ltd, 29(9), pp. 1477–1488. doi: 10.1038/emboj.2010.44.

Martínez-Solís, M. et al., 2016. A novel baculovirus-derived promoter with high activity in the baculovirus expression system. *PeerJ*, 4, pp.e2183–e2183. Available at: https://www.ncbi.nlm.nih.gov/pubmed/27375973.

Matsuda, A. et al., 2012. Genome-wide screen for Escherichia coli genes involved in repressing cell-to-cell transfer of non-conjugative plasmids. *Biochemical and Biophysical Research Communications*, 428(4), pp.445–450. Available at: http://www.sciencedirect.com/science/article/pii/S0006291X12020906.

Mattanovich, D. et al., 2012. Recombinant Protein Production in Yeasts BT - Recombinant Gene Expression. In A. Lorence, ed. Totowa, NJ: Humana Press, pp. 329–358. Available at: https://doi.org/10.1007/978-1-61779-433-9_17.

Matuszewska, M. et al., 2005. The Small Heat Shock Protein IbpA of Escherichia coli Cooperates with IbpB in Stabilization of Thermally Aggregated Proteins in a Disaggregation Competent State. *Journal of Biological Chemistry* , 280(13), pp.12292–12298. Available at: http://www.jbc.org/content/280/13/12292.abstract.

Méjean, V. et al., 1994. TMAO anaerobic respiration in Escherichia coli: involvement of the tor operon. *Molecular Microbiology*, 11(6), pp.1169–1179. Available at: https://doi.org/10.1111/j.1365-2958.1994.tb00393.x.

Meyer, R.R. & Laine, P.S., 1990. The single-stranded DNA-binding protein of Escherichia coli. *Microbiological reviews*, 54(4), pp.342–380. Available at: https://www.ncbi.nlm.nih.gov/pubmed/2087220.

Mori, H. & Ito, K., 2001. The Sec protein-translocation pathway. *Trends in Microbiology*, 9(10), pp.494–500. Available at: https://doi.org/10.1016/S0966-842X(01)02174-6.

Mould, R.M. & Robinson, C., 1991. A proton gradient is required for the transport of two lumenal oxygen-evolving proteins across the thylakoid membrane. *Journal of Biological Chemistry* , 266(19), pp.12189–12193. Available at: http://www.jbc.org/content/266/19/12189.abstract.

Mullis, K. et al., 1986. Specific Enzymatic Amplification of DNA In Vitro: The

Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology* , 51, pp.263–273. Available at: http://symposium.cshlp.org/content/51/263.short.

Natale, P., Brüser, T. & Driessen, A.J.M., 2008. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane—Distinct translocases and mechanisms. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1778(9), pp.1735–1756. Available at: http://www.sciencedirect.com/science/article/pii/S0005273607002775.

Naves, P. et al., 2008. Measurement of biofilm formation by clinical isolates of Escherichia coli is method-dependent. *Journal of Applied Microbiology*, 105(2), pp.585–590. Available at: https://doi.org/10.1111/j.1365-2672.2008.03791.x.

Nett, J.H. et al., 2011. A combinatorial genetic library approach to target heterologous glycosylation enzymes to the endoplasmic reticulum or the Golgi apparatus of Pichia pastoris. *Yeast*, 28(3), pp.237–252. Available at: https://doi.org/10.1002/yea.1835.

Neumann, M. et al., 2009. A periplasmic aldehyde oxidoreductase represents the first molybdopterin cytosine dinucleotide cofactor containing molybdo-flavoenzyme from Escherichia coli. *The FEBS Journal*, 276(10), pp.2762–2774. Available at: https://doi.org/10.1111/j.1742-4658.2009.07000.x.

Nguyen, V.D. et al., 2011. Pre-expression of a sulfhydryl oxidase significantly increases the yields of eukaryotic disulfide bond containing proteins expressed in the cytoplasm of E.coli. *Microbial cell factories*, 10, p.1. Available at: https://www.ncbi.nlm.nih.gov/pubmed/21211066.

Ochsner, U.A. et al., 2002. Effects of the twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), pp.8312–8317. Available at: https://www.ncbi.nlm.nih.gov/pubmed/12034867.

Overton, T.W., 2014. Recombinant protein production in bacterial hosts. *Drug Discovery Today*, 19(5), pp.590–601. Available at: http://www.sciencedirect.com/science/article/pii/S1359644613004029.

Palmer, I. & Wingfield, P.T., 2004. Preparation and extraction of insoluble (inclusion-body) proteins from Escherichia coli. *Current protocols in protein science*, Chapter 6, p.Unit-6.3. Available at: https://www.ncbi.nlm.nih.gov/pubmed/18429271.

Palmer, T. & Berks, B.C., 2012. The twin-arginine translocation (Tat) protein export

pathway. *Nat Rev Micro*, 10(7), pp.483–496. Available at: http://dx.doi.org/10.1038/nrmicro2814.

Patel, R., Smith, S.M. & Robinson, C., 2014. Protein transport by the bacterial Tat pathway. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1843(8), pp.1620–1628. Available at: http://www.sciencedirect.com/science/article/pii/S016748891400069X.

Pierce, J.J. et al., 1997. Factors determining more efficient large-scale release of a periplasmic enzyme from E. coli using lysozyme. *Journal of Biotechnology*, 58(1), pp.1–11. Available at: http://www.sciencedirect.com/science/article/pii/S0168165697001168.

Pradel, N. et al., 2003. Contribution of the Twin Arginine Translocation System to the Virulence of Enterohemorrhagic Escherichia coli O157:H7. *Infection and Immunity*, 71(9), p.4908 LP-4916. Available at: http://iai.asm.org/content/71/9/4908.abstract.

Prinz, W.A. et al., 1997. The Role of the Thioredoxin and Glutaredoxin Pathways in Reducing Protein Disulfide Bonds in the Escherichia coliCytoplasm . *Journal of Biological Chemistry* , 272(25), pp.15661–15667. Available at: http://www.jbc.org/content/272/25/15661.abstract.

Ramasamy, S. *et al.* (2013) 'The Glove-like Structure of the Conserved Membrane Protein TatC Provides Insight into Signal Sequence Recognition in Twin-Arginine Translocation', *Structure*. Elsevier, 21(5), pp. 777–788. doi: 10.1016/j.str.2013.03.004.

Retallack, D. et al., 2006. Pseudomonas fluorescens -- a robust expression platform for pharmaceutical protein production. *Microbial Cell Factories*, 5(1), p.S28. Available at: https://doi.org/10.1186/1475-2859-5-S1-S28.

Reuten, R. et al., 2016. Maltose-Binding Protein (MBP), a Secretion-Enhancing Tag for Mammalian Protein Expression Systems. *PloS one*, 11(3), pp.e0152386–e0152386. Available at: https://www.ncbi.nlm.nih.gov/pubmed/27029048.

Rodrigue, A. et al., 1999. Co-translocation of a Periplasmic Enzyme Complex by a Hitchhiker Mechanism through the Bacterial Tat Pathway. *Journal of Biological Chemistry* , 274(19), pp.13223–13228. Available at: http://www.jbc.org/content/274/19/13223.abstract.

Rodriguez, F. *et al.* (2013) 'Structural model for the protein-translocating element of the twin-arginine transport system', *Proceedings of the National Academy of Sciences*,

110(12), p. E1092 LP-E1101. doi: 10.1073/pnas.1219486110.

Sanger, F., Nicklen, S. & Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), pp.5463–5467. Available at: https://www.ncbi.nlm.nih.gov/pubmed/271968.

Sargent, F., Berks, B. C., & Palmer, T. (n.d.). The Twin-Arginine Transport System. In Protein Movement Across Membranes (pp. 71–84). Springer US. https://doi.org/10.1007/0-387-30871-7_6

Sargent, F. et al., 1998. Overlapping functions of components of a bacterial Sec-independent protein export pathway. *The EMBO journal*, 17(13), pp.3640–3650. Available at: https://www.ncbi.nlm.nih.gov/pubmed/9649434.

Schäfer, F. et al., 2015. Chapter Nine - Purification of GST-Tagged Proteins. In J. R. B. T.-M. in E. Lorsch, ed. *Laboratory Methods in Enzymology: Protein Part D*. Academic Press, pp. 127–139. Available at: http://www.sciencedirect.com/science/article/pii/S0076687914000706.

Siggers, K.A. & Lesser, C.F., 2008. The Yeast Saccharomyces cerevisiae: A Versatile Model System for the Identification and Characterization of Bacterial Virulence Proteins. *Cell Host & Microbe*, 4(1), pp.8–15. Available at: https://doi.org/10.1016/j.chom.2008.06.004.

Smales, C.M. & James, D.C., 2005. *Therapeutic Proteins: Methods and Protocols*, Humana Press. Available at: https://books.google.es/books?id=NZOnoS2PX18C.

Stanley, N.R. et al., 2001. Escherichia coli strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope. *Journal of bacteriology*, 183(1), pp.139–144. Available at: https://www.ncbi.nlm.nih.gov/pubmed/11114910.

Stanley, N.R., Palmer, T. & Berks, B.C., 2000. The Twin Arginine Consensus Motif of Tat Signal Peptides Is Involved in Sec-independent Protein Targeting in Escherichia coli . *Journal of Biological Chemistry* , 275(16), pp.11591–11596. Available at: http://www.jbc.org/content/275/16/11591.abstract.

Stanley, P., Taniguchi, N. & Aebi, M., 2017. N-Glycans. In *Essentials of Glycobiology*. p. Chapter 9.

Stevenson, G. et al., 1996. Organization of the Escherichia coli K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *Journal*

of bacteriology, 178(16), pp.4885–4893. Available at: https://www.ncbi.nlm.nih.gov/pubmed/8759852.

Stewart, E.J., Åslund, F. & Beckwith, J., 1998. Disulfide bond formation in the Escherichia coli cytoplasm: an in vivo role reversal for the thioredoxins. *The EMBO Journal*, 17(19), p.5543 LP-5550. Available at: http://emboj.embopress.org/content/17/19/5543.abstract.

Su, X. et al., 2012. Chapter One - Heterologous Gene Expression in Filamentous Fungi. In G. M. Gadd & S. B. T.-A. in A. M. Sariaslani, eds. *Advances in Applied Microbiology*. Academic Press, pp. 1–61. Available at: http://www.sciencedirect.com/science/article/pii/B9780123943828000010.

Sutherland, G.A. et al., 2018. Probing the quality control mechanism of the Escherichia coli twin-arginine translocase with folding variants of a de novo-designed heme protein. *The Journal of biological chemistry*, 293(18), pp.6672–6681. Available at: https://www.ncbi.nlm.nih.gov/pubmed/29559557.

Tarry, M. J. *et al.* (2009) 'Structural analysis of substrate binding by the TatBC component of the twin-arginine protein transport system', *Proceedings of the National Academy of Sciences*, 106(32), pp. 13284 LP – 13289. doi: 10.1073/pnas.0901566106.

Taubert, J. *et al.* (2015) 'TatBC-Independent TatA/Tat Substrate Interactions Contribute to Transport Efficiency', *PLOS ONE*. Public Library of Science, 10(3), p. e0119761. Available at: https://doi.org/10.1371/journal.pone.0119761.

Thomas, J.G. & Baneyx, F., 1996. Protein folding in the cytoplasm of Escherichia coli: requirements for the DnaK-DnaJ-GrpE and GroEL-GroES molecular chaperone machines. *Molecular Microbiology*, 21(6), pp.1185–1196. Available at: https://doi.org/10.1046/j.1365-2958.1996.651436.x.

Thomas, J.G. & Baneyx, F., 1996. Protein Misfolding and Inclusion Body Formation in Recombinant Escherichia coli Cells Overexpressing Heat-shock Proteins . *Journal of Biological Chemistry* , 271(19), pp.11141–11147. Available at: http://www.jbc.org/content/271/19/11141.abstract.

Tinker, J.K., Erbe, J.L. & Holmes, R.K., 2005. Characterization of fluorescent chimeras of cholera toxin and Escherichia coli heat-labile enterotoxins produced by use of the twin arginine translocation system. *Infection and immunity*, 73(6), pp.3627–3635. Available at: https://www.ncbi.nlm.nih.gov/pubmed/15908392.

Ulett, G.C. et al., 2007. Functional Analysis of Antigen 43 in Uropathogenic Escherichia coli Reveals a Role in Long-Term Persistence in the Urinary Tract. *Infection and Immunity*, 75(7), p.3233 LP-3244. Available at: http://iai.asm.org/content/75/7/3233.abstract.

Veenendaal, A.K.J., van der Does, C. & Driessen, A.J.M., 2004. The protein-conducting channel SecYEG. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1694(1), pp.81–95. Available at: http://www.sciencedirect.com/science/article/pii/S0167488904000916.

Vieira Gomes, A.M. et al., 2018. Comparison of Yeasts as Hosts for Recombinant Protein Production. *Microorganisms*, 6(2), p.38. Available at: https://www.ncbi.nlm.nih.gov/pubmed/29710826.

Walsh, G., 2014. Biopharmaceutical benchmarks 2014. *Nature Biotechnology*, 32, p.992. Available at: https://doi.org/10.1038/nbt.3040.

Walsh, G. & Jefferis, R., 2006. Post-translational modifications in the context of therapeutic proteins. *Nature Biotechnology*, 24, p.1241. Available at: https://doi.org/10.1038/nbt1252.

Walther, T. H. *et al.* (2013) 'Folding and Self-Assembly of the TatA Translocation Pore Based on a Charge Zipper Mechanism', *Cell*. Elsevier, 152(1), pp. 316–326. doi: 10.1016/j.cell.2012.12.017.

Wexler, M. et al., 2000. TatD Is a Cytoplasmic Protein with DNase Activity: NO REQUIREMENT FOR TatD FAMILY PROTEINS IN Sec-INDEPENDENT PROTEIN EXPORT . *Journal of Biological Chemistry* , 275(22), pp.16717–16722. Available at: http://www.jbc.org/content/275/22/16717.abstract.

Wildt, S. & Gerngross, T.U., 2005. The humanization of N-glycosylation pathways in yeast. *Nature Reviews Microbiology*, 3(2), pp.119–128. Available at: https://doi.org/10.1038/nrmicro1087.

Wolff, S. et al., 2008. Complementary analysis of the vegetative membrane proteome of the human pathogen Staphylococcus aureus. *Molecular & cellular proteomics : MCP*, 7(8), pp.1460–1468. Available at: https://www.ncbi.nlm.nih.gov/pubmed/18460691.

Woude, M.W. Van Der & Henderson, I.R., 1980. Regulation and Function of Ag43 (Flu). *J. Bacteriol*, (141), pp.858–867.

Yang, Z. et al., 2011. Highly Efficient Production of Soluble Proteins from Insoluble

Inclusion Bodies by a Two-Step-Denaturing and Refolding Method. *PLOS ONE*, 6(7), p.e22981. Available at: https://doi.org/10.1371/journal.pone.0022981.

Yin, J. et al., 2007. Select what you need: A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *Journal of Biotechnology*, 127(3), pp.335–347. Available at: http://www.sciencedirect.com/science/article/pii/S0168165606006237.

Yoon, S.H. & Kim, S.K.K. and J.F., 2010. Secretory Production of Recombinant Proteins in Escherichia coli. *Recent Patents on Biotechnology*, 4(1), pp.23–29. Available at: http://www.eurekaselect.com/node/93672/article.

Yue, J. et al., 2017. A new maltose-inducible high-performance heterologous expression system in Bacillus subtilis. *Biotechnology Letters*, 39(8), pp.1237–1244. Available at: https://doi.org/10.1007/s10529-017-2357-7.

Zheng, M., Åslund, F. & Storz, G., 1998. Activation of the OxyR Transcription Factor by Reversible Disulfide Bond Formation. *Science*, 279(5357), p.1718 LP-1722. Available at: http://science.sciencemag.org/content/279/5357/1718.abstract.

Zhu, J., 2012. Mammalian cell protein expression for biopharmaceutical production. *Biotechnology Advances*, 30(5), pp.1158–1170. Available at: http://www.sciencedirect.com/science/article/pii/S0734975011001674.

Ziemienowicz, A. et al., 1993. Both the Escherichia coli chaperone systems, GroEL/GroES and DnaK/DnaJ/GrpE, can reactivate heat-treated RNA polymerase. Different mechanisms for the same activity. *Journal of Biological Chemistry* , 268(34), pp.25425–25431. Available at: http://www.jbc.org/content/268/34/25425.abstract.

Zoufaly, S. *et al.* (2012) 'Mapping Precursor-binding Site on TatC Subunit of Twin Arginine-specific Protein Translocase by Site-specific Photo Cross-linking', *Journal of Biological Chemistry* , 287(16), pp. 13430–13441. doi: 10.1074/jbc.M112.343798 .

Annex 1: Published version of the paper *Escherichia coli* 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway

Escherichia coli 'TatExpress' strains export over 5g/L human growth hormone to the periplasm by the Tat pathway

Isabel Guerrero Montero[1†], Kirsty L. Richards[1†], Chillel Jawara[1], Douglas F. Browning[2], Amber R. Peswani[1], Mickael Labrit[3], Matthew Allen[3], Cedric Aubry[3], Emma Davé[3], David P. Humphreys[3], Stephen J. W. Busby[2] and Colin Robinson[1]*

[1] School of Biosciences, University of Kent, Ingram Building, Canterbury CT2 7NJ, UK

[2] Institute of Microbiology and Infection, School of Biosciences, University of Birmingham, Birmingham, B15 2TT, UK

[3] Discovery Biology, UCB Celltech, 216 Bath Road, Slough SL1 4EN, UK

[†] These authors contributed equally to this work.

* For correspondence:

Colin Robinson : Tel: +44 (0)1227- 823443. Email: C.Robinson-504@kent.ac.uk

Running title: High-level export of human growth hormone by the *E. coli* Tat pathway

Abstract

Escherichia coli is a heavily used platform for the production of biotherapeutic and other high-value proteins, and a favoured strategy is to export the protein of interest to the periplasm in order to simplify downstream processing and facilitate disulphide bond formation. The Sec pathway is the standard means of transporting the target protein but it is unable to transport complex or rapidly folding proteins because the Sec system can only transport proteins in an unfolded state. The Tat system also operates to transport proteins to the periplasm, and it has significant potential as an alternative means of recombinant protein production because it transports fully folded proteins. Here, we have tested the Tat system's full potential for the production of biotherapeutics for the first time using fed-batch fermentation. We expressed human growth hormone (hGH) with a Tat signal peptide in *E. coli* W3110 'TatExpress' strains that contain elevated levels of the Tat apparatus. This construct contained 4 amino acids from TorA at the hGH N-terminus. We show that the protein is efficiently exported to the periplasm during extended fed-batch fermentation, to the extent that it is by far the most abundant protein in the periplasm. The protein was shown to be homogeneous, disulphide bonded and active. The bioassay showed that the yields of purified periplasmic hGH are 5.4 g/L culture whereas an ELISA gave a figure of 2.39 g/L. Separate analysis of a TorA-hGH construct lacking any additional amino acids likewise showed efficient export to the periplasm, although yields were approximately 2-fold lower.

Key words

Biopharmaceuticals, recombinant protein, E. coli, Tat system, protein secretion

1.0 Introduction

Many high-value proteins are produced in *E. coli,* and a favoured strategy is to export the protein to the periplasm, usually by the well-characterised 'Sec' pathway (Walsh, 2014). Several strategies have been used, including expression of soluble proteins in the cytoplasm, expression in the form of insoluble inclusion bodies, or export to the periplasm. The latter approach is particularly favoured for the production of proteins that contain disulphide bonds, since the periplasm is the only oxidising compartment in wild type cells (Pooley *et al.*, 1996). Downstream processing is also simplified due to the lowered levels of cytoplasmic contaminants, including DNA (Balasundaram *et al.* (2009). Export of proteins to the periplasm is usually achieved by targeting via the well-characterised Sec pathway, whereby a cleavable Sec-specific signal peptide is present at the N-terminus of the target protein. However, the Sec pathway cannot transport some heterologous proteins and attention has focused on the alternative Tat pathway, which transports fully folded proteins by a completely different pathway (reviewed by Natale *et al.*, 2008). Initial studies showed that Tat could export a model protein, GFP, at high levels and that the cells were robust under fermentation conditions (Matos *et al.*, 2012). This work was followed by additional studies that showed Tat to be capable of exporting severa; biotherapeutics including human growth hormone, single chain antibody fragments and interferons (DeLisa *et al.*, 2003; Alanen *et al.*, 2015; Tullman-Ercek *et al.*, 2007; Browning *et al.*, 2017; Matos *et al.*, 2014).

Human Growth Hormone (hGH) is a 22 kDa protein consisting of 191 amino acids (Pooley *et al.*, 1996) used to treat burn injuries, wound healing, hypopituitarism and obesity (Isaksson *et al.*, 1985) since it is safe to use even in high doses (Van, 1998). Since hGH only has 2 disulphide bonds (Cys53-Cys165 and Cys182-Cys189) (Ultsch *et al.*,

1994) and no glycosylation, most of its large-scale production has used *E. coli* platforms, although a wide range of eukaryotic host systems have been tested (e.g. Hahm and Chung, 2001; Ecamilla-Trevino *et al.,* 2000). The highest recorded yield is 5 g of recombinant hGH per litre of milk in transgenic cows (Salamone *et al.,* 2006).

A major difficulty in producing hGH in *E. coli* is its aggregation into insoluble bodies that complicate the production process. Until recently, most commercial hGH has been obtained from insoluble bodies that have been solubilized, with the hGH refolded into its active form (Olsen *et al.,* 1981; Patra *et al.,* 2000). Recent, novel approaches have managed to produce high quantities of cytosolic hGH in fed-batch fermentation, with reported yields of up to 678 mg/L protein (Song *et al.,* 2017).

Although studies exporting hGH to the periplasm have reported relatively low yields (Sockolosky and Szoka, 2013), this is an attractive option due to the ease of downstream processing. In recent studies, we have shown that the Tat system can export hGH with high efficiency and that it is disulphide-bonded (Alanen *et al.,* 2015). Out study concluded that, although the protein was presumably synthesised and exported in the reduced state, the periplasmic protein was fully disuphide bonded by the normal DSB disulphide bonding system. We have also reported that a new series of engineered *E. coli* strains, which overexpress the Tat system components (termed TatExpress), demonstrate even higher export efficiencies (Browning *et al.* 2017). These strains bear a modified chromosomal *tatABC* operon, the expression of which is induced by IPTG, so that the Tat system is over-expressed at the same time as the IPTG-induced target protein. However, the activity of the hGH was not assayed and our studies (Alanen *et al.,* 2015; Browning *et al.,* 2017) used laboratory shake flask systems that bear little resemblance to

industrial production processes. Here, we have assessed the feasibility of this system for industrial production using TatExpress *E. coli* strains under fed-batch fermentation conditions. We report that the system is able to produce 5 g/L of purified, disulphide bonded, active hGH.

2.0 Materials and Methods

Bacterial strains and plasmids

Strains and plasmids are summarised in Table 1. *E. coli* strain W3110 was used, together with W3110 TatExpress which was previously described in Browning *et al.,* 2017 alongside the plasmid pKRK7 (pEXT22 expressing TorA-hGH-6His). pKRK38 was derived from pKRK7 to eliminate the TorA signal peptide that precedes the hGH. pKRK38 was made by PCR amplification using pKRK7 as a template and primers Mature_hGH_F (5'-ATGTTCCCAACCATTCCCTTATCCA-3') and Mature_hGH_R (5'- CATACATGTTCCTCTGTGGTAGGGT-3') as forward and reverse primers. PCR solutions, enzymes and reaction conditions were as described in Guerrero-Montero *et al.* (in press).

Fed-batch fermentation

Initial cultures were grown in 50 mL of 2xP media (16g/L of Bacto-tryptone/peptone, 10g/L of Yeast extract, 10 g/L NaCl) in 250 mL shake flasks for 6 h at 37°C, 200 rpm with 1:1000 antibiotic (5 µL, 1 M kanamycin). 1 mL of culture was transferred to 200 mL of SM6 defined media (Humphreys *et al*., 1998) and grown aerobically overnight at 30 °C, 200 rpm in shake flasks with 1:1000 antibiotic (5 µL, 1 M kanamycin). On the following day an equivalent of 300 OD_{600} was used to inoculate fresh defined SM6 media to a final volume of 500 mL in Infors Multifors 1.5L fermenters (Infors UK Ltd., Reigate,

UK). The pH was kept at 7 using 25% (v/v) ammonia solution and 25% (v/v) sulphuric acid. Dissolved oxygen tension (DOT) was kept at 30% using gas blending with 100% oxygen where necessary and the culture was maintained at 30°C until both stirrer and airflow was maximal and then dropped to 25°C. Supplementation of $MgSO_4$ occurred when the OD_{600} reached 38-42 (8 mL/L of 1M $MgSO_4$) and of Na_2HPO_4 when the OD_{600} reached 54-58 (5 mL/L of 232.8 g/L Na_2HPO_4) and when the OD_{600} reached 66-77 (7 mL/L of 232.8 g/L Na_2HPO_4). A glycerol feed containing 80% w/w glycerol at a rate of 0.35% pump capacity was started at OD_{600} 70. Induction with 9 ml/L of IPTG at a concentration of 4.31 g/L occurred at OD_{600} 75.

Fractionation

Cytoplasm, membrane and periplasm fractions were prepared by an osmotic-shock method. Cells were centrifuged at 3000 x g, 4°C, for 10 min and resuspended in 500 µL of Buffer 1 (100 mM Tris-acetate pH 8.2, 500 mM sucrose, 5 mM EDTA pH 8.0) and 500 µL dH_2O before addition of 40 µL hen egg white lysozyme (1 mg/mL) and incubation on ice for 5 min. 20 µL $MgSO_4$ (1M) was added and the suspension was centrifuged at 14,000 rpm, 4 ° C, 2 min. The supernatant containing the periplasmic fraction was removed and the pellet washed once in 750 µL Buffer 2 (50 mM Tris-acetate pH 8.2, 250 mM sucrose, 10 mM $MgSO_4$). The cell pellet was resuspended in 750 µL Buffer 3 (50 mM Tris-acetate pH 8.2, 2.5 mM EDTA pH 8.0) and sonicated for 6 × 10 s, amplitude 8 µm to disrupt membranes (Soniprep 150plus, Sanyo Gallenkamp, Loughborough, UK). The suspension was centrifuged at 70,000 rpm, 4 °C, 30 min to sediment the insoluble fraction. 500 µL of the supernatant was removed (designated cytoplasmic fraction). The remainder of the supernatant was discarded and the pellet resuspended in 500 µL Buffer 3 and designated membrane (or insoluble) fraction.

Protein purification

For purification of 6x Histidine-tagged (C-term) proteins by Nickel IMAC, 10 mL of the culture post induction was taken, centrifuged at 3000 rpm, 45 min, 4°C. Cell pellet was resuspended in 10 mL/g of chilled Buffer 1 without EDTA and 10 mL/g of milliQ H_2O and incubated on ice for 30 min before centrifuging at 14 000 rpm, 20 min, 4°C (Beckman Avanti J- 25, JA 25.5 rotor). Supernatant was taken as the periplasmic fraction and placed into SnakeSkin® dialysis tubing (Thermo scientific) and dialysed at 4°C overnight into 50 mM sodium phosphate, 150 mM NaCl, pH 7.2. Using an ÄKTA™ pure protein purification system and a HisTrap HP histidine-tagged protein column (GE Healthcare, Buckinghamshire, UK) the protein was purified: storage solution (20% ethanol) was washed off with 10 column volumes (CV) of milliQ H_2O before adding 5 mL 0.2 M NiCl, followed by another 2CV milliQ H_2O wash. Columns were equilibrated with 3CV equilibration buffer (50 mM sodium phosphate pH 7.2, 0.3 M NaCl) before loading Periplasmic sample and collecting Flow Through (FT). Unbound matter was removed with 6CV Wash buffer (50 mM sodium phosphate pH 7.2, 50 mM Imidazole, 0.3 M sodium chloride) and sample collected as Wash (W). Finally, the 6x Histidine-tagged protein was eluted with elution buffer with imidazole (50 mM sodium phosphate pH 7.2, 0.3 M NaCl with 250 mM Imidazole) and the peaks collected for analysis.

Protein expression analysis by Coomassie and Western blot

Protein samples were resolved by reducing SDS–PAGE and analyzed using Coomassie blue staining and Western blotting. All immunoblottingh procedures were as described in Guerrero-Montero *et al.* (in press). Periplasmic hGH activity was assayed using an hGH bioassay (PathHunter® Human Growth Hormone Bioassay Kit, Sigma) using the manufacturer's protocol. Standardized OD10 periplasmic fractions were diluted 1:500,000 using PBS and absorbance was read using a BMG Labtech Spectrostar microplate reader at 405 nm, with a reference wavelength at 490 nm. Concentrations were

calculated from two independent experiments and all samples were measured in triplicate when calculating periplasmic yield. Periplasmic hGH was also assayed using an hGH ELISA kit (Roche Diagnostics, Charles Ave, Burgess Hill, West Sussex RH15 9RY) using the manufacturer's protocol. Concentrations were calculated from two independent experiments and all samples measured in triplicate, and used to calculate average periplasmic yield (in mg/L) by reference to culture OD readings at 600 nm.

Intact Protein Electrospray LC-MS

The electrospray mass spectrum was recorded on a Bruker micrOTOF-Q II mass spectrometer. An aliquot of protein in solution, corresponding to approximately 20 picomoles of protein, was desalted on-line by reverse-phase HPLC on a Phenomenex Jupiter C4 column (5 μm, 300Å, 2.0 mm x 50 mm) running on an Agilent 1100 HPLC system at a flow rate of 0.2 ml/min using a short water, acetonitrile, 0.05% trifluoroacetic acid gradient. The eluent was monitored at 280 nm and directed into the electrospray source, operating in positive ion mode, at 4.5 kV and mass spectra recorded from 500-3000 m/z. Data were analysed and deconvoluted to give uncharged protein masses using Bruker's Compass Data Analysis software.

3.0 Results

3.1 Fed-batch fermentation of WT and TatExpress cells expressing TorA-hGH

We first expressed a construct comprising the TorA signal peptide linked to hGH (TorA-hGH) in W3110 *E. coli* TatExpress cells under fed-batch fermentation conditions as detailed in Materials and Methods. The construct was previously used in shake flask studies (Browning *et al.*, 2017) and comprises the signal peptide of *E. coli* TMAO

reductase (TorA) linked to hGH via a 4-residue linker (the first 4 residues of mature TorA; AQAA). The construct also contains a C-terminal 6-His tag. Parallel cultures were carried out with the construct expressed in wild type (WT) W3110 cells. Figure 1 shows the growth curves from the cultures; duplicate WT and TatExpress cultures reached ODs of about 100 and the Figure shows the point at which induction of TorA-hGH and TatABC synthesis (in TatExpress) was induced using IPTG.

Samples were removed 13, 17 and 21h after induction and the cells were fractionated to generate cytoplasmic, membrane and periplasmic samples. The periplasmic samples were analysed by immunoblotting to detect the hGH and the same samples were analysed using Coomassie stained SDS PAGE gels to analyse the proteome of the periplasmic samples. Figure 2 shows the 13h and 21h samples from the two WT cultures, which indicate the presence of a clear hGH protein band in the periplasm. The Coomassie-stained gel shows the presence of a 22 kDa protein, the relative abundance of which matches the strength of the blot signals. This was confirmed to be hGH (see below). 13h, 17h and 21h samples from the two TatExpress cultures were also analysed and the data show that the hGH blot signals are significantly higher than those of the WT cultures, confirming that TatExpress cells export this construct with much higher efficiency as observed by Browning *et al.* (2017) in shake flask culture. The abundance of the 22 kDa protein is correspondingly greater in the Coomassie stained gels, providing further evidence that this is indeed hGH.

Figure 3 shows a time course analysis of the abundance of the periplasmic hGH after induction of synthesis in TatExpress cells. Samples were taken from 2 - 51 h after induction and the blot shows a steady increase in hGH level over this time period. Equal numbers of cells were used in each sample, so this reflects an increase in the amount of

hGH per cell. The Coomassie gel confirms that the abundant 22 kDa protein is indeed hGH, since its abundance increases in parallel and the protein is virtually absent in the induction time point samples. Analysis of the later time points demonstrates that hGH is by far the most abundant periplasmic protein, which shows that Tat is exporting high amounts of protein.

We have carried out controls for fractionation artefacts in previous studies, and the periplasmic proteome is clearly distinct from that of the cytoplasm, suggesting that contamination of cytoplasmic proteins is minimal (e.g. Matos *et al.,* 2012). However, to confirm this point we fractionated samples at the 21 h time point and the data are shown in Figure 4. The results show that the fractions are indeed 'clean' with the periplasmic proteome clearly distinct from those of the cytoplasm and membrane fractions. To further confirm that hGH does not reach the periplasmic fraction 'spontaneously' we expressed hGH lacking any form of signal peptide. Mature hGH was expressed in parallel with Tor-hGH and samples were fractionated after 1h and 18 h induction (Figure 5). The immunoblot shows that the bulk of hGH is in the periplasm in the TorA-hGH culture, as expected, with a minimal level of protein present in the cytoplasm at the 18h point. In contrast, hGH is found exclusively in the cytoplasm in the sample expressing mature hGH, with none detected in the periplasmic fraction. These data confirm that the presence of hGH in the periplasm of the TorA-hGH cultures is due to Tat-dependent export and not contamination by cytoplasmic fraction or spontaneous transfer across the plasma membrane.

It is notable that the total level of hGH in the culture producing mature hGH is about 6-fold lower than that in the TorA-hGH culture. This reflects a phenomenon observed

previously in our shake flask studies: that hGH is subjected to rapid turnover in the cytoplasm (Alanen *et al.*, 2015). Clearly, this protein is more stable in the periplasm, and this is another advantage of using an effective protein export strategy for the production of this particular protein.

3.2 Tat-exported hGH is homogeneous and cleaved at the correct signal peptidase site

To assess the homogeneity of the exported hGH, we performed mass spectrometry analysis of the purified protein, using purified commercial hGH as a standard (provided with the Pathfinder Bioassay - see below). The exported, periplasmic hGH construct used in this study should contain additional amino acids when compared to commercial hGH, specifically the 4 amino acid linker from the mature TorA protein (TMAO reductase) at the N-terminus, the initiation methionine (cleaved from hGH *in vivo*) and the 6-His tag at the C-terminus (Figure 6). The combined molecular weight of the additional amino acids is 609.7 for the N-terminal residues and 822.86 for the C-terminal His tag (1432.56 in total). The predicted molecular weights are 22129.05 for the commercial hGH, and 23561.5 for the exported hGH. As shown in Figure 6, mass spectrometry analysis of commercial hGH gives a single prominent peak corresponding to 22124.3 Da, which matches almost exactly with the predicted molecular mass, having only a difference of 4.75 Da, 4 of which are a consequence of the formation of 2 disulphide bonds. We purified the Tat-exported periplasmic hGH using affinity chromatography and an analysis of this protein again shows a single prominent peak, and the mass of 23555 Da is again very close to, but 6 Da smaller than the predicted mass of 23561 Da. The presence of 2 disulphide bonds again accounts for the difference. In both cases, the 1-2 Da deviation from the predicted protein masses is within the error range for the mass spectrometer.

Other smaller peaks are observed in both samples, which appear to be an artefact of the hGH mass spectrometry since they are identical in the two samples. This implies that the hGH produced in this study is (i) as homogeneous as the commercial protein (ii) processed at the correct site, and (iii) fully disulphide bonded; there is no indication of a peak corresponding to the reduced form.

To confirm that the hGH is indeed disulphide-bonded, we ran samples of the periplasmic samples from the 51 h time point shown in Figure 3 on an SDS-PAGE gel in the presence and absence of reducing agent. The oxidised, disulphide-bonded form of hGH was shown to migrate more rapidly than the reduced form (Alanen *et al.*, 2015) and Figure 7 shows the periplasmic hGH from the fed-batch fermentation likewise runs more rapidly under oxidising conditions, further confirming that the protein is disulphide-bonded. It is unclear why the intensity of the blot signal is lower; presumably, the C-terminal His tag is less exposed in the oxidised protein and the antibodies bind less effectively.

3.3 High yields of active protein are obtained from fed-batch fermentation using TorA-hGH

To quantify the yield of periplasmic hGH and to assess its activity, we used a commercially available hGH bioassay (PathHunter® Human Growth Hormone Bioassay Kit, Sigma). This uses engineered cells in which one fragment of β-galactosidase is present on the hGH receptor and the complementary fragment is present on a phosphor-tyrosine SH2-domain-containing protein that is only able to bind the hGH receptor once it is activated. Binding of hGH to its receptor results in receptor phosphorylation by a

cytosolic tyrosine kinase such as JAK1, enabling the SH2-EA fusion protein to bind the phosphorylated receptor and generate active β-galactosidase. Enzymatic activity is quantitatively measured using a chemiluminescent substrate, and the expected result is a dose-response sigmoidal function where the phosphorylation of the receptor will be dependent on the amount of hGH in the sample. A standard curve was produced with the supplied commercial hGH and all samples were done in triplicate and with dilution factors in the range of the standard curve (see Supplementary Figure 2). This assay gave a figure of 5.4 g hGH per litre of fed-batch culture for the purified periplasmic hGH shown in Supplementary Figure 1.

We also calculated hGH levels using a commercial ELISA, which uses pre-bound antibodies to hGH on the surface of the microplate modules. The sample to analyse was added in triplicate (with various dilution factors) to the wells, and the hGH present in the samples binds to the anti-hGH antibodies. Afterwards a digoxigenin-labelled antibody to hGH is added which binds to the hGH as well. An antibody to digoxigenin conjugated to peroxidase is added and binds to the digoxigenin, and finally the peroxidase substrate ABTS is added. The peroxidase catalyses the cleavage of the substrate yielding a coloured reaction product that can be measured using a plate reader that reads at 405 and 490 nm. The absorbance is directly correlated to the level of hGH present in the sample and can be determined by comparison to the calibration curve (shown in Supplementary Figure 3). Using this ELISA the concentration of the purified periplasmic hGH was calculated to be 2.39 g/L culture. This is lower than the figure obtained using the bioassay, but since the assays are so different it is difficult to determine which is more accurate.

3.4 Export of TorA-hGH lacking the 4-residue linker between the signal peptide and hGH

The above experiments were all carried out using a construct that comprises the TorA signal peptide, 4 residues from the N-terminus of mature TorA and hGH. This construct was used in order to compare results with shake flask studies carried out using the same construct (Browning *et al.*, 2017), but we considered it important to assess the export of a 'linker-less' construct lacking any linker residues, since industrial processes are unlikely to incorporate such linkers. We therefore removed the 4 residues and carried out fed-batch fermentation studies on this 'clean' construct; the periplasmic fractions are analysed in Figure 8). The immunoblot and gel data show that hGH production is induced and that high levels of hGH are exported to the periplasm over the 38 h induction period. As with the original construct, periplasmic hGH accumulates to the extent that it is an abundant periplasmic protein. The data thus show that removal of the 4-residue linker does not block export of hGH.

Figure 8C shows a test in which fermentation samples were fractionated into cytoplasm, membrane and periplasm extractss, which were then blotted for the presence of hGH. The data show that in each case, the majority of hGH is present in the preiplasm, confirming that both constructs are efficiently exported. However, the hGH signal after export of the linker-less construct is less intense than that of the original construct containing the additional 4 residues. In addition, analysis of the stained gel in Figure 8A shows that while the periplasmic hGH is a major band, the abundance is not as striking as that of thye periplasmic accumulation of hGH obser4ved in Figures 2 and 3. We therefore believe that the linker-less hGH is exported with lowered efficiency, and quantification

of the blots suggests that, one average, its export is in the region of 40% as efficient as that of the original TorA-hGH construct.

4.0 Discussion

The Tat system has been proposed to offer a viable alternative to the Sec pathway for the export of high-value proteins to the bacterial periplasm, but its potential for the production of biotherapeutic proteins has not been fully explored in previous studies. This is primarily because the majority of those studies were carried out using laboratory shake flask culture systems, whereas industrial processes almost invariably use fed-batch fermentation. Here, we have used fed-batch fermentation systems to test the robustness of the TatExpress cells and the yields of hGH that are obtained after export to the periplasm.

A major aim in this work was to test for any deleterious effects due to the increased expression of the TatABC membrane proteins in the TatExpress strain. While the TatABC proteins are expressed to a much lower extent than TorA-hGH, the increased expression of any membrane protein can potentially lead to cell stress and lowered productivity in extended fed-batch fermentation systems. Here, we directly compared the growth characteristics of the TatExpress strain with its parental W3110 strain and we observed no significant differences. Clearly, TatExpress strains are viable production hosts.

In terms of target protein yield, we consistently observe that the cells export TorA-hGH throughout an extended induction period and it is notable that the exported hGH is by far

the most abundant periplasmic protein by the end of the induction period. In control tests, mature-size hGH is not exported at all and we can conclude that TorA-hGH is being exported by the TorA signal peptide as reported in shake-flask studies (Alanen *et al.*, 2015). The yields of protein are high: the bioassay indicates a yield of over 5 g/L active purified protein while the ELISA gives a figure of 2.39 g/L. This is one of the highest yields reported and clear evidence that this platform has potential for industrial use. Nevertheless, further study would serve to illustrate the potential in more detail, and it is notable that, while a 'clean' fusion of a Tat signal peptide and hGH is efficiently exported, the presence of a short linker region appears to enhance export. An optimal signal peptide-passenger protein junction may be an important factor for efficient Tat-dependent export, as previously suggested by Tullman-Ercek *et al.*, 2007.

In summary, we have shown that a model biotherapeutic protein can be exported by the Tat system in high amounts, and that it is homogeneous, disulphide-bonded and active. It will be of interest to conduct further studies to assess the full capability of the Tat system, and in particular to explore its potential for the export of more complex proteins.

ACKNOWLEDGEMENTS

This work was generously supported by Industrial Biotechnology Catalyst (Innovate UK, BBSRC, EPSRC) grant BB/M018288/1 to support the translation, development and commercialisation of innovative Industrial Biotechnology processes. The work was also supported by BBSRC Global Challenges Research Fund grant BB/P02789X/1. We are grateful to Richard Davies and Geoff Brown for fermentation expertise (USP development 1, Biotech Sciences, UCB Celltech).

CONFLICT OF INTEREST

The authors declare no competing financial or other interests.

REFERENCES

Alanen HI, Walker KL, Lourdes Velez Suberbie M, Matos CF, Bonisch S, Freedman RB, Keshavarz-Moore E, Ruddock LW, Robinson C. 2015. Efficient export of human growth hormone, interferon alpha2b and antibody fragments to the periplasm by the *Escherichia coli* Tat pathway in the absence of prior disulfide bond formation. Biochim Biophys Acta 1853(3):756-63.

Balasundaram B, Harrison S, Bracewell DG. 2009. Advances in product release strategies and impact on bioprocess design. Trends Biotechnol 27(8):477-85.

Browning DF, Richards KL, Peswani AR, Roobol J, Busby SJW, Robinson C. 2017. *Escherichia coli* "TatExpress" strains super-secrete human growth hormone into the bacterial periplasm by the Tat pathway. Biotechnol Bioeng. 114, 2828-2836.

DeLisa MP, Tullman D, Georgiou G. 2003. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. Proc Natl Acad Sci U S A 100(10):6115-20.

Guererro Montero, I., Dolata, K.M., Schlüter, R., Malherbe, G., Sievers, S., Zühlke, D., Sura, T., Dave, E., Riedel, K. and Robinson, C. (2019). Comparative proteome analysis in an Escherichia coli CyDisCo strain identifies stress responses related to protein production, oxidative stress and accumulation of misfolded protein. *Microb. Cell Fact.* (in press).

Ecamilla-Trevino LL, Viader-Salvado JM, Barrera-Saldana HA, Guerrero- Olazaran M. 2000. Biosynthesis and secretion of recombinant human growth hormone in Pichia pastoris, Biotechnol. Lett. 22; 109–114.

Hahm MS, Chung BH. 2001. Secretory expression of human growth hormone in Saccharomyces cerevisiae using three different leader sequences, Biotechnol. Bioprocess Eng. 6; 306–309.

Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, Ohtsubo E, Baba T, Wanner BL, Mori H and others. 2006. Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. Mol Syst Biol 2:2006.0007.

Humphreys DP, Vetterlein, OM, Chapman AP, et al. 1998. F(ab0)2 molecules made from E. coli produced Fab0 with hinge sequences conferring increased serum permanence times in an animal model. J. Immunol. Meth. 217:1 – 10.

Isaksson OG, Edén S, Jansson JO. 1985. Mode of action of pituitary growth hormone on target cells. Ann Rev Physiol; 47:483-99

Matos CF, Branston SD, Albiniak A, Dhanoya A, Freedman RB, Keshavarz-Moore E, Robinson C. 2012. High-yield export of a native heterologous protein to the periplasm by the tat translocation pathway in *Escherichia coli*. Biotechnol Bioeng 109(10):2533-42.

Matos CF, Robinson C, Alanen HI, Prus P, Uchida Y, Ruddock LW, Freedman RB, Keshavarz-Moore E. 2014. Efficient export of prefolded, disulfide-bonded recombinant proteins to

the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains. Biotechnol Prog 30(2):281-90.
Natale P, Bruser T, Driessen AJ. 2008. Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms. Biochim Biophys Acta 1778(9):1735-56.
Olson KC, Fenno J, Lin N, Harkins RN, Snider C, Kohr WH, Ross MJ, Fodge D, Prender G, St ebbing N. 1981. Purified human growth hormone from *E. coli* is biologically active. Nature 293:408-11.
Patra AK, Mukhopadhyay R, Mukhija R, Krishnan A, Garg LC, Panda AK. 2000. Optimization of inclusion body solubilization and renaturation of recombinant human growth hormone from *Escherichia coli*. Protein Exp Purification 18:182-92;
Pooley H. M., Merchante R., Karamata, D. 1996. Overall Protein Content and Induced Enzyme Components of the Periplasm of Bacillus subtilis. Microbial Drug Resistance, 2 (1): 9-15.
Salamone D, Baranao L, Santos C, Bussmann L, Artuso J, Werning C, Prync A, Carbonetto C, Dabsys S, Munar C, Salaberry R, Berra G, Berra I, Fernandez N, Papouchado M, Foti M, Judewicz N, Mujica I, Munoz, L, Alvarez, SF. Gonzalez E, Zimmermann J, Criscuolo M, Melo C. 2006. High level expression of bioactive recombinant human growth hormone in the milk of a cloned transgenic cow, J. Biotechnol. 124; 469–472.
Sambrook J, Russell DW. 2001. Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.
Sockolosky JT, Szoka FC. 2013. Periplasmic production via the pET expression system of soluble, bioactive human growth hormone. Protein Expression and Purification, 87(2), 129–135
Song H, Jiang J, Wang X, Zhang J. 2017. High purity recombinant human growth hormone (rhGH) expression in *Escherichia coli* under *phoA* promoter, Bioengineered, 8:2, 147-153
Tullman-Ercek D, DeLisa MP, Kawarasaki Y, Iranpour P, Ribnicky B, Palmer T, Georgiou G. 2007. Export pathway selectivity of *Escherichia coli* twin arginine translocation signal peptides. J Biol Chem 282(11):8309-16.
Ultsch MH. Somers AA, Kossiakoff AM, Devos AM. 1994. The crystal-structure of affinity-matured human growth-hormone at 2-Angstrom resolution, J. Mol. Biol. 236; 286–299.
Van LK. 1998. Safety of high doses of recombinant human growth hormone. Hormone Res; 49:78-81;
Walsh G. 2014. Biopharmaceutical benchmarks 2014. Nat Biotechnol 32:992-1000.

Figure Legends

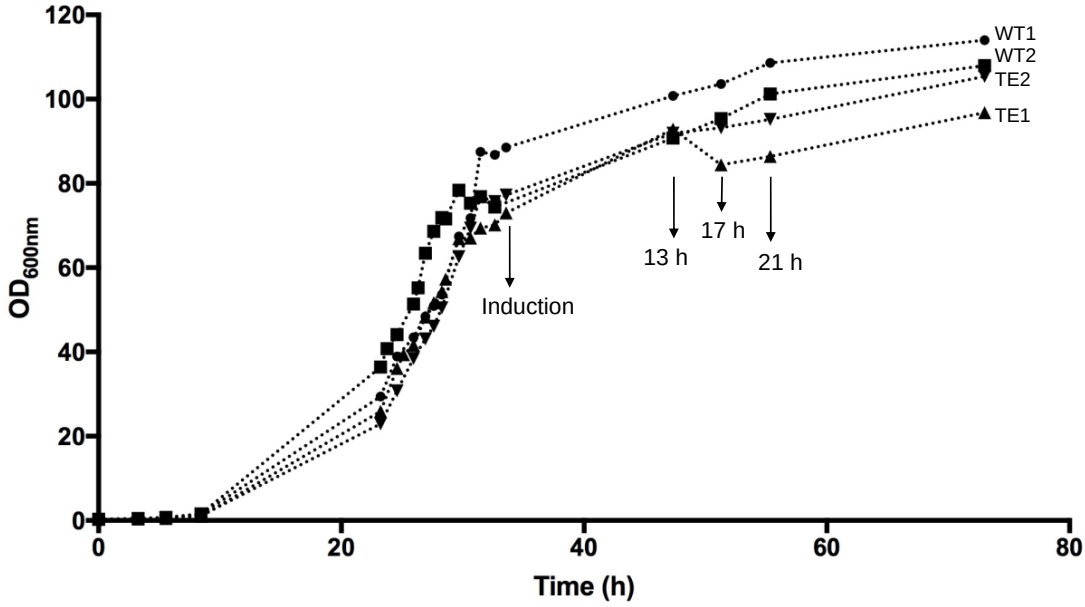


Figure 1. Growth data during fed-batch fermentation of *E. coli* W3110 WT and TatExpress cells expressing TorA-hGH. Duplicate cultures were analysed of WT cells (WT1, WT2) and TatExpress (TE1, TE2), with $OD_{600\,nm}$ values shown. Fermentation was carried out at 30°C as detailed in Materials and Methods. At the indicated times, the cultures were induced with IPTG (0.1 mM) and samples were removed for analysis.

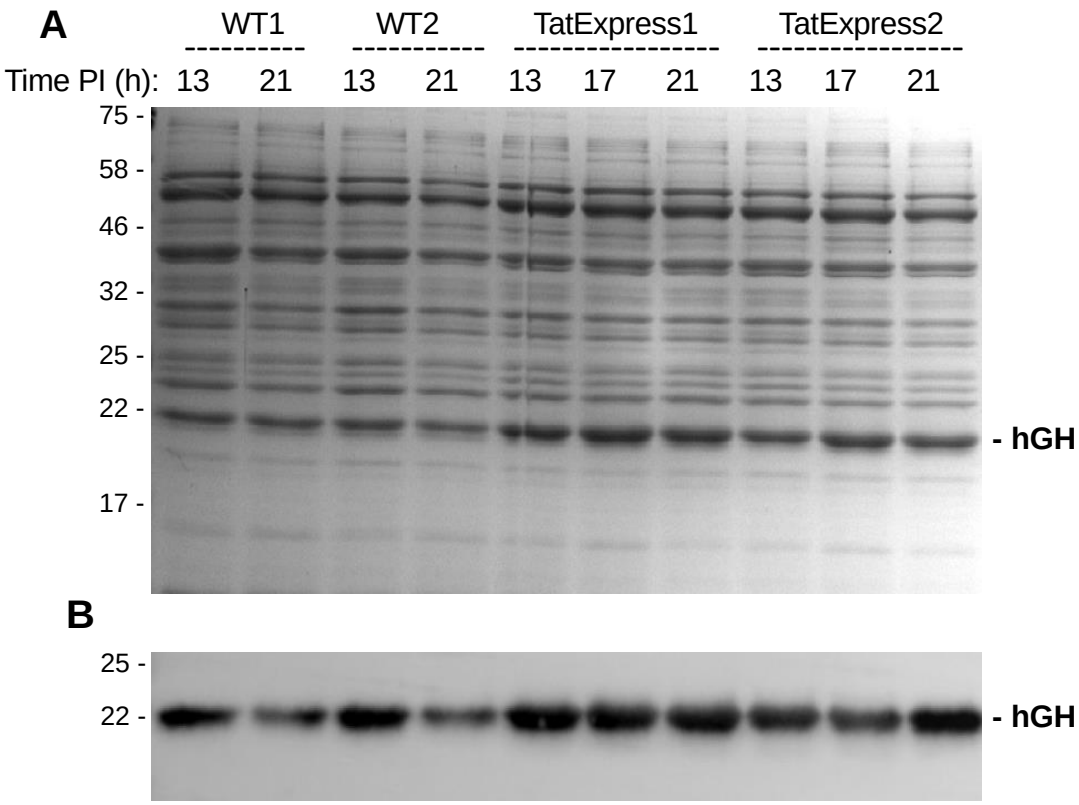


Figure 2. Tat-dependent export of TorA-hGH in W3110 wild type and TatExpress W3110 cells. Samples from duplicate fed-batch fermentation cultures of WT and TatExpress expressing TorA-hGH were removed at the indicated times post induction and fractionated to yield periplasmic samples. The periplasmic samples were analysed by Coomassie blue stained gels (A) and immunoblotting using antibodies to the His tag on hGH (B). For each lane a normalized amount of protein has been loaded, equivalent to OD_{600} 0.08 AU total cells (Coomassie blue stained gels), and OD_{600} 0.008 AU (immunoblots). Mobilites of molecular weight markers (in kDa) are shown on the left.

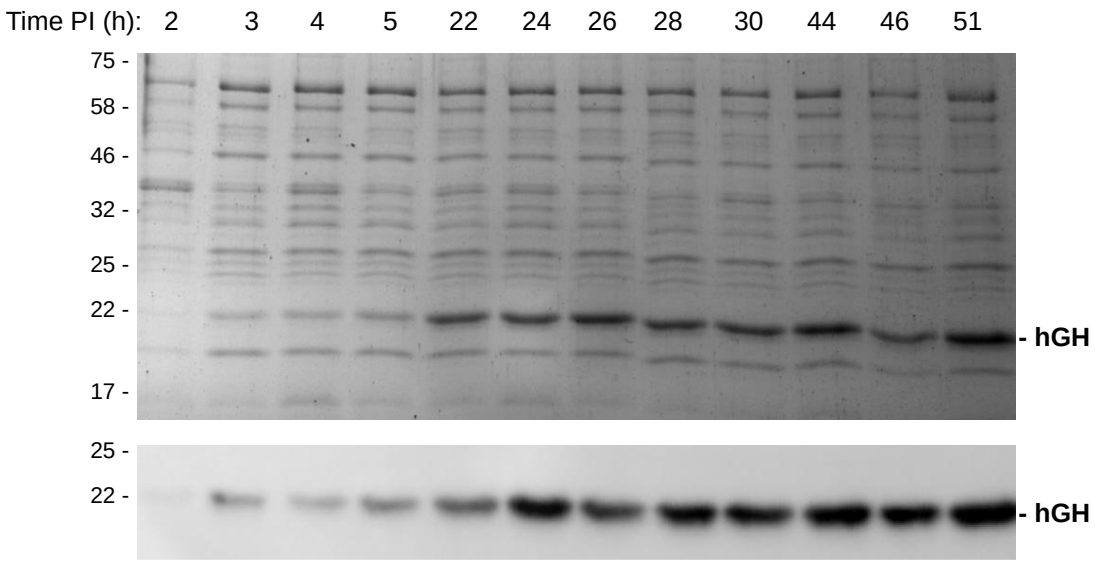


Figure 3. High-level export of TorA-hGH during extended fed-batch fermentation of TatExpress cells. TorA-hGH was expressed in TatExpress cells in a fed-batch fermentation system and samples were removed at the indicated times post induction (PI). The samples were fractionated to generate periplasm samples which were then analysed on Coomassie blue stained gels (A) and by immunoblotting using antibodies to the His tag on hGH (B).

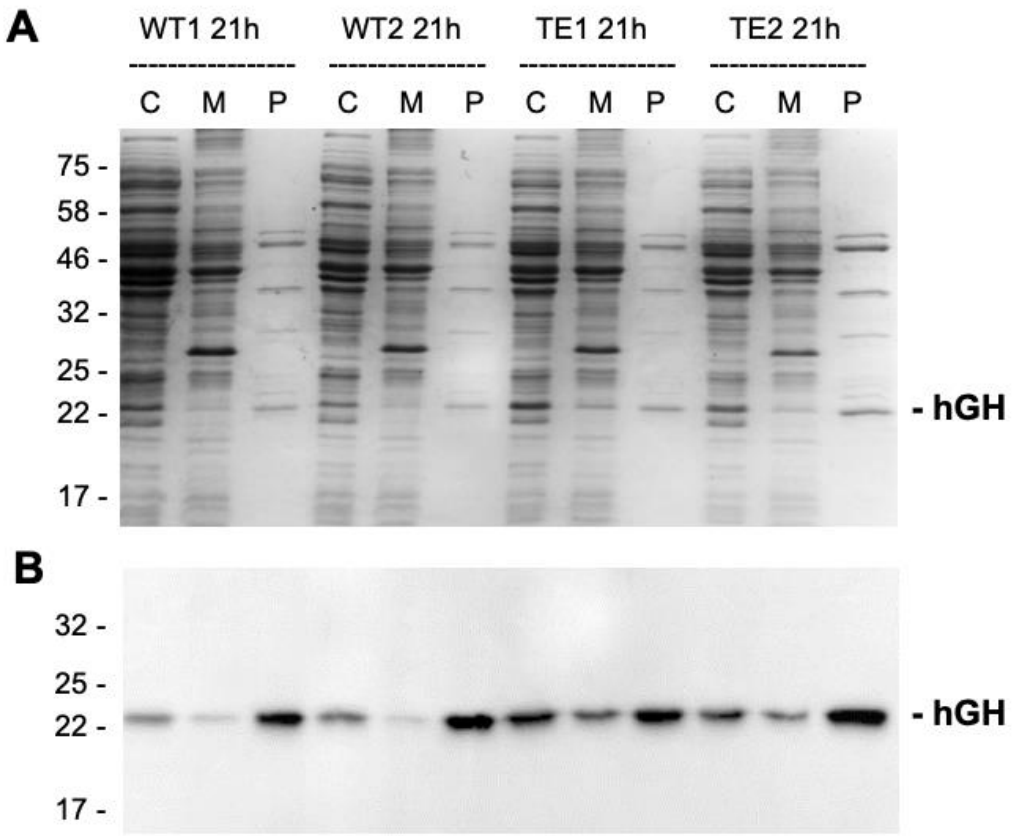


Figure 4. Fractionation of samples from fed-batch fermentation of W3110 wild type and TatExpress cells expressing TorA-hGH. Samples from the 21 h time point of fed-batch fermentation cultures of WT and TatExpress (TE) expressing TorA-hGH (shown in Figure 2) were fractionated to yield cytoplasm, membrane and periplasm samples (C, M, P). The samples were analysed by Coomassie blue stained gels (A) and immunoblotting using antibodies to the His tag on hGH (B). For each lane a normalized amount of protein has been loaded, equivalent to OD600 0.08 AU total cells (Coomassie blue stained gels), and OD600 0.008 AU (immunoblots). Mobilites of molecular weight markers (in kDa) are shown on the left

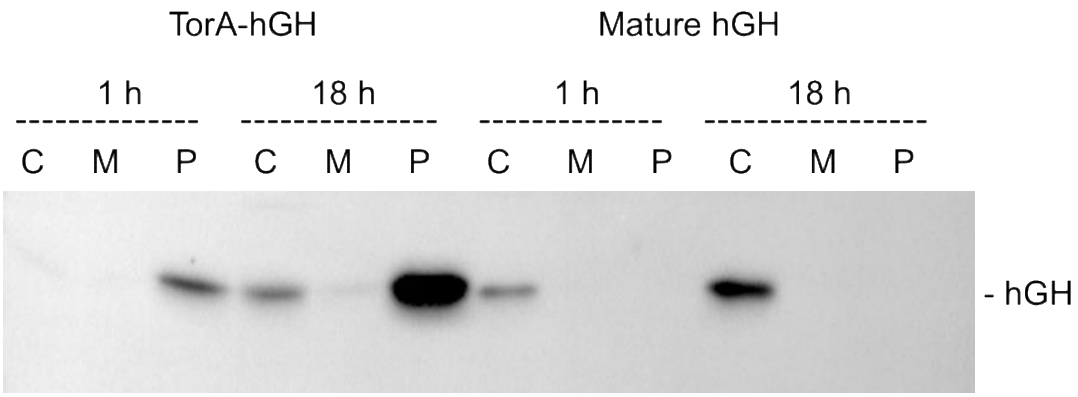
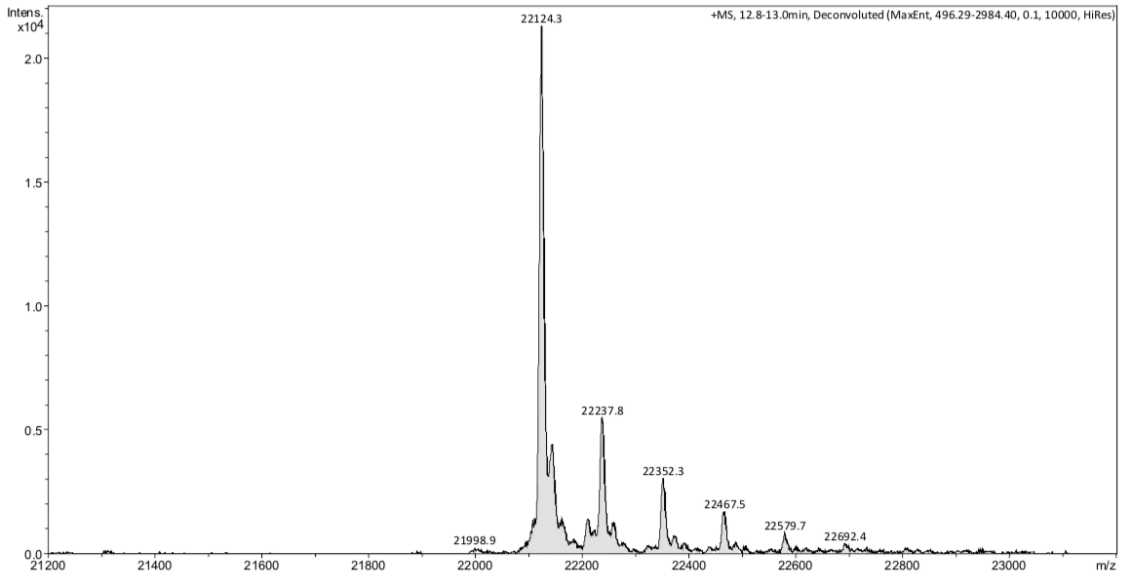
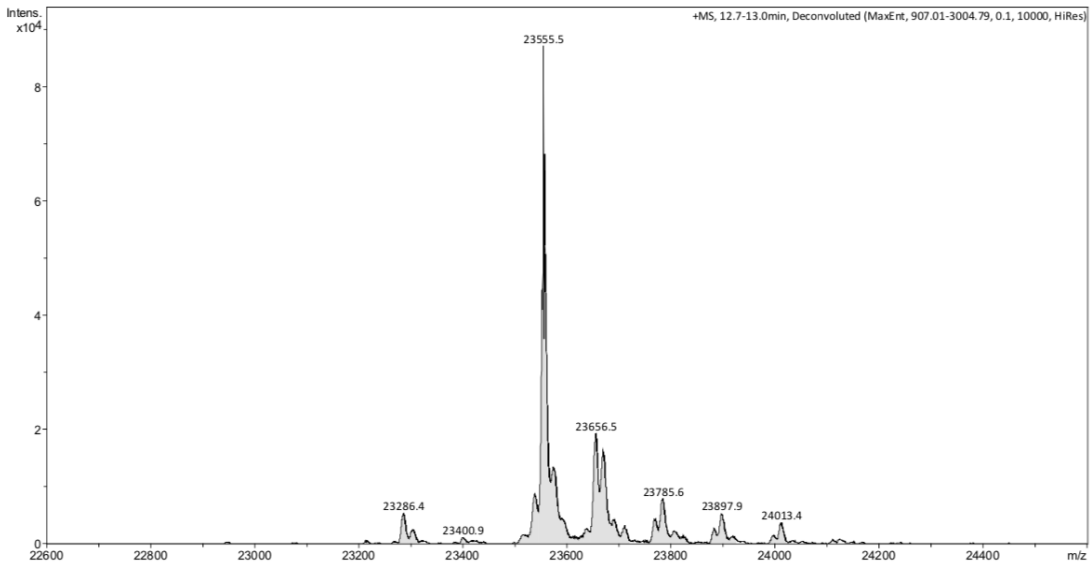


Figure 5. Mature hGH is not exported during fed-batch fermentation of TatExpress. Mature-size hGH and TorA-hGH were expressed in TatExpress cells using fed-batch fermentation systems. Samples were removed 1 h and 18 h post induction and fractionated to generate cytoplasm, membrane and periplasm samples (C, M, P) which were then analysed by immunoblotting using antibodies to the His tag on hGH.

A



B



C

NNNDLFQASRRRFLAQLGGLTVAGMLGPSLLTPRRATA♦AQAAHMFPTIPLSR LFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLCFSESIPTPSN REETQQKSNLELLRISLLLIQSWLEPVQFLRSVFANSLVYGASDSNVYDLLKDLE EGIQTLMGRLEDGSPRTGQIFKQTYSKFDTNSHNDDALLKNYGLLYCFRKDMD KVETFLRIVQCRSVEGSCGF**HHHHHH**

Figure 6: Exported periplasmic hGH is homogeneous and cleaved at the correct site. Periplasmic hGH was purified by affinity chromatography (see Supplementary Figure 1) and subjected to mass spectrometry analysis as detailed in Materials and Methods. Commercial hGH was analysed in an identical manner. The mass spectra for commercial hGH is shown in panel A and for periplasmic hGH in B. C: amino acid sequence of TorA-hGH-H6. The C-terminal half of the TorA signal peptide is shown, along with the first five amino acids of the mature TMAO protein and the initial methionine for the hGH protein. The signal peptidase cleavage site between ATA and

AQA is denoted by ♦. The sequence for the commercial hGH is underlined, and this is followed by the 6-histidine tag.

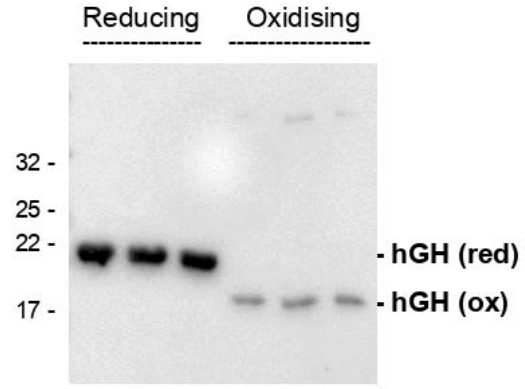


Figure 7. Periplasmic hGH is disulphide-bonded after fed-batch fermentation. Samples of the periplasmic fraction from the 51 h time point shown in Figure 3 were run (in triplicate) on an SDS polyacrylamide gel under standard reducing conditions (‘reducing’ or in the absence of reducing agent (oxidising’). The samples were immunoblotted using antibodies to the C-terminal His tag on hGH. hGH ‘ox’ and ‘red’: oxidised and reduced forms of hGH, respectively. Mobilties of molecular mass markers (in kDa) are shown on the left.

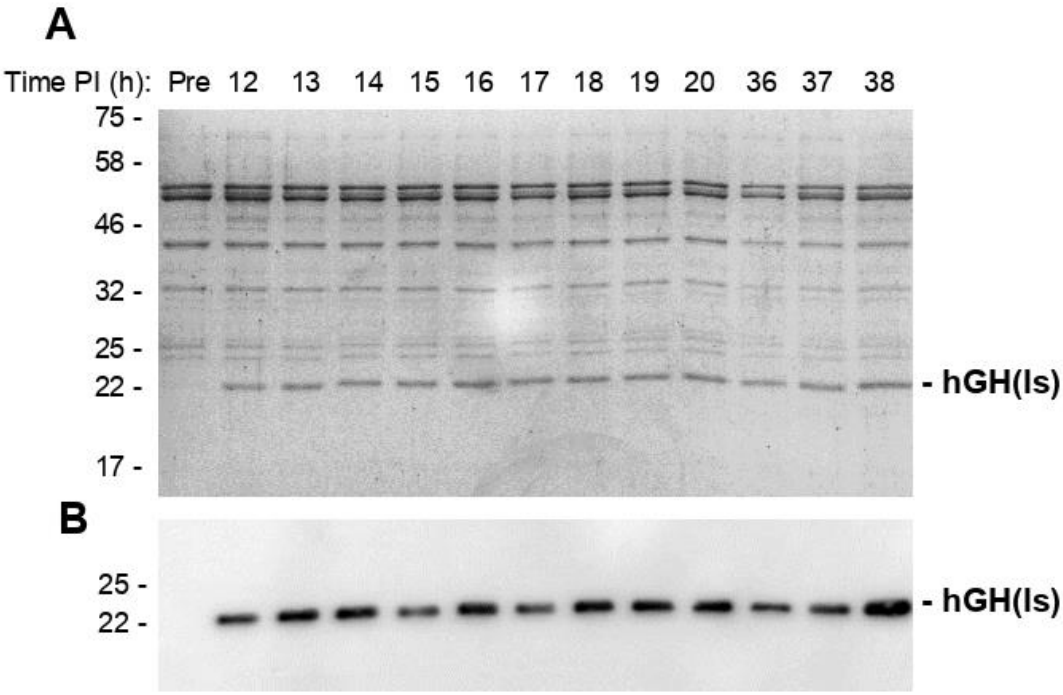


Figure 8. Export of 'linkerless' TorA-hGH during fed-batch fermentation of TatExpress cells. TorA-hGH lacking a 4-residue linker was expressed in TatExpress cells in a fed-batch fermentation system and samples were removed at the indicated times post induction (PI). The samples were fractionated to generate periplasm samples which were then analysed on Coomassie blue stained gels (A) and by immunoblotting using antibodies to the His tag on hGH (B). C: TorA-hGH constructs with or without the 4-residue linker were expressed in fed-batch fermentation with 40 h induction and fractionated to generate cytoplasm, membrane and periplasm samples (C, M, P). The samples were analysed by immunoblotting using antibodies to the C-terminal His tag. hGH(ls) denotes linker-less mature hGH

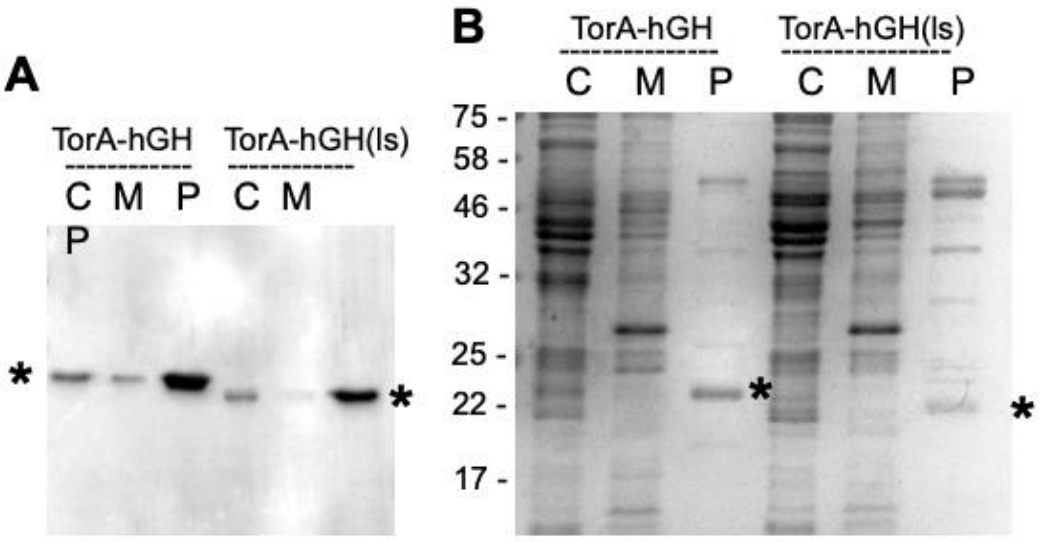
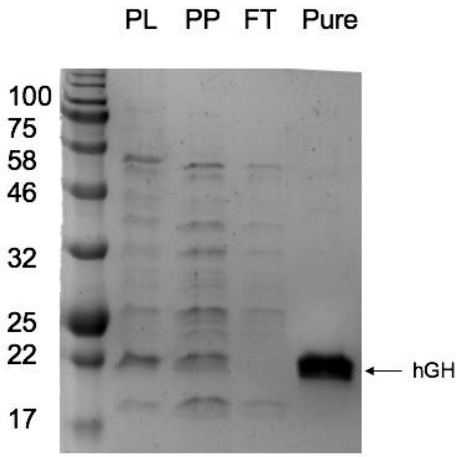
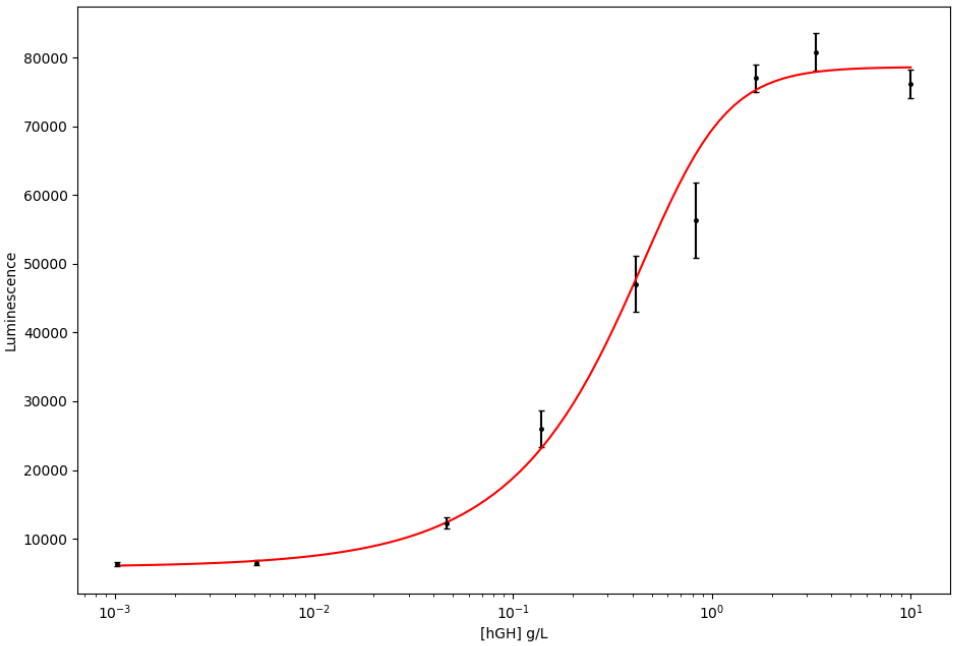


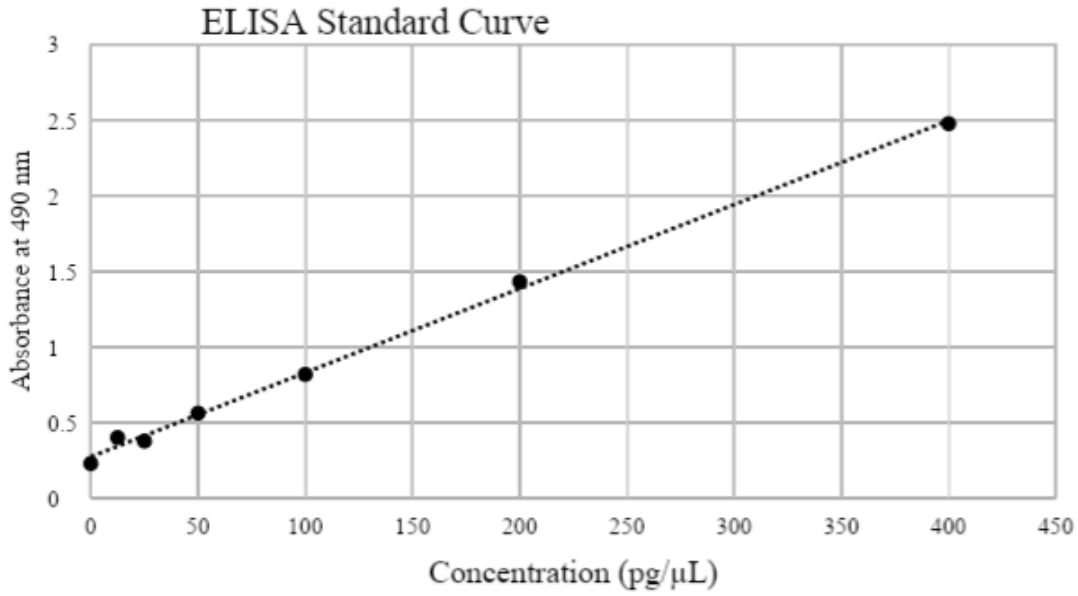
Figure 9. Comparison of export levels achieved during the export of TorA-hGH with and without a 4-residue linker during fed-batch fermentation of TatExpress cells: TorA-hGH constructs with or without the 4-residue linker were expressed in fed-batch fermentation with 40 h induction and samples were fractionated to generate cytoplasm, membrane and periplasm samples (C, M, P). The samples were analysed by immunoblotting using antibodies to the C-terminal His tag (A) or by Coomassie staining of SDS polyacrylamide gels (B). hGH(ls) denotes linker-less mature hGH construct; asterisks denote the mature hGH forms



Supplementary Figure 1. Purification of periplasmic hGH. Purification of hGH from periplasmic sample of fed-batch fermentation after 51 hours post induction via Nickel affinity column. Coomassie showing periplasmic released with lysozyme (PL), periplasmic release with lysozyme free method (PP), flow through after purification (FT) and purified hGH (Pure). Molecular weight markers are shown on the left (in kDa) and mature hGH height is indicated on the right.



Supplementary Figure 2. Quantification of hGH concentration using the PathFinder bioassay. An hGH concentration standard curve was constructed using the supplied commercial hGH and all samples were analysed in triplicate using dilution factors in the range of the standard curve. The standard curve was adjusted using numpy, matplotlib.pyplot and scipy in Python.



Supplementary Figure 3: Calibration curve and quantification of fed-batch fermentation samples obtained using an hGH ELISA kit. The calibration curve obtained from the ELISA kit using the standard hGH provided is shown, with concentration shown on the X axis (picogram per µL) and Absorbance at 490 nm shown on the Y axis. The purified periplasmic hGH was found to be present at a concentration that corresponds to 2.39 g hGH per litre of culture.

Annex 2: List of proteins of interest expressed at shake flask level with unsuccessful export to the periplasm.

Protein	Protein description	Export to Periplasm
Fab Fragment	Fab fragment obtained from Fujifilm.	None
Fab Fragment	Fab fragment obtained from Lloyd Ruddock	None
Fab Fragment	Fab fragment obtained from MedImmune	None
Sel1	Part of a protein complex required for the retrotranslocation or dislocation of misfolded proteins from the endoplasmic reticulum lumen to the cytosol, where they are degraded by the proteasome in a ubiquitin-dependent manner.	None
Bgal	Ype II membrane-bound glycoproteins	None
SaSodA	Superoxide Dismutase	None
WxL1	Hypothetical protein: leucine Rich repeat-containing WxL domain surface cell wall-binding protein	Low
WxL2	Hypothetical protein: WxL domain surface cell wall-binding	None
Glp1	Glucagon-like peptide	None

The proteins of interest were cloned between a TorA signal peptide and a 6xHis-tag in a pET23 vector. The vector was then transformed into W3110 cellas and cultures were grown at 30 °C, 200 rpm in 250mL Erlenmeyer flask. At OD_{600} of 0.4-0.6 plasmids were induced with 0.5 mM IPTG (25 µL, 1 M IPTG). After 3h induction, cells equivalent to a density of OD_{600} of 10 (~8 mL) were taken and fractioned into cytoplasmic (C), insoluble/membrane (M), periplasmic (P) and inclusion bodies (IBs) samples. Periplasmic (P) fractions were collected. Protein expression was analysed by western blot.