Exploring the Tat pathway for the production of disulphide-bonded recombinant proteins using CyDisCo in *Escherichia coli*

Claudia Arauzo-Aguilera

A thesis submitted for the degree of Doctor in Philosophy

University of Kent

Department of Biosciences

2022

Table of contents

Declaration

The work presented in this thesis is original and was conducted by me (unless otherwise stated) under the supervision of Professor Colin Robinson. All sources of information have been acknowledged by means of references. None of this work has been used in any previous application for a degree.

SECRETERS project is funded by the EU Horizon 2020 research and innovation programme under the Marie Skłodowska Curie grant agreement No 813979.

I would like to thank and acknowledge the following people for their help in producing data used in this thesis:

Dr. Stefan Krahulec, Dr. Matthias Berkemeyer, Dr. Karin Koch and Process Science Downstream and Upstream Development departments at Boehringer-Ingelheim (Austria) for allowing and helping in a collaboration that lead to a manuscript.

Dr. Mirva Saaranen and Professor Lloyd Ruddock from the University of Oulu (Finland) for the conceptualization and constant supervision of YebF work and the following useful discussions that ended in a manuscript.

Dr. Douglas Browning from the University of Birmingham for designing the mutations in tac promoter.

Some of the results presented in this thesis have been submitted to, or are in preparation to be submitted to, the following journals:

Klaudia Arauzo-Aguilera, Luisa Buscajoni, Karin Koch, Gary Thompson, Colin Robinson, Matthias Berkemeyer (2022). Yields and product comparison between *Escherichia coli* BL21 and W3110 in industrially relevant conditions: anti-c-Met scFv as a case study. **In preparation to be submitted** to Microbial Cell Factories.

Klaudia Arauzo-Aguilera, Mirva J Saaranen, Colin Robinson and Lloyd W Ruddock (2022). Highly efficient export of a disulphide-bonded protein to the periplasm and medium by the Tat

pathway using CyDisCo in *Escherichia coli*. **Submitted** to MicrobiogyOpen: MBO32022100306.

Rettenbacher, L. A., **Arauzo-Aguilera, K**., Buscajoni, L., Castillo-Corujo, A., Ferrero-Bordera, B., Kostopoulou, A., Moran-Torres, R., Núñez-Nepomuceno, D., Öktem, A., Palma, A., Pisent, B., Puricelli, M., Schilling, T., Tungekar, A. A., Walgraeve, J., Humphreys, D., von der Haar, T., Gasser, B., Mattanovich, D., Ruddock, L., & van Dijl, J. M. (2021). Microbial protein cell factories fight back? *Trends in Biotechnology 40*(5), 576–590. https://doi.org/10.1016/j.tibtech.2021.10.003 - **Published** (review).

Covid Impact Statement

The research experiments presented in this thesis started in mid-September 2019 and finished in mid-September 2022. During this 3-year-period, Covid-19 pandemic hit the whole world for more than a year and affected the research presented below. Some examples of how this affected my results are:

- Impact on access to the facilities and normal work environment. University of Kent Bioscience School was closed from mid-March until beginning of September in 2020 for non-essential work. From September onwards work was carried out in shifts to minimise the risk of spreading the virus.
- Impact on research and production of preliminary data, development of collaboration or methodological/technique training and experience. Being such a collaborative research program where most of the collaborators work in other countries in Europe, secondments were postponed, shortened (running out of time and quarantine periods) and flow of information was hampered due to covid restrictions.
- Change in personal circumstances. Contracting Covid-19 twice and mandatory quarantine periods slowed the research plan.

Overall, Covid-19 pandemic adversely affected the research plan and delayed enormously the progress of the research in this study.

Acknowledgements

This has been a bumpy and very exciting journey, and I would not have gotten until the end without the help of the great people that I have been surrounded by, some from the start and some others from the midway.

Firstly, I would like to thank my supervisor, Professor **Colin Robinson**. Thanks for believing I could get through a PhD and permitting me to get the most out of this journey. Thanks as well for our scientific discussions. I am truly grateful for this opportunity.

I also want to acknowledge CR lab. A big thanks to **Alex Jones**, **Amber Peswani**, **Hollie Scarsbrook** and **Conner Webb**, you made my life more pleasant while in Canterbury. Especially you, Conner: thanks a lot for putting so much effort in teaching me and in making me laugh.

I feel deeply grateful for being part of the SECRETERS consortium, and I could not be more excited to see you getting your degree! Each of you have been an inspiration and saying goodbye to you is by far the saddest thing about finishing my PhD. Special thanks to my Spanish-speaking gang for keeping me motivated. **Luisa Buscajoni**: you are just amazing, and you deserve the best, your kind nature will take you far. Thanks for making my days bright and full of laughter.

To **Mirva Saaranen** and **Lloyd Ruddock**. Many thanks for adopting me in your lab, you taught me a lot and I felt very supported during and after my stay in Oulu.

I also want to thank **Karin Koch** for welcoming me and our project at Boehringer-Ingelheim in Vienna. I know it was not the easiest project, but I am really thankful for your patience and all the useful discussions we had.

A shout-out to the very best friend anyone could have: **Claudia Niubó**. You took the crazy decision of coming to live with a first-year PhD student to Canterbury. Thank you because you cannot imagine what a support you were and how fun you made my days.

How not, thanks a lot to all my family and friends that kept asking when I am finishing and when I am coming back. You have been a tremendous support and I missed you a lot over these years. Also, I do not forget my few friends in Canterbury, that indirectly took care of my mental health

during the worst part of the pandemic. Of course, I want to thank my parents, **Claudia Aguilera** and **Carlos Arauzo** for their endless support in everything I decide to do in my life. And yes, it's time to go back home.

Finally, I want to deeply thank **Aatir Tungekar.** Thanks for your time, for supporting me, for our science discussion until late night, and for always being available to listen to me. You made me a better person, and this is what I value the most.

Abstract

The production of heterologous proteins of interest in microbial hosts such as *Escherichia coli* is a cost-effective and vital tool for the biopharmaceutical and enzyme industries. Research and development efforts over the past few decades have majorly focused on the Sec pathway for recombinant protein secretion in *E. coli*. However, the Twin-arginine translocation (Tat) pathway of *E. coli* has recently garnered interest for the periplasmic export of folded biopharmaceuticals as it possesses a unique proofreading ability to export correctly folded proteins. Tat-based translocation of biopharmaceuticals to the periplasm has the potential to greatly ease the downstream processing by reducing the amount of contaminating proteins and nucleic acids, thereby lowering the costs and time needed for their production. Furthermore, coupling Tat-dependent protein secretion with the CyDisCo technology (cytoplasmic disulphide bond formation in *E. coli*), which enables disulphide bond formation in the cytoplasm of *E. coli*, provides a powerful platform for the production of industrially relevant and difficult-to-express proteins.

Initially, a new experimental design for the export of multiple disulphide bond containing proteins to the periplasm by Tat in the presence of CyDisCo was empirically tested and developed as a starting point for future experiments. Later, a comparative study between Sec and Tat pathways for periplasmic export of a single chain variable fragment (scFv) in *E. coli* BL21 and W3110 was carried out. This led to an interesting comparison between the two strains based on soluble protein yields and quality when exported by the Sec pathway in fermenters. We observed a stark difference in the productivity profile of these two strains and a similar heterogeneity profile in the final product in both strains that could not detected by mass spectrometry-based analysis but confirmed by Differential Scanning Calorimetry (DSC) and 1D-1^{H}-NMR spectroscopy. Finally, in order to understand the limitations and maximize the applicability of the Tat pathway for proteins that fail to get exported efficiently with the TorA signal peptide, we focused on investigating alternative signal peptides. Here, we found that alternative signal peptides namely AmiC and MdoD allow highly efficient secretion of a disulphide bond-containing protein (YebF) to the periplasm of *E. coli* via Tat with CyDisCo. We report that these signal peptides are far more

efficient than the well-known Tat specific TorA signal peptide, and hence they are potentially more suitable for large scale protein production than this more rigorous Tat-specific signal peptide. Overall, the research work presented in this thesis suggests that the Tat pathway and the CyDisCo system are attractive platforms for biotechnology and establishes highly efficient Tat-dependent secretion of disulphide-bonded protein YebF to the *E. coli* periplasm.

List of abbreviations

°C	Degrees Celsius
µL	Microlitre
µM	Micromolar
µm	Micrometer
aas	Amino acids
AIM	Autoindution media
AmiC-	AmiC signal peptide fused to following protein
ATP	Adenosine triphosphate
AxA	Alanine-any-Alanine
bp	Base pair
C	Cytoplasmic fraction
CEX	Cation Exchange Chromatography
CHO	Chinese hamster ovary cells
C-terminal	Carboxyl-terminal
CueO	Multicopper oxidase
CyDisCo	Cytoplasmic Disulphide formation in *E. coli*
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotides
DSB	Disulphide bond
DSC	Differential Scanning Calorimetry
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EoF	End of fermentation
Erv1p	*Saccharomyces cerevisiae* mitochondrial thiol oxidase
Fab	Fragment antigen-binding
FDA	The United States Food and Drug Administration
Fc	Fragment crystallizable
Ffh	Fifty-four homologue
GFP	Green fluorescent protein
GTP	Guanosine-5'-triphosphate
H	Histidine (his)
h	Hours
hGH	Human growth hormone
I	Insoluble fraction
IB	Inclusion bodies
IL	Interleukin
IM	Inner membrane
IMAC	Immobilized Metal Affinity Chromatography
IPTG	Isopropyl β-D-thiogalactoside
kDa	Kilodaltons
L	Litre
LB	Lysogeny broth
LBA	Lysogeny agar
mAbs	Monoclonal antibodies
MB	Membrane fraction

MBP	Maltose Binding Protein
MdoD-	MdoD signal peptide fused to following protein
min	Minute
mL	Millilitre
mM	Millimolar
MS	Mass spectrometry
ng	Nanogram
NMR	Nuclear magnetic resonance
o/n	Overnight
OD	Optical density
OM	Outer membrane
P	Periplasmic fraction
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PDI	Protein Disulphide Isomerase
PCN	Plasmid copy number
PCR	Polymerase chain reaction
PhoA	Alkaline phosphatase
PTM	Post-translational modification
REMP	Redox enzyme maturation protein
RNA	Ribonucleic acid
rpm	Revolutions per minute
RR	Twin-Arginine motif
scFv	Single-chain variable-fragment from an antibody
SDS	Sodium dodecylsulphate
Sec	General Secretory pathway
SOP	Standard operating procedure
SP	Signal peptide
SRP	Signal recognition particle
T1-6SS	Type 1-6 secretion system
Tat	Twin-Arginine Translocase
TE	TatExpress strain
TMAO	Trimethylamine-N-oxide
TorA	Trimethylamine-N-oxide reductase
TorA-	TorA signal peptide fused to following protein
Tris	Tris (hydroxymethyl) aminomethane
Tween20	Polyoxythylenesorbitan monolaurate
UV	Ultraviolet
US	United States
V	Volts
W	Wash
WB	Western Blot
WT	Wild-type strain

CHAPTER 1: Introduction

1.1. General introduction

Recombinant proteins are required in academic research, as well as the pharmaceutical, textile and food industries and have several uses such as therapeutics, research proteins, industrial enzymes and proteases in laundry detergents (Adrio & Demain, 2014). Since the production and successful entry of the first recombinant protein into the market, several protein products have been developed and approved for human use. Human insulin was the first recombinant protein drug to be licensed and was also the first drug produced using recombinant DNA technology. The so-called "Humulin" was approved by the US Food and Drug Administration (FDA) in 1982. Insulin is composed of two peptide chains, insulin A and insulin B chain, which are linked together by two disulphide bonds (DSB). At the time, insulin was expressed in a two-plasmid system, the two peptide chains were purified and then bioactive insulin was obtained through *in vitro* DSB formation (Johnson, 1983; Baeshen et al., 2014). Insulin accumulates within the cell and after the growth process, the bacteria are lysed to purify the protein (Nilsson et al., 1996; Baeshen et al., 2014). Before using *Escherichia coli (E. coli)* for the production of insulin, insulin used to be purified from the pancreatic tissue of cows and pigs (Lens & Evertzen, 1952; Vecchio et al., 2018). At that time, insulin quality was variable, and the yields were low (Lens & Evertzen, 1952). Additionally, the pancreatic tissue had to be kept frozen to be suitable for purification.

Since the early commercialization of the recombinant human insulin, the later recombinant human growth hormone (hGH), and the recombinant bovine growth hormone, the field has expanded to include over 174 US-approved therapeutic proteins, including 98 antibodies (FDA, 2020). The year 2012 saw the highest approval of peptide drugs and protein-based drugs, reaching a global market value of $174.7 billion in 2015 at a compound annual growth rate of 7.3% (Agyei et al., 2017). Hence, the rising demand for new biopharmaceuticals requires increased production capacities as well as new production processes, which also implies the use and development of suitable expression systems (Müller et al., 2006). Bacterial expression systems are historically the first developed hosts for heterologous protein production. They are attractive because of their

ability to grow rapidly and at high density on inexpensive media, their well-characterized genetics and the availability of a large number of cloning vectors and mutant host strains (Terpe, 2006). However, due to their inability to adequately process complex proteins and due to their insufficient protein secretion capabilities, prokaryotic expression systems nowadays are mainly used to produce simple proteins and peptides (Müller et al., 2006). Many engineering approaches have emerged in the past decades to overcome these limitations in order to render bacteria even more attractive for recombinant protein production, but a tendency towards mammalian cells exists nowadays (Rettenbacher et al., 2021). Mammalian cells are the hosts of choice for proteins that require humanized post-translational modifications (PTMs). The main drawbacks of these hosts are the high costs brought by the slow cell growth, expensive media and specialized equipment needed for culture conditions (Aricescu et al., 2006).

Bacterial systems can be broadly classified into Gram-positive and Gram-negative. Gram-positive bacteria's cell wall consists of a thin periplasmic space between a thick peptidoglycan layer and the cell membrane (Figure 1) (Matias & Beveridge, 2005). On the other hand, the inner and outer cell membranes of Gram-negative bacteria surround a periplasmic space that contains a thin peptidoglycan layer (Silhavy et al., 2010) (Figure 1). In terms of recombinant protein production, this difference is translated by the ease of protein secretion in Gram-positive compared to Gram-negative bacteria. Indeed, Gram-positive bacteria such as *Bacillus subtilis* are frequently used for the production of proteins to be secreted into the extra-cellular medium (Van Dijl et al., 2002). Other bacterial expression hosts have emerged like the Gram-positive *Lactococcus lactis* and the Gram-negative *Pseudomonas* (Chen, 2012; Rettenbacher et al., 2021). For instance, protein production in *Pseudomonas fluorescens* is less oxygen-dependent, does not accumulate acetate and can give a yield 30 times higher compared to production in *E. coli* (Huang et al., 2007).

Even though these emerging microorganisms present a promising alternative to the better-established microorganisms in the field, *E. coli* is still the most preferred microorganism to express heterologous proteins for therapeutic use. Around 30% of the approved therapeutic proteins, particularly non-glycosylated proteins, are being produced using *E. coli* as a host (Baeshen et al., 2015). These include insulin, interferons, monoclonal antibodies and interleukins

among other therapeutics (Kamionka, 2011; Baeshen et al., 2015; Jozala et al., 2016; Sanchez-Garcia et al., 2016). *E. coli* offers several additional advantages including fast growth, high yield of the product, cost-effectiveness, easy scale-up and the availability of non-pathogenic species (Rosano & Ceccarelli, 2014; Baeshen et al., 2015).

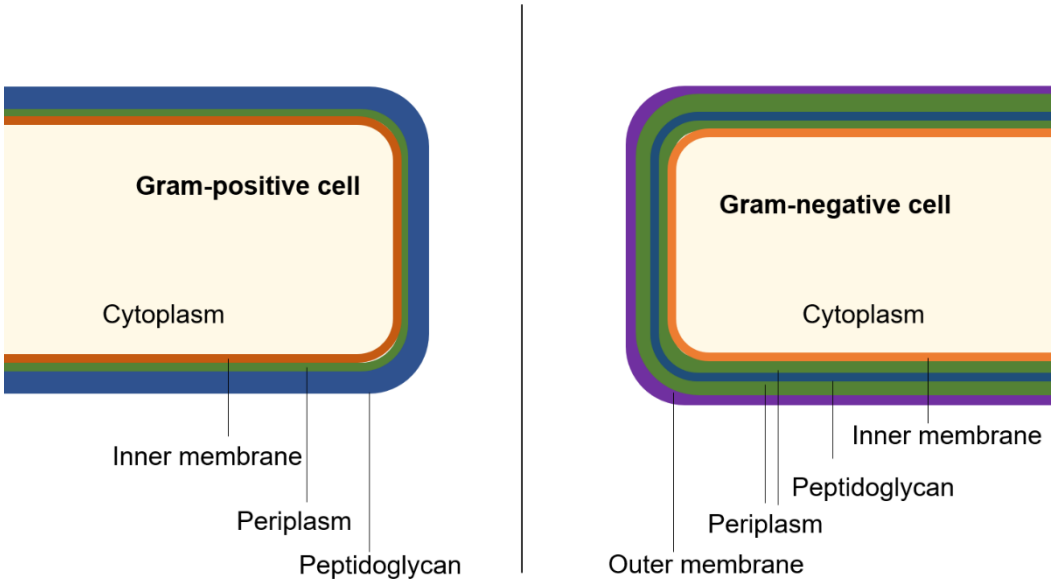


Figure 1. Diagram of Gram-positive and Gram-negative envelopes. Schematic representation of a Gram-positive cell (left) and Gram-negative cell (right). The cytoplasm of both cell types is contained by the inner membrane, enclosed by periplasm and the peptidoglycan wall. In comparison to Gram-negative cell, Gram-positive cell has a thicker peptidoglycan wall. The outer membrane is an extra feature of Gram-negative cell.

1.2. *Escherichia coli* strains in the biopharmaceutical industry

Nowadays almost all biopharmaceuticals are produced in bacteria, yeast and mammalian cells, with the former two grouped as microbial systems. In recent years, due to the rising dominance of monoclonal antibodies (mAbs), which require humanized PTMs, a tendency towards mammalian cell lines has arisen. This recent microbial system displacement by mammalian cells is very well reviewed in Rettenbacher et al. (2021) (Annex 1), which discusses the future of recombinant production with a holistic comparison of four of the most lucrative microbial production hosts (*E. coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae* and *Pichia pastoris*),

mammalian hosts, new upcoming biotherapeutics and the impact of emerging tools such omics and systems biology.

E. coli has been extensively studied and is the most used microorganism in biological research laboratories and the biopharmaceutical industry. Its physiology and molecular biology have been very well characterized, and therefore a large number of cloning or expression vectors, mutants and protocols of cultivation have been developed. This knowledge about *E. coli* allows researchers to better employ it for the production of recombinant proteins as well as for the production of other metabolites such as biofuels, amino acids, diamines and others. Moreover, and due to this extensive use in biotechnology, many different strains have been developed (Marisch et al., 2013; Blount, 2015; Idalia & Bernardo, 2017; Pham et al., 2019). Choosing the ideal host to produce a specific protein of interest (POI) is a critical step when developing a production process. Among all the *E. coli* strains available, some of them pathogenic to humans (Croxen et al., 2013), the most widely used *E. coli* strains for recombinant protein production have been those derived from B and K-12 strains. Both representative strains have distinctive differences in genotypic and phenotypic features associated with key cellular functions including cell growth, protein production and quality, motility, and cellular resistance (Yoon et al., 2012; Rosano & Ceccarelli, 2014). *E. coli* B, especially BL21, accumulates less acetate in minimal media, grows faster and lacks flagella and some proteases. These characteristics make it highly suitable for the enhanced production of recombinant proteins, as it does not use the extra energy required by the bacterial flagellar motor (Studier et al., 2009; Lozano Terol et al., 2019; Rosano et al., 2019). Moreover, the B strain has a more permeable membrane and an additional type II secretion system, making it more favourable for protein secretion (Yoon et al., 2012). On the other hand, the K-12 strains have a solid membrane structure with flagellar motility and higher levels of proteins that respond to stress, making them more resistant to stress conditions (Han et al., 2014). *E. coli* BL21 (DE3) is commonly used as a host because it is believed that higher protein yields can be achieved in *E. coli* B strains compared to *E. coli* K-12 derived strains (Rosano & Ceccarelli, 2014). However, researchers in both academia and industry have encountered obstacles for protein production in BL21 (DE3), such as for example plasmid loss,

which limits or completely prevents protein production. As a result, protein production in non-B strains has gathered a lot of interest in recent years (Pan & Malcolm, 2000; Rosano & Ceccarelli, 2014). *E. coli* K-12 have given rise to strains such as MG1655 and W3110 and nowadays, these are routinely used for recombinant protein expression (Daegelen et al., 2009). Many examples of successful protein expression in *E. coli* K-12 strains can be found at laboratory scale such as interleukin 3 (Hercus et al., 2013), myosins (Pylypenko et al., 2016), amylases (Zafar et al., 2016), and therapeutics at industrial scale such as hGH (de Oliveira et al., 1999) and Fab fragments (Goel & Stephan, 2010), among others.

Due to the aforementioned advantages that B strains seem to offer in recombinant protein production, it was selected as a host for most of the research experiments in this thesis. However, in Chapter 3, an insight of the differences in yield and product quality found in the production of a single-chain variable-fragment (scFv) in industrially relevant conditions can be found.

Nevertheless, the interaction of the target protein with cell metabolism remains little understood, and this slows an efficient process design. Even though the selection of the host strain is very important for enhancing the production of the POI, a plethora of other parameters should be considered when optimizing the recombinant protein production process. A summary of such parameters is reviewed by Tungekar et al. (2021). These parameters will vary for each process and for each target protein and will differ if seeking for soluble production or insoluble production for refolding (section 1.4.).

1.3. Post-translational modifications in *E. coli*

One of the major limitations in the production of recombinant proteins are post-translational modifications (PTMs). PTMs occur either after or during protein synthesis and they change the activity and physicochemical properties of the protein. They range from simple chemical modifications in the amino acid chain, to the addition of complex branching structures on the protein backbone. The most widely found PTMs in recombinant proteins are glycosylation and

DSB formation. These two are frequently required for correct protein folding and bioactivity (Bhatwa et al., 2021; Ramazi & Zahiri, 2021).

Glycosylation is one of the most challenging PTMs and consists of the addition of a sugar to the polypeptide chain or lipid. Its nomenclature is based on its linkage type, and the most widely distributed glycosylations are *N*- and *O*- glycosylations. As *E. coli* does not own a natural glycosylation pathway, it makes this predominant eukaryotic PTM difficult to replicate (Wong. 2005; Nothaft & Szymanski, 2010). Some glycoproteins still can be expressed without their native glycans in their active form, such as interleukin-2 (IL-2) (Roifman et al., 1985). However, successful attempts to transfer the *Campylobacter jejuni* bacteria (Wacker et al., 2002) or yeast glycosylation systems (Valderrama-Rincon et al., 2012) to *E. coli* have been made so it can *N*-glycosylate to some extent. Successful efforts to achieve *O*-glycosylation were also made by the expression of human genes in *E. coli* (Du et al., 2019).

On the other hand, *E. coli* does have an endogenous system for DSB formation in the periplasm. Many pharmaceutically relevant proteins have DSBs in their structure, and *E. coli* can struggle with the production of disulphide-rich proteins with complex folding patterns (Rettenbacher et al., 2021). Nonetheless, it is possible to achieve high titers of disulphide-bonded proteins in *E. coli* (Guerrrero Montero et al., 2019). DSB formation in the target protein is especially important for the correct folding of the POI and its lack can cause the unfolded protein to become insoluble and create inclusion bodies (IBs). Thankfully there are some ways to circumvent this issue, the main one being to target the POI to the periplasmic space of the cell, as the cytoplasmic space is a reducing environment where DSB formation is difficult (Rosano and Ceccarelli, 2014). The periplasm, however, is an oxidizing compartment rich in DSB oxidoreductases, which catalyse the disulphide bridge formation (Mergulhão et al., 2005; de Marco, 2009). In an industrial process, a signal peptide (SP) recognized by the general secretory (Sec) pathway is attached to the N-terminus of the POI, and the POI is translocated from the cytoplasm to the periplasm in an unfolded state (section 1.5.). Later in the periplasmic compartment, the protein acquires its DSBs (Mirzadeh et al., 2020) (Figure 2). For example, in a case-study, the overexpression of *E. coli*'s periplasmic chaperones SurA and FkpA in combination in the same plasmid with a copy of DsbC

and DsbA that catalyse the formation and isomerization of disulphide bridges, improved the multi-DSB containing antibody fragment yields by correcting improper DSB formation in *E. coli* (Shriver-Lake et al., 2017). As expected, and since the biotherapeutics list is long and varied, many have a fast-folding mechanism that will disqualify them for export via Sec pathway.

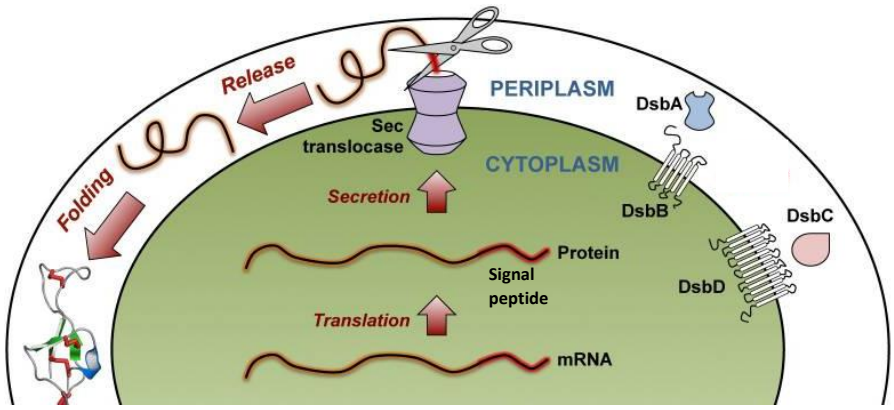


Figure 2. Schematic representation of periplasmic translocation for the production of DSB containing proteins in *E. coli*. After translation in the cytoplasm, the protein is translocated into the periplasm *via* the Sec translocase system. During translocation, the signal peptide is removed (red) and the mature protein (orange) is released to the periplasm. DsbA, DsbB, DsbC, and DsbD proteins aid in disulphide-bond formation. Adapted from Klint et al. (2013).

When the protein cannot be translocated by Sec system to the periplasm, DSBs need to be formed within the cytoplasm. As mentioned before, this cannot naturally occur. First and foremost, the cytoplasm's thiol-disulfide redox potential is insufficient to provide a driving force for the formation of stable disulphides. Second, there are no enzymes that can catalyse protein thiol oxidation under physiological conditions. The cytoplasm of *Escherichia coli* contains two thioredoxins, TrxA and TrxC, as well as three glutaredoxins. These proteins, in their oxidized form, can stimulate the formation of disulphide bonds of products. However, both thioredoxins and glutaredoxins are kept in a reduced state in the cytosol by the actions of thioredoxin reductase (TrxB) and glutathione, respectively. Glutathione in *E. coli* is synthesized by the *gshA* and *gshB* genes. The *gor* gene product, glutathione oxidoreductase, is necessary to reduce oxidized glutathione and complete the catalytic cycle of the glutathione-glutaredoxin system (Bessette et al., 1999). To overcome the impossibility of DSB formation in the cytoplasm, originally, the dysfunction of the natural reducing pathways was carried out. The removal or mutation of *gor* and *trxB* genes (Bessette et al., 1999) resulted in the commercially available Origami strain

(Novagen). Later, SHuffle strain (New England Biolabs) was launched, which expresses a chromosomal copy of the periplasmic DsbC in the cytoplasm in addition to the mutations in the reducing pathways thus, further improving cytoplasmic DSB formation (Lobstein et al., 2012). Even though they both help in the DSB formation of some heterologous proteins in the cytoplasm, yields are rather low and industrially dismissible due to their growth fitness (Ren et al., 2016). In the last decade, CyDisCo system (cytoplasmic disulphide bond formation in *E. coli*), which bases its technology on the expression of a eukaryotic (*Saccharomyces cerevisiae*) catalyst of DSB formation, sulfhydryl oxidase (Erv1p) and a catalyst of disulphide isomerization, human protein disulphide isomerase (PDI) from a plasmid, leaves the natural reducing pathways of the cells intact (Figure 3) (Matos et al., 2014). This system has allowed high yield production of different industrially relevant proteins such as avidin, single chain IgA_1 antibody fragment, hGH, interleukin 6 and other scFvs and Fabs (Gąciarz et al., 2016; Gąciarz et al., 2017; Gąciarz & Ruddock, 2017). Even a highly complex 44-DSB containing extracellular matrix protein was correctly folded by this system (Sohail et al., 2020).

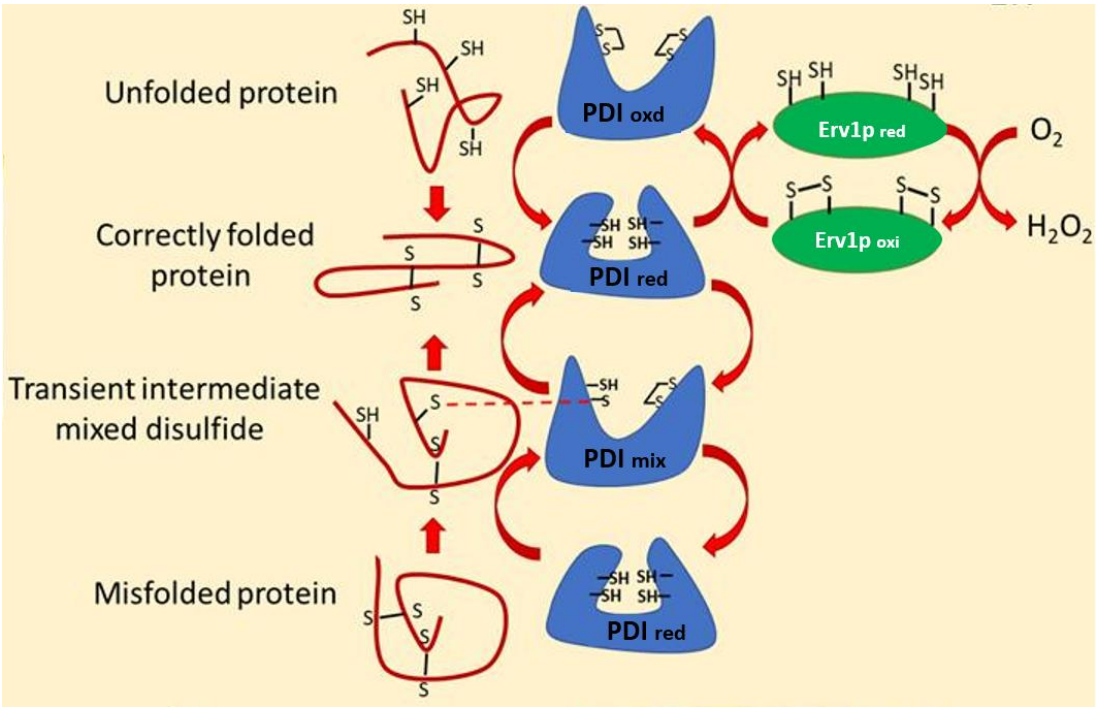


Figure 3. Schematic representation of the folding of a protein by CyDisCo system. PDI should be available in an oxidized state for the correct folding of the target protein. In order to maintain PDI in oxidized form, Erv1p must be present. Figure adapted from Tiphany de Bessa (2018).

1.4. Protein harvest in *E. coli*

1.4.1. Protein harvest from the cytoplasm

Once the correctly folded and disulphide-bonded protein has been obtained, harvesting and purifying are the next steps. As the main purpose of recombinant protein production is to harvest as much of the target protein from the cytoplasm, one of the major obstacles found is the protein accumulation in insoluble deposits (Thomas & Baneyx, 1996). The reasons for the formation of these IBs are varied and include insufficient chaperones to assist the protein folding, absence of other essential PTMs, especially DSB formation due to the reducing environment of the cytoplasm, and a high local protein concentration that results in precipitation. The literature describes different approaches to try to solubilise IBs and yield active protein when produced in the cytoplasm. Overall, they can be divided into procedures where the production protocol or protein target is modified to obtain the soluble expression in the cytoplasm and procedures to refold and solubilize the protein from IBs (*in vitro*) (Sandomenico et al., 2020).

As summarised by Correa & Oppezzo (2015), there are different ways to improve protein solubility *in vivo*. However, there is especially one that represents in some cases an advantage for structural biology, simplifying protein purification and allowing protein detection too: fusion partners. There are many fusion tags available, many of them under development and novel, and their optimal performance is highly dependent on the heterologous protein to be expressed which often requires a trial-and-error approach. Maltose Binding Protein (MBP) (40 kDa) and NusA (54.8 kDa) are ranked as two of the best tags for producing soluble proteins. Both have been used for the solubilization of highly insoluble scFvs in the cytoplasm of *E. coli* (Costa et al., 2013). MBP, for example, can also be used as a purification tag as it binds to the amylose resin (Reuten et al., 2016). Trx fusion partner, apart from enhancing the solubility of the recombinant protein, can also be useful in protein crystallization of certain target proteins because it forms several crystals itself, and it sticks rigidly to the target protein. The *FLAG* tag works both for purification and protein detection. All these tags can be removed from its passenger protein, as usually a protease recognition site allows the tag removal when the protein is used as a therapeutic, because

the tag can potentially interfere with the proper conformation and function of the target protein (Costa et al., 2014).

Quite the opposite, when IBs formation of the target protein is to be enhanced, IBs can be harvested by differential centrifugation (Palmer & Wingfield, 2004). Separating the POI from the rest of the cytoplasmic content can be beneficial to protect the cells from over-expressed toxic POIs or *vice versa*: protect the target proteins that are prone to degradation. The solubilization of IBs can be a simple operation in which they are solubilized by using chaotropic agents and detergents (Yang et al., 2011). The major drawbacks are the low yield of refolding, the need to optimise refolding conditions for each POI, and the possibility that the integrity of refolded proteins is compromised in the solubilisation procedures (Yamaguchi & Miyazaki, 2014). Moreover, the purification of highly expressed soluble proteins is more affordable and less time-consuming than refolding and purification of IBs. Maximising the production of soluble heterologous proteins is consequently an attractive alternative to *in vitro* refolding techniques (Sørensen & Mortensen, 2005).

1.4.2. Secretion to the periplasm and extracellular medium

The cytoplasm contains more host cell proteins including proteases which in turn impacts product purification. Periplasmic or medium secretion, on the other hand, improves folding, minimizes proteolysis and improves purification. Previously, periplasmic expression was associated with low yield of the product (mainly due to the small compartment that the periplasmic space is and the intrinsic limitation of the translocation machinery) (Kleiner-Grote et al., 2018), but recent studies demonstrated that up to 2.4 g/L of a Fab can be harvested using an engineered strain and co-expressing a chaperone (Ellis et al., 2017).

E. coli possesses several pathways to transport protein to the periplasm and to the extra-cellular medium. These are the Sec pathway (Sec), Twin-arginine translocation (Tat), Type I, Type II, Type III, Type IV, Type V and Type VI secretion systems (T1-6SS). Sec and Tat pathway (section 1.5. and 1.6. respectively) have been widely exploited for the production of heterologous proteins and will be extensively discussed in individual sections due to their high importance in this thesis

research. The rest are known to be involved in bacterial virulence, communication or interactions in the environment (Kleiner-Grote et al., 2018). Recently, T3SS has been repurposed as a protein delivery tool for a broad range of biomedical applications (Bai et al., 2018). T1SS, which translocates unfolded substrates across the cell envelope from the cytoplasm to the extracellular medium in a one-step process, has also been exploited for heterologous protein production (Linton et al., 2012).

Some recombinant proteins have demonstrated different secretory capacities that have not yet been explained. For example, L-asparaginase, an enzyme used to cure leukaemia, has been produced in *E. coli* (Khushoo et al., 2004). The protein was produced in fusion with the Sec-exported PelB SP and was found in high yields in the extracellular medium in cultures with a high cell density. The cause of the recombinant protein's translocation across the outer membrane (OM) is still unknown (Khushoo et al., 2004). This phenomenon does not seem to be unique, as it has also been seen when translocating other recombinant proteins into the periplasm in high-cell density cultures (Rinas & Hoffmann, 2004) and even in shaken *E. coli* cultures promoted by low aeration (Ukkonen et al., 2013). Numerous theories have been proposed as to why periplasmic leakage occurs, including cell division before the formation of outer membrane, the build-up of recombinant proteins in the periplasm that disrupt the membrane due to an increase of osmotic pressure, or lysis induced by periplasmic secretion, particularly in older cultures (Mergulhão et al., 2005).

1.5. The general secretory pathway (Sec)

Approximately 98% of the *E. coli* secretome uses the Sec route, of which 60% are proteins from the inner membrane (IM), while the remaining 40% are from the OM, periplasm, and extracellular proteins (Yu et al., 2010; Orfanoudaki & Economou, 2014). The Sec system permits the translocation of unfolded substrates to the periplasmic space or into the inner membrane by using two different mechanisms: the SecB and signal recognition particle (SRP) pathways (Pugsley, 1993). Since its discovery, the Sec system in *E. coli* has been exploited for the export of

recombinant proteins to the periplasm. According to the nature of the Sec SP attached to the N-terminus of the POI, heterologous proteins can be translocated either by the SecB or the SRP route. For instance, the T cell receptor glycoprotein, parathyroid hormone, and the active human epidermal growth factor were all effectively transported to the periplasm by SecB utilizing the native SP of the *E. coli* outer membrane protein, OmpA (OmpA SP) (Wong & Sutherlan, 1993). Remarkably high yields of secreted proteins (in the 5-10 g/L range) after an exhaustive optimization of high cell density fermentation are well documented in the literature. Nonetheless, for most of the target proteins compatible with *E. coli* expression and folding machinery, reasonable titers starting from 0.5-0.8 g/L can be expected (Georgiou & Segatori, 2005).

1.5.1. Sec-specific signal peptides

Sec translocated proteins carry a Sec specific SP at the N- terminus (Green & Mecsas, 2016). While SecB detects less hydrophobic SPs, SRP targets highly hydrophobic SPs to the SecYEG channel. Sec-specific SPs may be divided into three parts, as shown in Figure 4, and are generally 18–27 residues long (Gouridis et al., 2009). The n-region, located in the amino terminal region, is distinguished by a net positive charge. One or more basic residues are responsible for the positive charge(s) of the region and permit the insertion of the SP into the SecYEG translocon. The net charge varies from substrate to substrate, and it has been found that it can influence the synthesis and translocation rates of the protein (Caspers et al., 2010). Furthermore, the translocation rate may be significantly reduced if the net charge is not maintained at a positive value (Inouye et al., 1982).

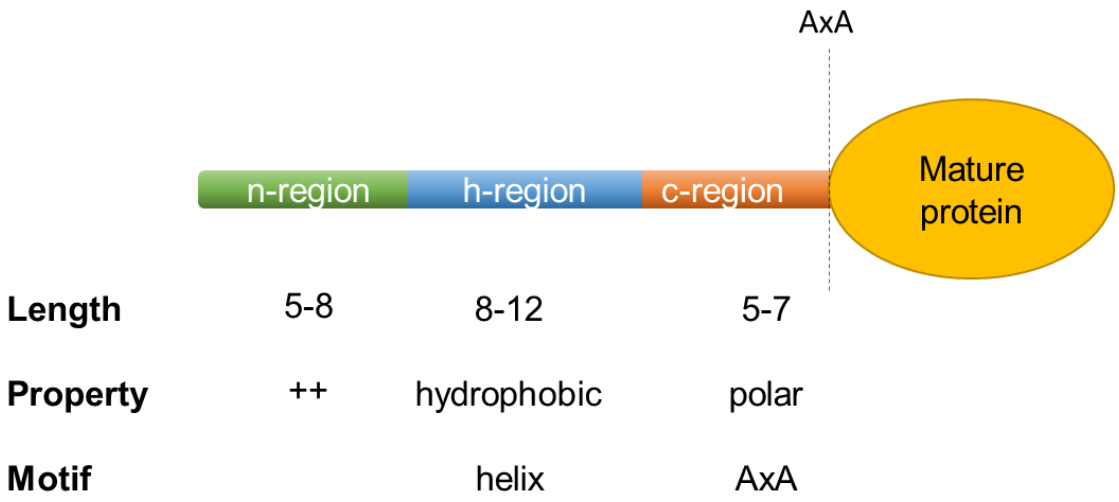


Figure 4. General Sec-specific signal peptide features. A Sec signal peptide is composed of a positively charged n-region (green), a central hydrophobic helix-shape h-region (blue), and a polar c-region (orange) at the carboxyl-terminal that contains the signal peptidase cleavage motif (AxA). The length is indicated in amino acid residues and "++" means positively charged.

The next region, named as h-region, possesses 8 to 12 amino acid residues and is characterized by its hydrophobicity and its helical propensity (Figure 4). When decreasing the hydrophobicity of this region, by mutation or deletion of amino acids, protein export decreases considerably often resulting in accumulation of the misfolded protein in the cytoplasm (Suominen et al., 1995). On the contrary, mutating the h-region to make it more hydrophobic has proven to increase the translocation efficiency (Chen et al., 1996). Adams et al. (2002) saw that introducing a helix breaker in this region reduces the affinity of SRP for the highly hydrophobic SPs. When extending the helix, the affinity of the SP shifted towards the SRP route.

The Sec SP's "c-region" at the C-terminus has a cleavage site that allows the mature domain to be released into the periplasm or into the IM (Figure 4). To ensure export and cleavage effectiveness, this region must be at least five residues long (Suominen et al., 1995). The SP is inserted into the SecYEG translocon for export, and when the nascent polypeptide has fully been translocated into the periplasm, the signal peptidase I on the outer leaf of the IM, LepB, cleaves it (Date, 1983). This peptidase recognizes a three amino acid motif referred as AxA, named this way since alanine (A) is the typical residue at these locations (Karamyshev et al., 1998). This region tends to have an almost neutral charge in Sec specific SPs (Berks et al., 2003). Protein

export levels can be optimized by the alteration of the c-region of the SP, as this region has a significant impact on SP cleavage efficiency (Low et al., 2013). Overall, it is important to note that there is no universal SP that will export all heterologous proteins through Sec, Tat or extracellular-medium pathways. Hence, a screening and optimization process to find the most adequate SP for each unique target protein to obtain the highest yields is required (Freudl, 2018).

1.5.2. The SecB pathway

Sec substrates are transported post-translationally through the IM via the SecB pathway to the periplasm or to the extracellular medium (Tsirigotaki et al., 2016). SecB protein is a non-essential protein; although its deletions show a hypersensitive phenotype to both cold and hot temperatures (bellow 23 °C and above 45 °C) and antibiotics (Ullers et al., 2007). The SecB protein, a homotetrameric chaperone located in the cytoplasm, binds to the SP of nascent polypeptides (Green & Mecsas, 2016). This interaction is the reason why the polypeptide remains unfolded and then, the complex binds to SecA. Translocation is initiated by the essential cytoplasmic translocation motor protein SecA that pulls the complex (the substrate and SecB) into the SecYEG channel (Figure 5). This process requires ATP hydrolysis as the energy source. In the later stage of the substrate translocation, SecA detaches from the SP and is returned to the cytoplasm (Tsirigotaki et al., 2016). Finally, the proton motive force, enabled by the accessory factor SecDF embedded in the inner membrane, pulls the polypeptide completely across the channel (Tsukazaki et al., 2011). The SP is then cleaved by the signal peptidase I (LepB) or by the membrane-anchored signal peptidase II (LspA) for lipoproteins, which enables the release of the polypeptide to the periplasmic space (Yu et al., 1984). Once in the periplasm, the mature protein is folded and can be further secreted to the extracellular medium.

1.5.3. The SRP pathway

The SRP pathway exports to the periplasm those polypeptides that SecB cannot maintain in an unfolded status or inserts into the membrane those substrates that possess hydrophobic regions (such as transmembrane domains) (Green & Mecsas, 2016). The SRP system is a ribonucleoprotein complex constituted by the fifty-four homologue (Ffh) protein and a small 4.5S

RNA. SRP attaches to ribosomes but is only kinetically stabilized by nascent polypeptide when recognizing a specific SP or transmembrane domain, and at that point it momentarily stops translation at the ribosome (Tsirigotaki et al., 2016). The SRP-polypeptide complex subsequently attaches to the docking protein FtsY, which is present in the IM and transports the complex to the SecYEG channel (Figure 5). The hydrolysis of GTP by the Ffh and FtsY GTPase domains provides the required energy for this process, which in turn causes the release of SRP and restarts translation. The polypeptide is directly synthesised within the SecYEG channel that employs the proton motive force to pull the developing chain through. When a membrane protein is inserted, YidC drives the integration of nascent transmembrane helixes into the inner membrane while PpiD chaperone is thought to provide a quality check on transmembrane domains that are leaving the SecYEG channel (Sachelaru et al., 2014). In the final steps, LepB cleaves the SP and the mature protein is released to the periplasm or inserted in the membrane (Auclair et al., 2012).

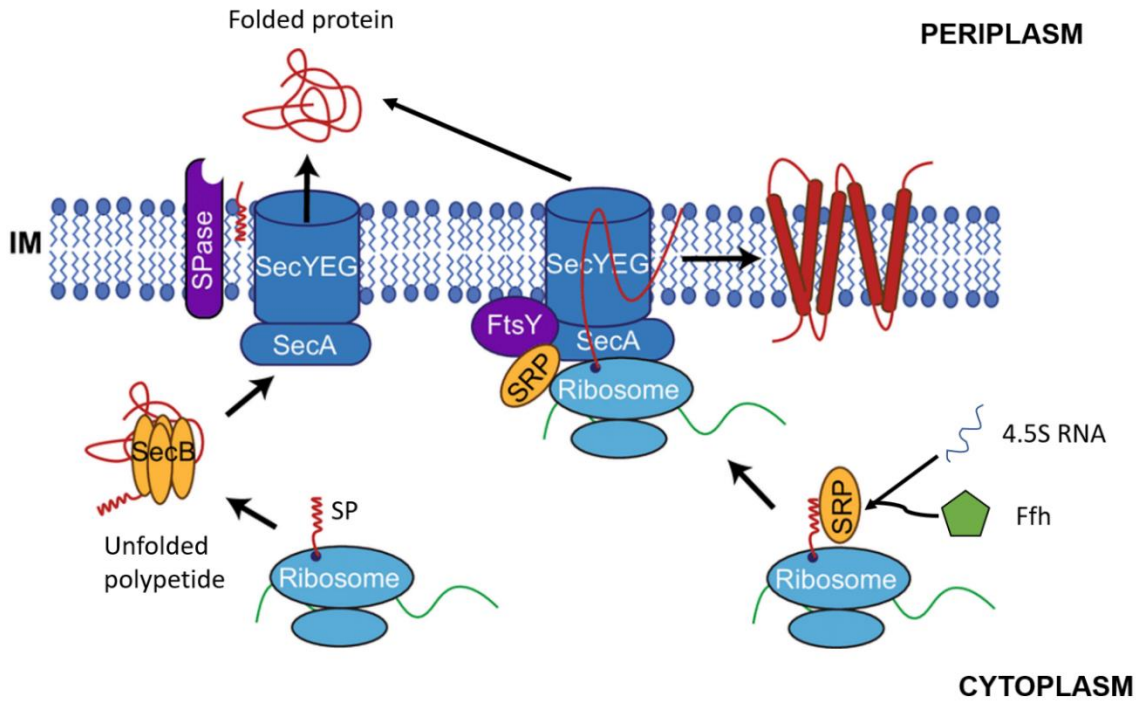


Figure 5. Sec mediated protein translocation in bacteria. The movement of unfolded proteins through the IM is mediated by the SecYEG complex. There are two recognition pathways that enable either the protein's translocation into the periplasm or its insertion into the IM. SecB recognizes post-translationally secreted polypeptides and keeps them in their unfolded condition while delivering them to the SecYEG complex via an interaction with SecA. The polypeptide is translocated by SecYEG. Signal peptidase I (SPase) cleaves the SP upon secretion, releasing the polypeptide into the periplasm where it can fold or be shielded by chaperones until OM insertion. When compared, SRP complex (made of the Ffh protein and the small 4.5S RNA) recognizes the extremely hydrophobic signal sequence of co-translationally secreted

proteins, and by interacting with FtsY, the SRP receptor, brings the translating ribosome to SecYEG. As the protein is translated, Sec introduces the transmembrane helixes of the IM proteins into the membrane or releases the mature protein to the periplasm. The figure excludes accessory Sec proteins for simplicity. Figure adapted from Silhavy & Mitchell (2019).

1.6. The Twin-Arginine Translocase pathway (Tat)

The twin-arginine translocation pathway (Tat pathway) is an alternative to the Sec pathway for protein translocation from cytoplasm to periplasm. The main distinction between the two routes is that Tat exclusively transports folded proteins, whereas Sec exports unfolded polypeptide chains into the periplasm (Patel et al., 2014). About 98% of the *E. coli* secretome is translocated through Sec and only the remaining 2% (around 30 substrates) are exported by the Tat pathway (Tullman-Ercek et al., 2007; Palmer & Berks, 2012).

The proteins exported by the Tat pathway are involved in important functions e.g., cell wall formation, phosphate, copper, and iron metabolism, pathogenicity and respiratory metabolism (Berks et al., 2005; Palmer & Berks, 2012). Tat translocation of the proteins provides a physiological advantage over Sec translocation in these five cases: (i) the insertion of complex cofactors, (ii) the prevention of periplasmic metal ions competing for insertion into the active site, (iii) the transport of hetero-oligomeric complexes that require cytoplasmic assembly, (iv) the requirement of cytoplasmic proteins for folding and maturation and (v) the instability of some unfolded proteins in the periplasm (Palmer & Berks, 2012). It's interesting to note that some redox proteins, despite not possessing their own signal sequence, are found to be transported through the Tat pathway. These proteins create a multimeric complex with another protein that has a Tat SP and use the so-called "hitchhiker mechanism" in order to be translocated by Tat (Rodrigue et al. 1999).

It is predicted that *E. coli* encodes around 30 Tat-dependent substrates, but there is a lack of experimental confirmation of the bioinformatic predictions. Moreover, some of the predicted Tat substrates are “promiscuous”, meaning they can either go to the periplasm by Tat or Sec or both

(Blaudeck et al., 2001; Stanley et al., 2002; Blaudeck et al., 2003; Tullman-Ercek et al., 2007). See Table 1 for a list of 27 Tat-secreted proteins.

TorA is an example of a protein exported by the Tat pathway. TorA, a trimethylamine N-oxide (TMAO) reductase encoded by the *torCAD* operon (Méjean et al., 1994), was one of the earliest Tat substrates found. TorA's SP has a twin-arginine motif (RR) and the protein is located in the periplasm (Weiner et al., 1998; Santini et al., 1998). The protein does, in fact, have a role in anaerobic respiration, which explains why it is found in the periplasm. For TorA to function, a cytoplasmically-inserted molybdenum cofactor is required, which justifies the need for Tat translocation. Additionally, TorA needs the chaperone TorD for cofactor binding and cytoplasmic maturation. Moreover, TorD also enhances targeting and translocation via the Tat pathway (Sargent et al., 2002; Ilbert et al., 2004; Jack et al., 2004; Lee et al., 2010) by providing activity against protein aggregation or/and proteolysis (Lee et al., 2010).

Even though the Tat protein export system is not essential for *E. coli*, the absence of a functional pathway results in serious growth abnormalities. Tat-dependent proteins become mislocalized when the Tat pathway is blocked. For example, amidases AmiA and AmiC are crucial for the cell division process and a functional cell envelope phenotype, and they are naturally located in the periplasm of *E. coli (*Bernhardt & de Boer 2003; Ize et al., 2003). Long chains of cells are formed as a result of their improper localisation; therefore, cells lack proper cell separation (Harrison et al., 2005).

From a biotechnology point of view, since Tat exports completely folded proteins, it means that it has a system for quality control and proofreading. Due to this unique ability, it is likely that the Tat system naturally produces high quality, active proteins, making this system a very valuable tool for heterologous protein production (Walker et al., 2015).

Table 1. List of Tat substrates in *E. coli*. List of 27 natural Tat substrates supplemented with UniProt entry numbers (www.uniprot.org) and a short protein description. The selected candidates for this research study are highlighted in green. The Tat substrate list was obtained from Sargent et al. (2005).

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
NapG	P0AAL3	Electron transfer to nitrate reductase
NrfC	P0AAK7	Electron transfer to nitrite reductase
PaoA/YagT	P77165	Aldehyde oxidoreductase iron-sulfur-binding subunit
YdhX	P77375	Uncharacterized ferredoxin-like protein
TorA	P33225	TMAO reductase catalytic subunit
TorZ	P46923	TMAO reductase-2 catalytic subunit
NapA	P33937	Nitrate reductase catalytic subunit
YnfE	P77374	Dimethyl sulfoxide (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
CueO	P36649	Multi-copper oxidase
SufI/FtsP	P26648	Cell division protein
YahJ	P77554	Uncharacterized protein
WcaM	P71244	Biosynthesis of colanic acid
MdoD/YdcG	P40120	Glucans biosynthesis protein D
EfeB/YcdB	P31545	Deferrochelatase/peroxidase
EfeO/YcdO	P0AB24	Iron uptake system component
YaeI	P37049	Phosphodiesterase
AmiA	P36548	N-acetylmuramoyl-L-alanine amidase (Cell wall amidase)
AmiC	P63883	N-acetylmuramoyl-L-alanine amidase (Cell wall amidase)
FhuD	P07822	Ferrichrome binding protein

1.6.1. Tat-specific signal peptides

Similar to Sec, Tat substrates have a particular N-terminal SP to target export (Patel et al., 2014). Both SP types share a tripartite structure: the n-, h-, and c-regions, including the AxA motif to release the mature domain upon translocation (Figure 6). The ability to selectively target the protein through its proper pathway is made possible by a number of distinctions between the SPs. Compared to Sec SPs, which have an average length of 24 amino acids, Tat SPs are typically 14 residues longer, with an average length of 38 amino acids (Cristobal et al., 1999). The decreased

hydrophobicity of the h-region, a conserved motif, and the basic residues of the c-region are the major characteristics that prevent Tat substrates from being mistakenly targeted to Sec (Berks et al., 2000, Palmer & Berks, 2012). It has been seen that an increase in the hydrophobicity of the h-region of a Tat SP can convert it into a substrate for the Sec pathway and at the same time prevent the export through Tat pathway (Cristobal et al., 1999).

At the intersection of the n- and h-regions lies the conserved S/T-R-R-X-F-L-K motif which contains the distinctive twin-arginine (RR) that gives the translocase its name (Figure 6) (Berks, 1996). When one arginine of the motif was mutated to lysine, the Tat efficiency decreased (Stanley et al., 2000). However, Tat-mediated export was totally prevented when both arginine residues were mutated to lysine on the native SufI's Tat SP (Yahr & Wickner, 2001). The presence of positively charged residues in the c-region is the final Sec-repelling characteristic. In fact, converting these residues to uncharged polar amino acids enables Sec export (Cristobal et al., 1999). Despite these "Sec avoidance" features, 16 out of the 27 predicted Tat SPs have been discovered to be promiscuous when fused to MBP as a target protein, meaning that these SPs fused to this POI can also be exported by Sec pathway (Tullman-Ercek et al., 2007).

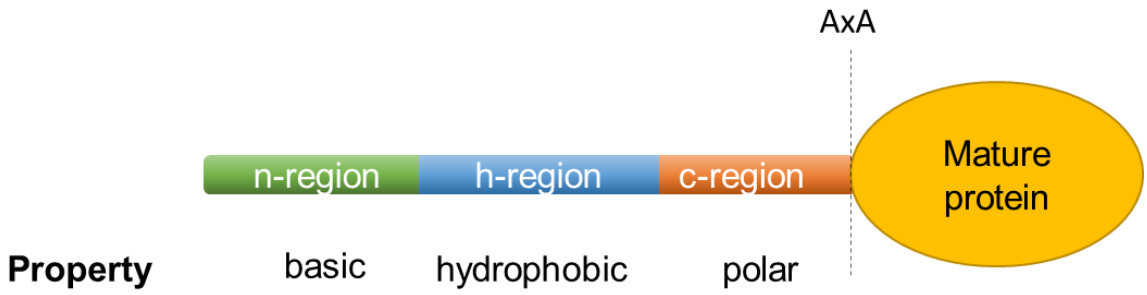


Figure 6. General Tat-specific signal peptide features. The Tat signal peptide is composed of a basic n-region (green), a hydrophobic h-region (blue), and a polar c-region (orange) at the carboxyl-terminal that contains the signal peptidase cleavage site (AxA). The S/T-R-R-X-F-L-K consensus motif is located at the junction between the n- and the h-regions.

1.6.2. Tat export mechanism

The Tat pathway in *E. coli* consists of five different proteins: four integral membrane proteins TatA, TatB, TatC and TatE and the cytoplasmic protein TatD. The first four are encoded in

tatABCD operon and TatE is encoded as a single gene (Sargent et al., 1998). There is a lot of uncertainty regarding the mechanism through which Tat exports folded proteins across the IM, and some of the Tat components' protein structures are being used to glean information. NMR has been used to determine the structure of the *E. coli* proteins TatA and TatB (Rodriguez et al., 2013; Zhang et al., 2014a; Zhang et al., 2014b), but no other structures of isolated Tat proteins have been published. Additionaly, TatBC complex of *E. coli* with and without substrate binding has also been published (Tarry et al., 2009).

TatA, TatB, and TatE are sequence-related and only have a single N-terminal transmembrane helix (Baglieri et al., 2012), whereas TatC is the most conserved of the Tat proteins and has six transmembrane helices (Behrendt et al., 2004). TatA and TatE are the putative pore-forming components that facilitate the actual protein translocation process, whereas TatC and TatB form a complex that serves as the first recognition site for Tat substrates (Fröbel et al., 2012). TatA and TatE have overlapping roles and their amino acid sequences are more than 50% identical (Sargent et al., 1998). It has been demonstrated that the TatA/B/C components' stoichiometry is essential for the Tat pathway's proper operation (Leake et al., 2008) and Tat-dependent protein flow can only increase when TatABC are concurrently overexpressed (Browning et al., 2017). TatB and TatC interact with a 1:1 stoichiometry and there are multiple copies of the two proteins (Bolhuis et al., 2001). In order to elucidate the function of Tat components, a variety of mutants have been constructed. It seems that TatB and TatC are both crucial parts of the Tat system, and their loss results in a total mislocalization of a particular Tat substrate (Lee et al., 2002; Wexler et al., 2000). ΔtatA or ΔtatE mutants, in contrast, cause a variety of cell abnormalities but do not completely impair the Tat mediated export. Finally, TatD protein was found to be nonessential as a TatD knock-out strain did not impact Tat translocation (Wexler et al., 2000), but it takes part in the quality control mechanism (Matos et al., 2009).

In the presence of folded Tat substrates in the cytoplasm containing the N-terminal Tat-specific SP, the reversible recruitment of the translocon components for translocation starts (Rose et al., 2013). The interaction of TatB and TatC proteins as dimers in the membrane, which serve as receptors for the substrates, is the initial stage of Tat translocation (Alami et al., 2003; Fröbel et

al., 2019). The interaction of these two proteins is crucial to avoid the substrate's premature release. Without TatB, the insertase TatC alone can also recognize Tat substrates by their SP but can prematurely trigger their cleavage (Fröbel et al., 2012). TatB is consequently necessary for the proper initiation of the translocation by preventing early cleavage of the SP. Together, they create a hydrophobic cavity in the lipid bilayer that allows the Tat signal peptides to dock there (Blümmel et al., 2017). Next, the Tat SP is fully inserted into a cavity that the TatBC complex has created (Figure 7) (Lausberg et al., 2012; Blümmel et al., 2015).

After substrate binding to the TatBC complexes, the TatA proteins are recruited and oligomerized (Alcock et al., 2013). This recruitment is mediated by the proton motive force (Patel et al., 2014). From this point, two models are currently being explored to explain the translocation of the substrate: the translocation pore model (Figure 7) (Palmer & Berks, 2012) and the membrane destabilisation model (Berks, 2015; Cline, 2015). According to the translocation pore theory, TatA proteins oligomerize at the TatBC/substrate complex to create a size-variable pore that is tailored for the bound substrate (Lange et al., 2007; Patel et al., 2014). In contrast, the second concept proposes that, following the oligomerization of TatA proteins, the membrane is thinned and destabilized as a result of these proteins' N-termini, and a temporary pore into the lipid bilayer is opened, allowing the substrate to pass through (Hou et al., 2018). The combination of the two models is likely to be the explanation for Tat translocation, as this system translocates diverse molecular weight substrates ranging from 10 to 150 kDa (Berks et al., 2000).

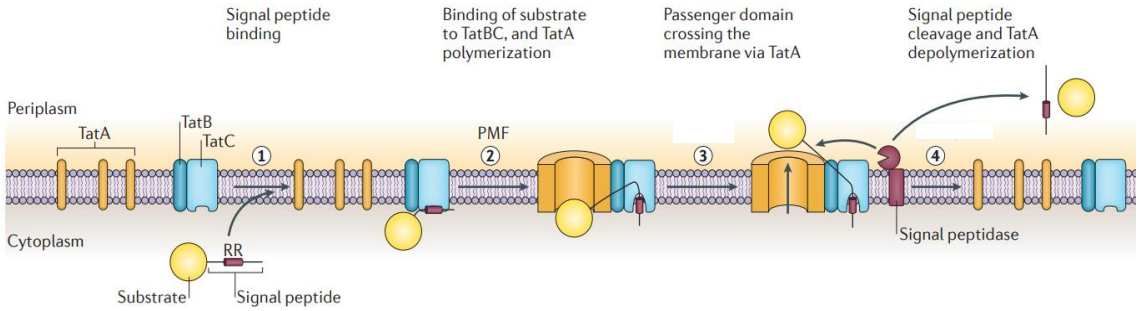


Figure 7. Translocation pore model of the Tat pathway of *E. coli*. At the start of the translocation, TatB and TatC are associated as a complex, while TatA is presented as dispersed protomers. In the first step (1) the TatBC complex binds the SP of the already folded substrate; the twin-arginine (RR) motif in the SP is recognized by a site in TatC. The rest of the SP and protein passenger domain are close to TatB. In the second step (2) the substrate SP becomes more tightly bound to the TatBC complex, however the N-

terminus of the SP remains in the cytoplasm. TatA protomers are recruited and polymerized to the TatBC complex. The resulting TatABC complex is the active translocation site and at this step, the SP is already in contact with these three Tat components. In the step three (3) the passenger domain of the protein crosses the membrane via TatA, while the SP remains bound to the TatBC complex. In the last step (4), the passenger domain has crossed the membrane and the SP is removed by a signal peptidase at the periplasmic face of the membrane. After successful translocation and SP cleavage, TatA dissociates from TatBC and depolymerizes back to free protomers. Adapted from Palmer & Berks (2012).

1.6.3. The proofreading and quality control mechanism

As mentioned before, Tat pathway has also a quality control and proofreading ability. Only fully folded substrates containing their designated cofactor (if necessary) can exit the cytoplasm via the Tat pathway (DeLisa et al., 2003; Sutherland et al., 2018). To ensure that cofactors are correctly assembled, specific cytoplasmic chaperones (redox enzyme maturation protein (REMP)) are involved. Without their chaperones, these Tat substrates cannot achieve a natural conformation which is detected by the Tat translocon preventing their transport (Turner et al., 2004). Therefore, we could define "quality control" or "proofreading" mechanism as the capacity to detect the folded state of Tat-targeted proteins. A misfolded substrate, alkaline phosphatase (PhoA), was fused to the SP from TorA, reached the Tat translocase but was not exported (Richter & Bruser, 2005). Pre-mature PhoA was associated with TatBC complex but did not trigger the recruitment of the TatA proteins. It seems that the Tat translocase “senses” the substrates that cannot reach their native conformation and triggers their degradation (Matos et al., 2008). Additionally, Rocco et al. (2012) identified specific mutations in the TatABC proteins which allows a misfolded model protein to be transported by Tat pathway to the periplasm.

The conformational tolerance of the Tat proofreading mechanism has been studied in recent investigations. The natural *E. coli* Tat substrate, multicopper oxidase (CueO), can be exported without its copper cofactor which is acquired in the periplasm. The apo form of CueO is nearly structurally identical to its holo form, what means the Tat senses this protein sufficiently folded to pass the proofreading mechanism (Djoko et al., 2010; Stolle et al., 2016). This tolerance has been further demonstrated many other times with heterologous proteins. Alanen et al. (2015) saw that hGH, interferon α2b and two scFvs were exported with high efficiency even in the absence

of cytoplasmic disulphide bond formation, while PhoA, AppA and a dimeric Fab construct employed in the study of DeLisa et al. (2003), which are larger proteins containing more DSBs, could not "fool" the Tat proofreading mechanism and hence were not transported by Tat. The ones employed by Alanen et al. (2015), on the other hand, were smaller, and did not contain more than two DSBs. It might be that the simple structure of the proteins enables them to acquire native or near-native structures in the absence of DSBs, therefore, allowing Tat to recognize them as fully folded and permitting its translocation to the periplasm. Sutherland et al. (2018) observed a correlation between Tat export efficiency and the increase of the substrate's rigidity when using a heme-binding protein known as BT6 maquette. They observed that the rising abundance of periplasmic protein was clearly correlated with increasing heme-binding capacity (three variants of BT6 binding either two, one or no heme(s)). This was already shown before by Jones et al. (2016). They realized that Tat preferentially exports the more rigid structure, thus, it senses the conformational flexibility: even when their model protein, an scFv, was stably folded and active even in its reduced form, the formation of its two DSBs (by CyDisCo) enhanced Tat-dependent export 10-fold. Furthermore, they found that this scFv was not translocated when fusing 26-residue unstructured tail at the C-terminus, suggesting that even this small change could be sensed by Tat quality control mechanism. Additionally, they claimed that the Tat proofreading system can tolerate major changes in surface charge or hydrophobicity, when they mutated the same model protein.

How certain proteins are rejected by the proofreading mechanism is still unclear. The co-purification of a mutant PhoA with the TatBC complex showed that misfolded proteins still interact with the Tat translocon (Richter & Bruser, 2005). Hence, Tat may not have an inbuilt mechanism for proofreading but might be instead connected to an efficient degradation system as it has been seen in Tat export in *Bacillus subtilis* (Frain et al., 2019).

1.6.4. A more efficient Tat export: TatExpress

The Tat pathway is only employed to export a small number of proteins to the periplasm, and because it is constitutively expressed in *E. coli* (Jack et al., 2001), the availability of Tat-forming proteins will always be a limitation when trying to exploit this system for heterologous protein

production. Trying to solve this limitation, a new strain was developed with a *tac* promoter right before the original promoter of Tat in the genome of the most used *E. coli* strains in industry (e.g., BL21 and W3110). This new-developed strain, name as TatExpress (TE) strain, significantly improved the Tat-dependent secretion of hGH and an scFv into the bacterial periplasm, boosting the commercial and industrial relevance of the Tat pathway (Browning et al., 2017; Guerrero Montero et al., 2019).

1.7. Aims of this project

The primary aim of this research project was to achieve efficient export of industrially usable prefolded, disulphide-bonded proteins by the Tat export pathway of *E. coli* CyDisCo strains; initially at shake-flask and later at fed-batch fermentation scale.

Export of DSB containing proteins to the periplasm, an oxidising environment, presents an advantage over the cytoplasm production, as DSBs can naturally form. The periplasm also contains chaperones that can enhance the correct folding of the POI, contains less proteases than the cytoplasm and facilitates downstream processing (Mergulhão et al., 2005; Balasundaram et al., 2009). Moreover, Tat pathway offers an attractive alternative to the Sec pathway, because it translocates cytoplasmic folded proteins and contains a proofreading mechanism that only permits correctly folded proteins to be exported. This pathway alongside CyDisCo technology might be a welcome alternative for quicker and cost-effective production of high-value proteins.

Firstly, the export of cytoplasmic prefolded proteins by CyDisCo was assessed by the Tat pathway. Until now, only small, simple biopharmaceutical proteins and AppA and PhoA model proteins have been exported with the help of CyDisCo, and there is a knowledge gap on how efficient CyDisCo is at promoting translocation of other target protein through the Tat pathway, especially when it comes to multi-DSB containing POIs (Chapter 2). In Chapter 3, we set out to test whether Tat + CyDisCo could be scaled-up for industrial fermentation and produce competing yields compared to its counterpart Sec pathway. For this, a previously tested Tat-dependent scFv was chosen and experiments were carried out in two different *E. coli* strains: BL21 and W3110.

Moreover, we compared those *E. coli* strains in terms of product yields and product heterogeneity when exporting by Sec pathway in 5 L bioreactors. Finally, in the Chapter 4, we report it is possible to get very high level of secretion to the periplasm and media of a disulphide-bonded protein (YebF) by Tat pathway when fusing AmiC and MdoD SPs in a CyDisCo-dependent manner.

CHAPTER 2: Proteins with multiple disulphide bonds as a substrate for the Tat pathway

2.1. Introduction

Escherichia coli is a favored choice of host to produce biopharmaceutical proteins, and protein export is almost invariably carried out by the Sec pathway. Sec transports proteins through the membrane-bound translocase in an unfolded form, and once in the periplasm, the DSBs of the protein are formed (Karyolaimos et al., 2019). Targeting the POI to the periplasm offers several advantages to produce therapeutic proteins, including simplification of downstream processing via a reduction in the amount of contaminating proteins and nucleic acids (Bagherinejad et al., 2016). The Sec system cannot efficiently export some industrially relevant proteins usually because of premature folding in the cytoplasm, or inability to correctly fold in the periplasm. The Tat pathway offers a potentially important alternative to the Sec system. This pathway also exports proteins to the periplasm. As with Sec substrates, Tat substrates are synthesised with N-terminal signal peptides, but these contain Tat-specific determinants including the presence of a highly conserved twin-arginine motif (Kleiner-Grote et al., 2018). Tat dependent native substrates range in size from 10 kDa to 150 kDa (Berks et al., 2000). The most important difference between the Sec and the Tat pathways to stress upon is that the Sec pathway exports unfolded proteins while the Tat pathway translocates proteins that have been folded in the cytoplasm (Green & Mecsas, 2016).

Moreover, the Tat machinery has an inherent preference not only to export folded proteins, but to export correctly folded proteins, in the sense that misfolded proteins purposely fail to get exported. This trait implies a quality control element that may lead to a more homogeneous product (DeLisa et al., 2003; Fisher et al., 2008). A recent study showed that Tat could export human growth hormone (hGH) at high levels and that the TatExpress (TE) cells could be employed under fermentation conditions (Guerrero-Montero et al., 2019). TE strains, over-express the *tatABC* genes (encoding the Tat system) from the chromosome and boost the industrial potential of this pathway (Browning et al., 2017). Prior to this work, studies reported

Tat to be capable of exporting several biotherapeutics, including single chain antibody fragments, and interferons (DeLisa et al., 2003; Tullman-Ercek et al., 2007; Alanen et al., 2015; Browning et al., 2017). Other simple proteins such as green fluorescent protein (GFP) have been also successfully exported by Tat (Thomas et al., 2001; Matos et al., 2012).

However, this feature poses problems for the export of disulphide-bonded proteins, because these types of proteins often only achieve a native conformation after formation of the DSBs. In wild-type *E. coli* cells, this only occurs in the periplasm (Matos et al., 2014). One study confirmed this by expressing multiple disulphide bond-containing proteins in a mutant strain, Δgor/ΔtrxB, that does allow DSB formation in the cytoplasm. It was shown that if a Tat-specific signal peptide was attached, the proteins could be exported by Tat, but the same constructs were not exported in wild-type strains. Δgor/ΔtrxB strains enable the formation of DSBs in proteins in the cytoplasm by the removal of the two naturally occurring reducing pathways (Prinz et al., 1997). Nevertheless, they are inefficient at making disulphide bond-containing proteins as they do not contain any DSB formation catalyst. CyDisCo (cytoplasmic disulphide bond formation in *E. coli*) mimics these natural systems by introducing a catalyst of disulphide bond formation, usually the sulfhydryl oxidase Erv1p, and a catalyst of disulphide isomerization, usually human protein disulphide isomerase (PDI). This system allows disulphide bond formation in wild-type *E. coli* with the reducing pathways intact and allows efficient production of disulphide bond containing eukaryotic proteins in the cytoplasm (Hatahet et al., 2010). In 2014, an alkaline phosphatase (PhoA, two DSBs), a phytase containing four disulphide bonds (AppA) and anti-interleukin 1β scFv proteins (two DSBs) were exported in an active form by Tat in the presence of CyDisCo components, and mass spectrometry showed that the majority of the scFv protein was disulphide-bonded and correctly cleaved (Matos et al., 2014). This evidence indicates that the combination of Tat + CyDisCo offers a novel means of exporting active, correctly folded disulphide bonded proteins to the periplasm.

Brazzein is a promising natural, low-caloric sweetener candidate appreciated for its intense sweetness, sugar-like taste (Assadi-Porter et al., 2000). Brazzein is the smallest sweet protein and has a molecular mass of 6.4 kDa. This protein contains one short α-helix and three strands of

antiparallel β-sheet held together by four disulphide bonds (Caldwell et al., 1998) (Annex 2). For food industry applications, high-level production of Brazzein is necessary. The recombinant expression of Brazzein was described in various hosts, such as *E. coli* (Assadi-Porter et al., 2000; Assadi-Porter et al., 2008; Lee et al., 2010), *Lactococcus lactis* (Berlec et al., 2006), *Pichia pastoris* (Poirier et al., 2012) and *Kluyveromyces lactis* (Jo et al., 2013). The functional production of Brazzein was frequently low because the protein contains four disulphide bonds and needs a specific terminal sequence (confers its sweet taste). In *E. coli*, the recombinant Brazzein was expressed as a fusion protein that was produced in an insoluble form in the cytoplasm what increased the difficulty of purification. Alternatively, Tat-mediated secretion was reported in *Bacillus subtilis* (Barnett et al., 2009).

The first Fc fusion protein to receive FDA approval was dulaglutide. In 2014, this therapeutic started to be sold for weekly administration for treatment of type 2 diabetes. It consists of two identical single chain subunits linked by two interchain DSBs and four intrachain DSBs (six DSBs in total). This protein has a mass of almost 60 kDa in oxidized conditions and is produced for commercialization in CHO cells (Jimenez-Solem et al., 2010; Sanford, 2014) (Annex 2).

Romiplostim, is a Fc-peptide fusion protein (peptibody) that binds to and activates the thrombopoietin receptor, thereby increasing platelet production. Similar to Dulaglutide, Romiplostim consists of two identical single chain subunits linked by two interchain DSBs and four intrachain DSBs (six DSBs in total). This protein has a mass of almost 60 kDa in oxidized conditions and is expressed intracellularly in *E. coli* as inclusion bodies using high cell density fed batch fermentation for commercialization (Frampton et al., 2009; Keating, 2012) (Annex 2).

Here, we examine whether CyDisCo can be combined with the Tat pathway for the efficient export of more complex, high-value, disulphide-bonded difficult-to-express proteins: Brazzein, Dulaglutide and Romiplostim. Tat secretion system was considered as it can secrete folded proteins which might simplify and lower the cost of recovery of the proteins. At the same time, the proofreading mechanism of Tat has never been tested for more complex proteins, so valuable information of Tat translocation mechanism will be obtained and stoichiometries for the CyDisCo, Tat and substrate proteins will be evaluated.

2.2. Materials and methods

2.2.1. Suppliers of chemicals

All chemicals used in this study were supplied by Fisher Scientific UK (Thermo Fisher Scientific Inc, USA), Sigma (Sigma-Aldrich, USA) or Formedium (UK) unless otherwise stated.

2.2.2. DNA Techniques

2.2.2.1. Preparation of plasmid DNA

All plasmids were isolated through mini preparation using QIAprep spin miniprep kit (Qiagen, Hilden, Germany) as per the manufacturer's instructions. Plasmids were recovered in 50 µL elution buffer (10 mM Tris-Cl pH 8.5) and concentration was determined using NanoDrop 2000c (Thermo Fisher Scientific Inc, USA) before use immediately or stored at -20°C until needed.

2.2.2.2. Polymerase Chain Reaction (PCR)

PCR reaction mixes were made using 1 µL template DNA (80-100 ng), 1 µL dNTP mix (200 µM, Promega, USA), 10 µL GC buffer, 0.3 µL of each primer (0.5 µM), 1.5 µL DMSO (3%), 0.5 µL Phusion high fidelity DNA polymerase (2 U/µL, New England Biolabs, UK) and made up to a final volume of 50 µL with Milli-Q H_2O (GenPure UV/UF water purification system). PCRs were generally carried out in a Bio-Rad T100^{TM} Thermal Cycler (Thermo Fisher Scientific Inc, USA) and then analysed through agarose gel electrophoresis (2.2.2.3.). The primers used in this study are presented in Table 2. An example cycle is given below:

Initial denaturation	98 ˚C	30 sec	
Denaturation	98 ˚C	10 sec	30x
Annealing	Primer T_m	30 sec	
Elongation	72˚C	30 sec/kb	
Final elongation	72˚C	5 min	
Hold	4 ˚C	∞	

Table 2. Cloning primers used in this work. Primers have been supplied by integrated DNA technologies.

Primer	**Sequence 5'→ 3'**
TorA_R	CGCAGTCGCACGTCGCGGC
His_F	CACCATCACCATCACCATTAATAAG
TorA + brazzein_F	CGCCGCGACGTGCGACTGCGCAGGACAAGTGCAAGAAA
His + brazzein_R	TAATGGTGATGGTGATGGTGATATTCGCAGTAATCGCA
TorA + romiplostim_F	CGCCGCGACGTGCGACTGCGATGGACAAGACCCATACCTGCCCGC
His + romiplostim_R	TAATGGTGATGGTGATGGTGCGCACGCGCCGCCAGCCACT
Tor A + dulaglutide_F	CGCCGCGACGTGCGACTGCGCACGGCGAGGGTACCTT
His + dulaglutide_R	TAATGGTGATGGTGATGGTGACCCAGGCTCAGGCTCAG
Plasmid + TorA_F	AACTTTAAGAAGGAGATATAATGAACAATAACGATCTCTTTCAG
Plasmid + His_R	CTAGTGAATTCGGATCCTTATTAATGGTGATGGTGATGGTG
Plasmid1_PMJS162_R	TATATCTCCTTCTTAAAGTTAAACAAAA
Plasmid2_PMJS162_F	CACCATCACCATCACCATTAATAAG

2.2.2.3. Agarose gel electrophoresis

DNA was run on agarose gels containing 1% (w/v) agarose dissolved in 1x TAE buffer. 1x TAE buffer was made from 50x TAE stock solution (242 g/L Tris, 57.1 mL/L glacial acetic acid, 100 mL/L 0.5 M EDTA pH 8). Typically, 0.5 g agarose and 50 mL 1 x TAE buffer were put into a conical flask and microwaved for 1 min until the agarose had dissolved. The solidified gel was then placed in the tank, submerged in 1 x TAE buffer. The DNA samples were mixed with SYBR Green Nucleic Acid Gel stain and 6x Gel loading buffer before loading into the wells of the solidified gel. Electrophoresis was carried out at 120 V for 30 to 50 mins depending on the expected sizes. For the visualisation of the agarose gels a Bio-Rad Gel doc (Bio-Rad Laboratories Ltd, USA) was used.

2.2.2.4. Purification of DNA from agarose gels

Appropriate bands were excised from agarose gels under UV transilluminescence using a sharp scalpel blade. Agarose was then removed using QIAprep gel extraction kit (Qiagen, Hilden,

Germany) as per manufacturer's instructions. DNA was recovered in 30 µL elution buffer before use or stored at -20 °C until needed.

2.2.2.5. Gibson assembly cloning

For constructs designed by Gibson cloning, a 20 µL reaction mix was prepared using a 1:10 dilution of the primers. PCRs were performed using Bio-Rad T100™ Thermal Cycler. Each contained, 0.3 µL template DNA (80–100 ng). Furthermore 0.1 µL Phusion high fidelity DNA polymerase (2U / mL), 0.4 µL dNTP mix (10 mM), 4 µL 5X Phusion HF buffer, and 0.8 µL each of forward and reverse primer (10 µM) were added and the mix was made up with 13.6 µL with autoclaved dH_2O. As previously described the samples were visualized using agarose electrophoresis. After adding 0.5 µL DpnI to each sample, the samples were incubated for 1 hour at 37˚C. The Gibson assembly mix was set up as described in Table 3 and the assembly mix was incubated for an hour at 50˚C before transformation into NEB Turbo competent cells.

Table 3. Gibson assembly composition.

DNA fragments	0.02 - 0.5 pmols
Gibson Assembly Master Mix (2x)	10 µL
Add dH_2O up to	20 µL

2.2.2.6. Sequencing of plasmid DNA

For this study, the sequencing service of GENEWIZ was used. Typically, 80 to 100 ng of DNA was mixed with 5 µL of 5 mM sequencing primer (pTacF: 5'-GAGCGGATAACAATTTCACACAGG- 3')

2.2.2.7. Constructs generated and used in this study

A variety of constructs has been used in this study, which are presented in Table 4.

Table 4. Construct used and generated in this study.

Plasmid name	Function	Tag	Promoter	Antibiotic resistance	Reference
pKA1	pEXT22 derivative / TorA-Romiplostim-His6	His6 (C-terminal)	*tac*	Kanamycin	This study
pKA2	pEXT22 derivative / TorA-Romiplostim-His6 + CyDisCo	His6 (C-terminal)	*tac*	Kanamycin	This study
pKA3	pEXT22 derivative / TorA-Dulaglutide-His6	His6 (C-terminal)	*tac*	Kanamycin	This study
pKA4	pEXT22 derivative / TorA-Dulaglutide-His6 + CyDisCo	His6 (C-terminal)	*tac*	Kanamycin	This study
pKA5	pEXT22 derivative / TorA-Brazzein-His6	His6 (C-terminal)	*tac*	Kanamycin	This study
pKA6	pEXT22 derivative / TorA-Brazzein-His6 + CyDisCo	His6 (C-terminal)	*tac*	Kanamycin	This study
pKA7	pET23 derivative / TorA-Brazzein-His6	His6 (C-terminal)	*tac*	Ampicillin	This study
pKA8	pET23 derivative / TorA-Dulaglutide-His6	His6 (C-terminal)	*tac*	Ampicillin	This study
pKA9	pET23 derivative / TorA-Romiplostim-His6	His6 (C-terminal)	*tac*	Ampicillin	This study
pAAT1	pET23 derivative / His6-Romiplostim	His6 (N-terminal)	*tac*	Ampicillin	Prof. Ruddock (University of Oulu)
pAAT7	pET23 derivative / His6-Dulaglutide	His6 (N-terminal)	*tac*	Ampicillin	Prof. Ruddock (University of Oulu)
PMJS256	pET23 derivative / His6-Brazzein	His6 (N-terminal)	*tac*	Ampicillin	Prof. Ruddock (University of Oulu)
PMJS205	CyDisCo: Erv1p and PDI	-	*tac*	Chloramphenicol	Gąciarz A et al. (2016)
pKRK7	pEXT22 derivative / TorA-hGH-His6	His6 (C-terminal)	*tac*	Kanamycin	Browning et al. (2017)

2.2.3. Growth and maintenance of *E. coli* cultures

2.2.3.1. Glycerol stocks

The constructs used in this study were transformed in NEB Turbo cells and stored as glycerol stocks at –80 °C. Glycerol stocks were prepared using 750 µL 50% glycerol and 750 µL stationary phase culture.

2.2.3.2. Preparation of competent cells

A 5 mL LB pre-culture, with appropriate antibiotics (at 1:1000 concentration), was inoculated with *E. coli* strain and incubated o/n at 30°C, 200 rpm in a shaking incubator (Multitron Pro shaking incubator, Infors, Switzerland). The next day 10 mL fresh LB was inoculated from the pre-culture and incubated at 30°C 200 rpm until OD600 = 0.3 – 0.4 at which point cells were pelleted by centrifugation (Thermo Fisher Scientific Inc, USA) at 1690 x g, 4°C, 10 min. Supernatant was discarded and the cell pellets were resuspended in 10 mL ice-cold 100 mM $MgCl_2$ and incubated on ice for 5 min. Cells were re-pelleted before resuspending in 1 mL ice-cold 100 mM $CaCl_2$ and stored at 4°C for 1 – 24 hrs before use.

2.2.3.3. E. coli strains used in this study

The *E. coli* strains presented in Table 5 were used in this study.

Table 5. Strains used in this study.

Strain	Genotype	Reference/ Source
NEB® Turbo Competent *E. coli*	*F' proA$^+$B$^+$ lacI^q ΔlacZM15 / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-210::Tn10) Tet^S endA1 thi-1 Δ(hsdS-mcrB)5*	NEB®
BL21	*E. coli* B strain. *fhuA2 [lon] ompT gal [dcm] ΔhsdS*	NEB®
BL21 TatExpress	BL21 carrying a pTac promotor upstream of *tatABCD*	Browing et al. (2017)

2.2.3.4. Media

For growth of *E. coli* in liquid media, Lysogeny Broth (LB) medium (10 g/L sodium chloride, 10 g/L tryptone and 5 g/L yeast extract) was used for the experiments in 2.3.1 as a main culture. For

growth on agar plates, Luria Bertani agar (LBA) (10 g/L sodium chloride, 10 g/L tryptone, 5 g/L yeast extract and 10 g/L bacto-agar) was used. At the end of the section 2.3.1 and onwards, the *E. coli* cells were grown in rich autoinduction medium as a main culture (12 g/L tryptone, 24 g/L yeast extract, 3.3 g/L $(NH_4)_2SO_4$, 6.8 g/L KH_2PO_4, 7.1 g/L Na_2HPO_4, 0.5 g/L glucose, 2 g/L lactose, 0.15 g/L $MgSO_4$ and 0.03 g/L Trace Elements, giving a total of 55.85 g/L). The final mix of 55.85 g of powdered medium was dissolved in 1 L of distilled water. A final concentration of 0.8% glycerol was added to the prepared media. Suitable antibiotics based on the vector were used at the following concentrations: Kanamycin (50 µg/mL), Ampicillin (100 µg/mL) and Chloramphenicol (35 µg/mL).

2.2.3.5. Transformation of competent E. coli cells

To transform *E. coli* competent cells, 1-2 µL of DNA (100 ng/µL) was gently mixed with 30 µL competent cells and incubated on ice for 30 minutes. The cells were then placed in a water bath (Grant Instruments, Cambridge, UK) at 42°C for exactly 30 seconds before placing back on ice for 5 min. Next 0.5 mL LB was added, and cells incubated at 37°C for 60 min. Total cell suspension was plated on LBA plates containing appropriate antibiotic(s) which were incubated in plate incubator (MLR-162-PE, Panasonic, Japan) inverted, at 37°C overnight.

2.2.4. Protein production

2.2.4.1. Culture of E. coli and plasmid induction

For fractionation of *E. coli* cells (section 2.2.4.2.) in LB media expression experiments, 5 mL LB with appropriate antibiotics was inoculated and grown aerobically at 30°C, 220 rpm overnight. The following day cultures were diluted to OD_{600}= 0.05 in 50 mL fresh LB with antibiotic(s) in 250 mL Erlenmeyer flask before incubation at 30°C until OD600= 0.4 – 0.6 (mid-log phase). At this point, the cells were induced with 100 µM IPTG for 3 hours or o/n (typically 16 hours).

For fractionation of *E. coli* cells grown in rich autoinduction media for protein expression, after plate overnight incubation at 37°C, the colonies were used to inoculate 2-5 mL of LB media supplemented with 2 g/L of glucose. These starter cultures were grown for 6 hours at 30°C, 250 rpm and were used to seed the main cultures in a 1:100 ratio. Main cultures cells were grown in

100 mL flasks with 10 mL culture in each flask. The flasks were covered with sponge as oxygen permeable membrane and grown at 30°C, 250 rpm for 22-24 hours and harvested for fractionation.

2.2.4.2. Fractionation of the E. coli cells

After the induction period, the cells equal to $10/OD_{600}$ were harvested by centrifugation at 1690 rpm for 10 min at 4 °C (Eppendorf centrifuge, 5417R, Hamburg) and EDTA/ lysozyme/ cold osmotic shock protocol for periplasmic extraction was carried out (Randall & Hardy, 1986) with some modifications (Pierce et al., 1997). Samples were placed on ice and the supernatant was discarded. The pellets were resuspended in ice cold 500 µL Buffer 1 (100 mM Tris-acetate pH 8.2, 500 mM sucrose, 5 mM EDTA). In addition, 500 µL ice cold MilliQ H_2O and 40 µL lysozyme (from egg-white, 1 mg/mL stock (Sigma Aldrich, UK)) was added, and the suspension was incubated on ice for 5 min. After 5 mins, 20 µL 1 M $MgCl_2$ was added to stabilise the cytoplasmic membrane, and cells were centrifuged at 20800 x g, 4 ºC for 2 mins. The supernatant, periplasmic fraction (P) was collected in a 1.5 mL Eppendorf without disturbing the cell pellet. The remaining supernatant around the pellet was discarded and the pellet was washed with 1 mL ice cold Buffer 2 (50 mM Tris-acetate pH 8.2, 250 mM sucrose, 10 mM MgSO4) and centrifuged at 20800 x g, 4 ºC for 5 mins. The supernatant was discarded, and the pellet was resuspended in 1 mL Buffer 3 (50 mM Tris-acetate pH 8.2, 2.5 mM EDTA). To fully lyse the cells, they were sonicated (10 secs on / 10 secs off 4x) and then ultra-centrifuged at 264,360 x g, 4 ºC for 30 mins. This guarantees the separation of the cytoplasm and insoluble material. The supernatant was collected as cytoplasmic fraction (C) in a 1.5 mL Eppendorf without disturbing the cell pellet. The remaining supernatant was discarded, and the pellet was resuspended in 1 mL Buffer 3 (50 mM Tris-acetate pH 8.2, 2.5 mM EDTA), giving the insoluble fraction (I). For further analysis of membrane fraction (MB), after sonication, samples are centrifuged at 20800 x g at 4 ºC for 15 mins. The supernatant obtained was ultra-centrifuged at 264,360 x g, 4 ºC for 30 mins (the final pellet was again resuspended 1 mL Buffer 3 giving MB fraction), and the pellet was resuspended in 1 mL Buffer 3 (inclusion bodies, IB). All cell fractions were stored frozen at – 20 °C.

2.2.5. Protein separation

2.2.5.1. SDS poly-acrylamide gel electrophoresis (SDS-PAGE)

Protein separation was carried out using a vertical gel apparatus from Bio-Rad Mini-PROTEAN® Tetra System (UK), the gels were cast and run according to manufactures instructions. Typically, 0.75 mm gels were used. The separation gel was composed of 15% acrylamide, 0.3% bis-acrylamide (37.5:1, BioRad, Herts, UK), 375 mM Tris-HCl pH 8.85 and 0.1 % SDS. Stacking gel was composed of 5% acrylamide, 0.0375 % bis-acrylamide 125 mM Tris-HCl pH 6.8 and 0.001% SDS. Before loading, samples were mixed with protein gel loading buffer (125 mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 0.02% bromophenol blue, 5% β-mercaptoethanol) and heated for 5 minutes at 95°C. Protein gel running buffer was 25 mM Tris, 192 mM Glycine and 0.1% SDS at a pH of 8.3. Gels were run until the dye-front had run off the gel, typically 40 minutes at 60 mA.

2.2.5.2. Detection of proteins with Coomassie

The protein profile of samples was visualised with Coomassie stain (10% acetic acid, 40% methanol and 1 g/L Coomassie Brilliant Blue R-250 blue (Bio-Rad, Herts, UK)). Newly run SDS gels were covered in Coomassie stain and incubated at room temperature for 1 h. The stain was removed by washing the gel in destaining solution (5% ethanol, 7.5% acetic acid).

2.2.6. Protein imaging

2.2.6.1. Western-blotting

Proteins of interest were also visualised using electrophoresis for wet-western blotting. Proteins were transferred to PVDF-membrane (GE Healthcare, Buckinghamshire, UK) using the Bio-Rad system. The SDS-PAGE was placed on top of the membrane in between two sheets of Whatman paper and two sponges whilst being soaked in western blot transfer buffer (192 mM Glycine, 25 mM Tris and 10% ethanol). Membrane was prepared by soaking in methanol prior to contact with the gel. Proteins were transferred at 90 V for 1 hr.

2.2.6.2. Detection of proteins by immunoblotting

After transfer, the PVDF membrane was immersed in a blocking buffer (2.5% (w/v) skimmed milk powder in 50 mL 1x PBS-tween20 (0.1%)) at 4°C overnight. On the following day, the blocked membrane was washed 3x 5 min with 1x PBS-tween20 (0.1%) before incubating with 1:5000 primary antibody (anti-His C-terminal, Invitrogen) in 20 mL 1x PBS-tween20 (0.1%) for 1 hour at room temperature on shaking table (GyroTwister, Labnet, New Jersey, USA). The membrane was then washed for 3x 5 min with 1x PBS-tween20 (0.1%) before incubating with 1:5000 secondary antibody (anti-Mouse-HRP, Promega) in 20 mL 1x PBS-tween20 (0.1%)) for 1 hour at room temperature on a shaking table. Finally, the membrane was washed for 6x 5 min with 1x PBS-tween20 (0.1%) before imaging. Immunoreactive bands were detected using enhanced chemiluminescence detection kit (ECL, Bio-Rad, Herts, UK) according to the manufacturer's instructions. Bands were visualised using BioRad Gel-doc chemiluminescence imager and associated software.

2.3. Results

2.3.1. Low copy number polycistronic one-plasmid based format (pEXT22) for Tat dependent export of disulphide-bonded proteins

Gene expression in bacteria is organised into operons that contains polycistronic mRNAs which encode several proteins. This has the advantage of allowing the simultaneous control of the production of several proteins using a single promoter and terminator. This avoids, for example, the multiple uses of identical promoters and terminators, which may diminish overall promoter activity and increase the metabolic burden on the cell during protein production (Müntjes et al., 2020). At the same time, many vectors used for expression of proteins are low copy number plasmids as having a reduced copy number (lower gene dosage) can decrease basal, or uninduced, intracellular levels of an expressed protein. This is advantageous because some foreign proteins are toxic even at very low levels and therefore can inhibit growth of the *E. coli* cells (Wood et al., 2017). Also, for the aim of this research, it was hypothesized that low copy number plasmid (pEXT22) would avoid overloading the Tat system and permit the translocation of the POI to the periplasm (Sutherland et al., 2018).

Previously, the successful export of various prefolded, disulphide-bonded recombinant proteins to the periplasm by the Tat pathway with CyDisCo strains has been reported (Matos et al., 2014). Based on this research, we aimed to successfully transport prefolded industrially relevant proteins namely Brazzein, Dulaglutide, Romiplostim as POIs from the cytoplasm to the periplasm by the Tat pathway.

For this aim, the TorA Tat specific SP was attached to the N-terminus of each POI for export and a C-terminal 6x-His tag was used for detection (referred as TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim). Constructs were expressed in a one-plasmid based format, specifically a polycistronic modified pEXT22 with a pTac promoter and kanamycin resistance (Dykxhoorn et al., 1996) that carries also genes for co-expression of Erv1p (*Saccharomyces cerevisiae*) and PDI (human), CyDisCo components (Matos et al., 2014). CyDisCo would catalyze disulphide-bond formation in the cytoplasm and Tat pathway would recognize the proteins as correctly folded and

export them to the periplasm from where they could be harvested and used for analytical and downstream experiments. As a negative control, a plasmid with the same features but without CyDisCo components was cloned for each POI.

For these first experiments, POIs were expressed with TorA SP in LB media at 30˚C for 3 h in an *E. coli* B strain (BL21) in shake-flask at 220 rpm and *tac* promoter was induced with 100 µM of IPTG in the presence and absence of CyDisCo. Strains employing CyDisCo system must be cultivated at 30°C or lower temperatures during protein induction phase because Erv1p shows temperature dependent folding. After harvesting, the cells were fractionated as described in the section 2.2.4.2. in Materials and Methods and protein production was assessed by SDS-PAGE. Protein identification was done by SDS-PAGE stained with Coomassie and visually, cells' fractions were compare with the aim of assessing fractionation technique quality. POI detection was carried out by blotting against the C-terminal His tag (Western blot, WB).

Figure 8A shows a WB of the fractionation of TorA-Brazzein in a polycistronic plasmid in the absence and presence of CyDisCo in LB media in the conditions described above. We expected that in the absence of CyDisCo, and considering that Brazzein has four essential DSBs (Caldwell et al., 1998) (see Annex 2 for structure and sequence data), the protein would be insolubly produced and not transported to the periplasm. In the presence of CyDisCo, disulphide bonds would be catalyzed, and Tat would transport the protein to the periplasm. The TorA SP does not appear to support Brazzein export, and no protein could be visualised on the WB either in the absence or in the presence of CyDisCo in any of the cell fractions. TorA SP-Brazzein-His (precursor) is expected to be 11.5 kDa in reducing conditions, while Brazzein-His (mature) is 7.3 kDa in reducing conditions.

Figure 8B shows a Coomassie-stained gel of the cytoplasm (C), insoluble fraction (I) and periplasm (P) from a 3 h induction of TorA-Dulaglutide and TorA-Romiplostim in the same polycistronic plasmid in the absence and presence of CyDisCo in the conditions described above (see Annex 2 for structure and sequence data). Our collaborators at University of Oulu report high soluble yields in the presence of CyDisCo of these two POIs expressed in the cytoplasm without a SP. In the Coomassie-stained gel in Figure 8B, it is evident that the protein profiles of each cell

fraction are distinct, and the pre-POI (pre-Dulaglutide and pre-Romiplostim) are presumably identified in the insoluble fraction of the cell. This suggests no export or low export efficiency of the POIs by Tat pathway with TorA SP, while the WB of the same gel as in Figure 8B, confirms the insoluble state of the pre-POIs and the absence of export to the periplasm (Figure 8C). TorA SP-Dulaglutide-His (precursor) is expected to be 34.4 kDa in reducing conditions, while Dulaglutide-His (mature) is 30.2 kDa in reducing conditions. TorA SP-Romiplostim-His (precursor) is expected to be 34.8 kDa in reducing conditions, while Dulaglutide-His (mature) is 30.6 kDa in reducing conditions.

Furthermore, expression of the same constructs was carried out under the same conditions as explained before but lowering the culture temperature to 25˚C. Lower expression temperatures have been shown to facilitate the production of folded, soluble proteins (Francis & Page, 2010). In the 25˚C expression experiments, samples were taken 3 hours post induction, and after overnight expression (approximately 16 hours post-induction). TorA-Brazzein was not found to be expressed at 25˚C as seen in the WB in Figure 9A and different harvesting points do not show any differences in the yields of this POI.

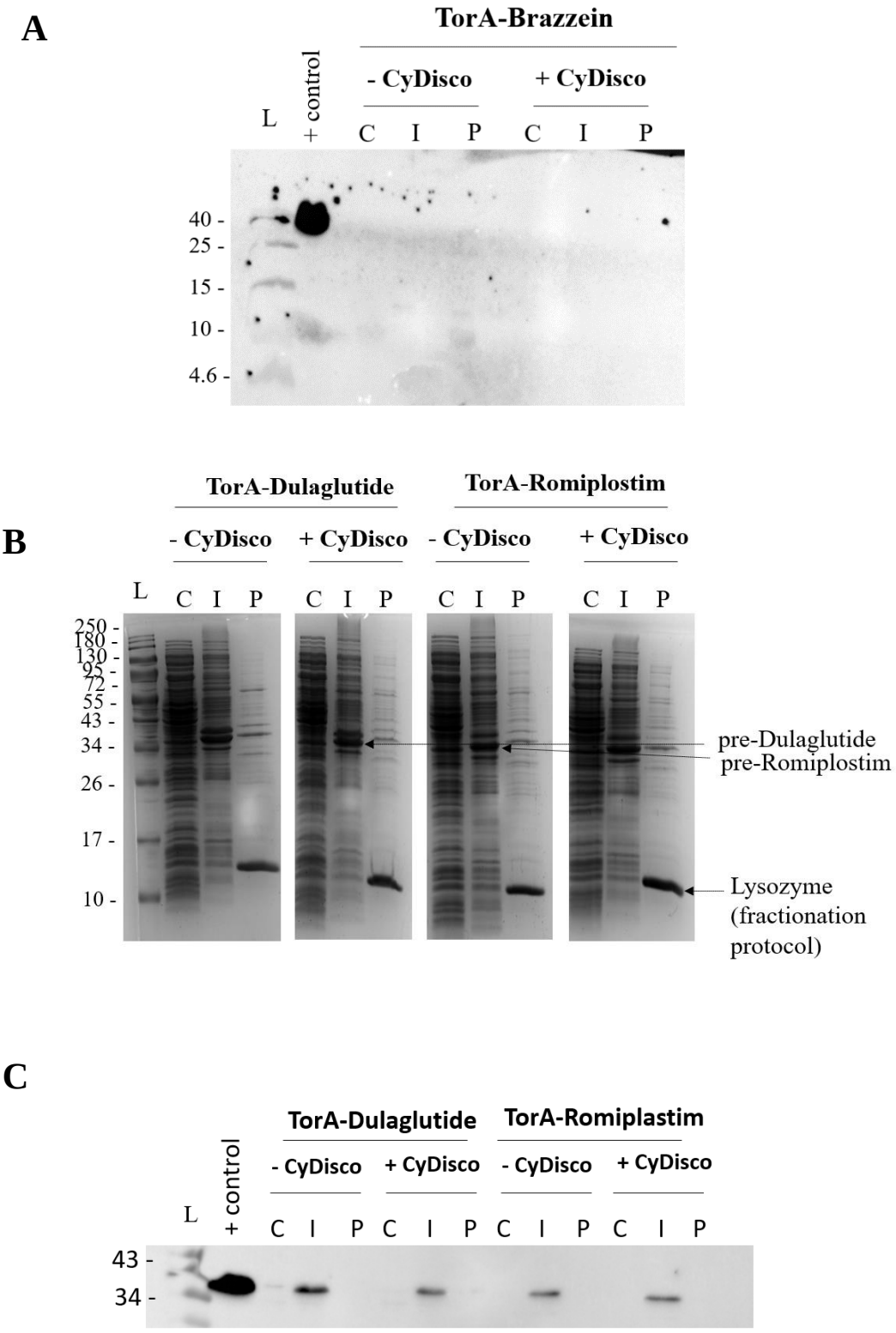


Figure 8. Expression of TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim in a polycistronic plasmid in the presence/absence of CyDisCo in LB media at 30˚C. C-terminal His-tagged TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim were expressed in the presence (+) and absence (-) of CyDisCo in BL21 strain in 250 mL shake flask with 50 mL media at 220 rpm with a sponge as a O_2 permeable membrane. After 3 h induction, cultures were fractionated into cytoplasm (C), insoluble fraction (I**)** and periplasm fraction (P) using EDTA/ lysozyme/ cold osmotic shock for periplasm extraction. The rest of the cells were lysed by sonication. **A:** Blank WB for fractionation of cells expressing TorA-Brazzein in the absence/presence of CyDisCo. **B:** Coomassie-stained gel of the fractionation of TorA-Dulaglutide

and TorA SP-Romiplostim highlighting the distinct profile of C/I/P fractions. With arrows the pre-POIs in the insoluble fraction (I) of the cell. Mature POIs (without the SP) are marked with an asterisk (*) in the insoluble fraction of the cell. **C:** WB for fractionation of cells expressing TorA-Dulaglutide and TorA-Romiplostim in the absence and presence of CyDisCo. In all cases samples were immunoblotted to the C-terminal 6x histidine-tag. Molecular weight ladder (L) is shown on the left in kDa and immunoblot positive control is labelled as "+ control". All the cells' fractions have been diluted in the same volume. The reproducibility of the experiments has not been ascertained.

TorA-Dulaglutide expression at 25˚C in the aforementioned conditions helped in the solubilization of the POI. In the WB in Figure 9B, after 3 h post-induction, a doublet band presumably corresponding to the uncleaved POI (pre-Dulaglutide) and to the mature protein (Dulaglutide) were visible in the cytoplasmic (C) and some in the periplasmic fractions (P) of the cells in the presence and absence of CyDisCo. In the insoluble fraction (I), a prominent band could be detected, presumably corresponding to the precursor of the POI (pre-Dulaglutide). Dulaglutide was found to be partially soluble in the cytoplasm of *E. coli* in the absence of CyDisCo components. This could be attributed to the highly soluble nature of the IgG Fc region. On the other hand, intermolecular DSBs that lead to dimerization, would not be catalyzed without the presence of CyDisCo (see Annex 2 for structure and sequence data). Therefore, a native state of the protein would not be reached, so we could not expect some export to the periplasm as Tat should recognize them as not properly folded and reject them for export. After o/n induction, pre-Dulaglutide is not visible any more in the C and P fraction and only the mature protein is detectable in the C fraction. Insoluble TorA-Dulaglutide seems to be overall the most abundant state of the POI (located in the insoluble fraction) in both time-points. In the Coomassie-stained gel in Figure 9C of the same WB experiment as in Figure 9B, it is possible to identify the two bands presumably corresponding to the uncleaved and mature POI. Also, in this gel we can identify that the periplasmic fraction after 3 h post-induction is cross contaminated with cytoplasmic proteins (compare with the o/n post-induction periplasmic fraction). This fact together with the detection of the uncleaved form of the protein in the periplasm reject any possible export to the periplasm by the Tat pathway.

In Figure 9D, when TorA-Romiplostim is expressed in the same conditions as previously explained, but at 25 ˚C, the POI is detected by WB in the insoluble fraction (I) of the cell in the presence and absence of CyDisCo in both time points. To sum up, lowering the temperature does not help in the solubility or export of TorA-Romiplostim under the tested conditions.

Overall, the expression of these three POIs with TorA SP in LB media in a polycistronic plasmid in the presence of CyDisCo was not successful. Lowering the temperature culture to 25˚C does not help the export by Tat pathway. In the following experiments, CyDisCo optimal performance should be a priority as its improper functioning could be the issue: without the proper activity of CyDisCo, these disulphide-bonded proteins will not be properly folded, therefore, the translocation through Tat system seems improbable.

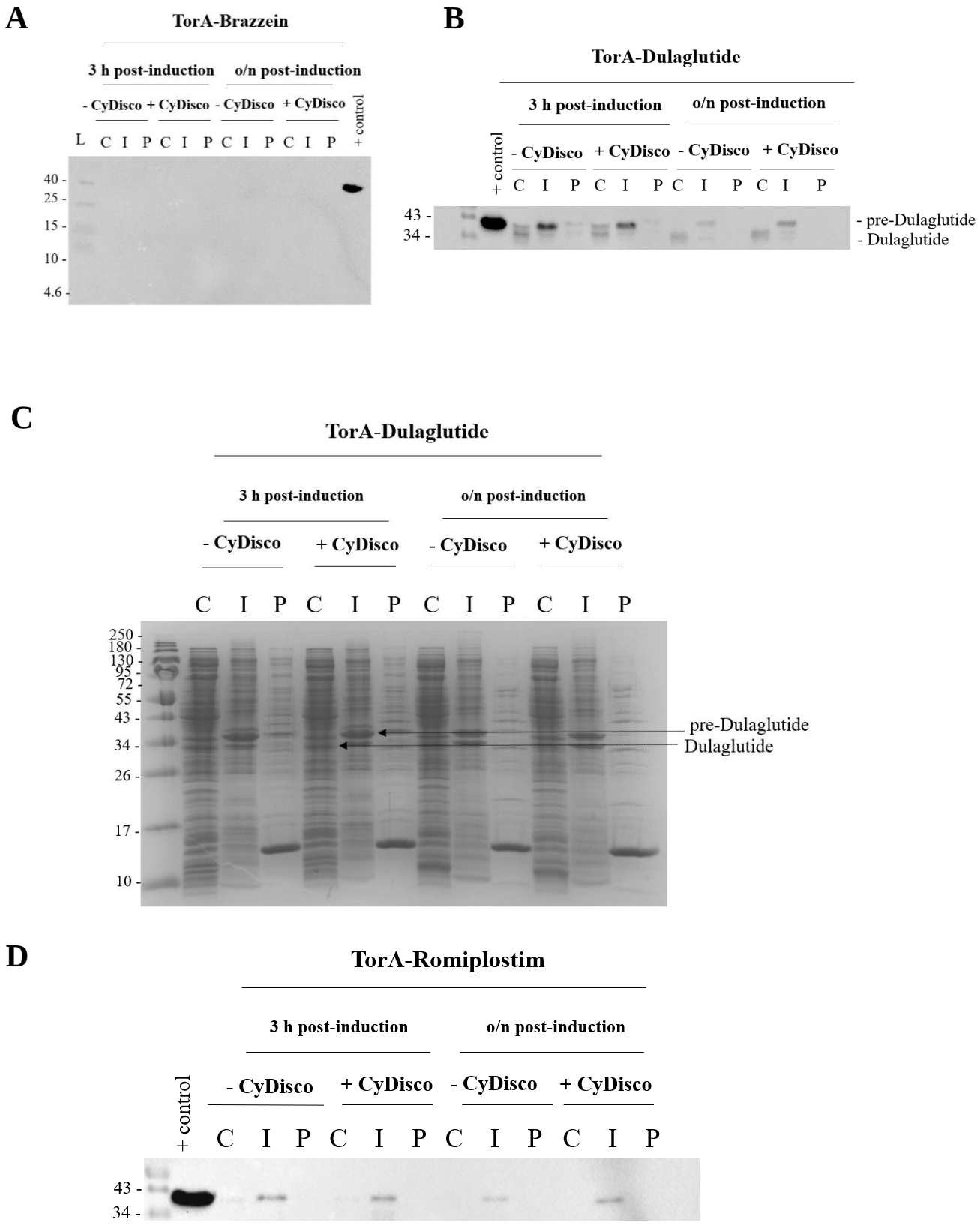


Figure 9. Expression of TorA-Brazzein, TorA-Dulaglutide and TorA -Romiplostim in a polycistronic plasmid in the presence/absence of CyDisCo in LB media at 25˚C. C-terminal His-tagged TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim were expressed in the presence (+) and absence (-) of CyDisCo in BL21 strain in 250 mL shake flask at 220 rpm with a sponge as a O_2 permeable membrane.

After 3 h induction and overnight (o/n) induction, cultures were fractionated into cytoplasm (C), insoluble fraction (I**)** and periplasm fraction (P) using EDTA/ lysozyme/ cold osmotic shock for periplasm extraction. The rest of the cells were lysed by sonication. **A:** Blank WB for fractionation of cells expressing TorA-Brazzein in the absence/presence of CyDisCo. **B:** WB for fractionation of cells expressing TorA-Dulaglutide in the absence and presence of CyDisCo. **C:** Coomassie-stained gel of TorA-Dulaglutide fractionation in the absence and presence of CyDisCo. **D:** WB for fractionation of cells expressing TorA-Romiplostim in the absence and presence of CyDisCo. In all cases samples were immunoblotted to the C-terminal 6x histidine-tag Molecular weight ladder (L) is shown on the left in kDa and immunoblot positive control is labelled as "+ control". All the cells' fractions have been diluted in the same volume. The reproducibility of the experiments has not been ascertained.

The production of recombinant proteins for research purposes is usually carried out in small volumes of culture media, in microtiter plates or shake flasks. Such cultivations are usually batch cultures in which all media components are added at the beginning of the culture. Changes in conditions such as pH, dissolved oxygen and substrate concentration are not controlled. In small scale cultivations, oxygen concentration in the media is often low due to insufficient agitation or inefficient permeability of the flask cover in proportion to the oxygen demand of growing cells (Vasala et al., 2006). *E. coli* is a facultative anaerobic organism that depending on the presence of oxygen can switch between aerobic and anaerobic metabolism (Unden et al., 1994). Moreover, oxygen is crucial for cultivation of strains employing CyDisCo system as Erv1p uses molecular oxygen as the electron acceptor (Gąciarz et al., 2017).

Autoinduction medium (AIM) has recently acquired popularity, and there are several examples in the literature of successful protein production by growing *E. coli* in this medium. At least two carbon sources are present in autoinduction medium: glucose and lactose (glycerol is also added to increase cell growth and hence yields). Glucose is the preferred carbon source, and once it is depleted (usually during exponential growth), lactose starts being consumed. This, in turn, induces protein production from lac promoter-based systems. Using autoinduction medium has various advantages: higher bacterial densities are achieved, the time point of induction is highly reproducible, and the culture no longer needs to be stopped for IPTG addition (Rosano et al., 2019). These advantages that autoinduction media provides are desirable for our research in many ways as it permits more similar bioreactor-like conditions and aeration is not stopped at any time

in the incubators, assuring that Erv1p gets the oxygen needed for its optimal performance. Along with changing the media, we also increased the shaking speed of our cultures to 250 rpm when expressing our constructs for better aeration purposes. At the same time, a longer culture duration will also permit CyDisCo components to be translated and to fully fold and start catalysing the formation of DSBs in our POIs.

Protein expression in rich AIM was tested for all six constructs. AIM was particularly suited to this system as it is formulated to auto induce IPTG inducible promoters, and all constructs are under *tac* promoter. Moreover, the media was developed for B strains (Studier, 2005). The cells were pre-cultured in LB media supplemented with 2 g/L of glucose and corresponding antibiotic for 6 hours and transferred to rich AIM for cell expression. After 22-24 h in this media at 30˚C, 250 rpm, the cells were fractionated into cytoplasmic (C), periplasmic (P) and insoluble fractions and the latter was further separated into inclusion bodies (IB) and membrane proteins (MB) with the aim of further understanding whether our POIs were becoming insoluble or getting stuck in the membrane on their way out to the periplasm. Figure 10 shows the results of these experiments.

Rich AIM allowed, for the first time, to identify TorA-Brazzein in inclusion bodies (IB) in a WB only in the presence of CyDisCo (Figure 10A). TorA-Dulaglutide (Figure 10B) and TorA-Romiplostim (Figure 10C) show very similar WB interpretation: both are expressed in inclusion bodies (IB) in the absence and presence of CyDisCo. Overall, these proteins were not expressed solubly in AIM with/without CyDisCo in a one-plasmid polycistronic system and no export to the periplasm can be identified.

To sum up, it has not been possible to export TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim by Tat pathway in pEXT22 (low copy number polycistronic one-plasmid based format). Different media and temperature did not help in the transport of the proteins to the periplasm.

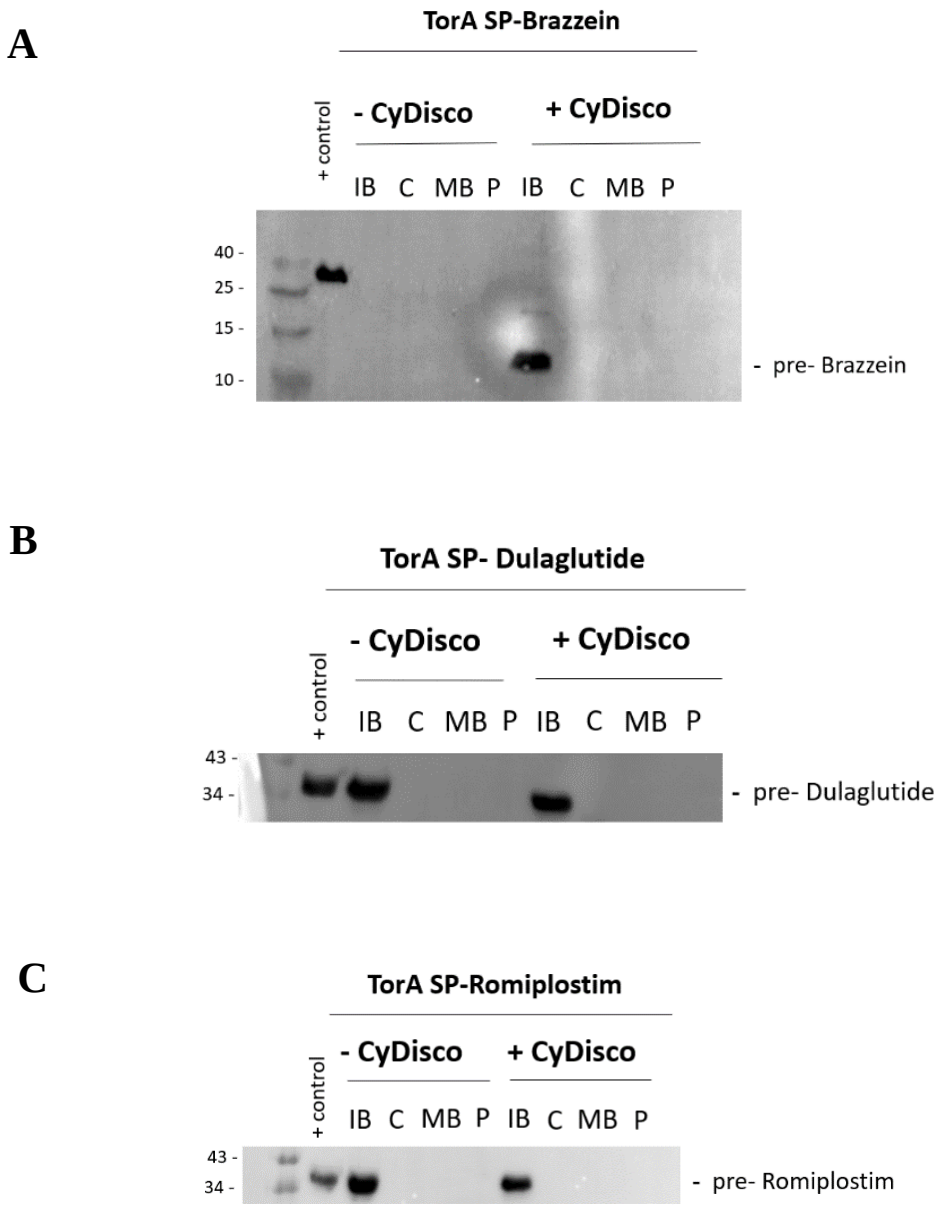


Figure 10. Expression of TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim in a polycistronic plasmid in the presence/absence of CyDisCo in AIM at 30˚C. C-terminal His-tagged TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim were expressed in the presence (+) and absence (-) of CyDisCo in BL21 strain in 100 mL shake flask at 250 rpm with a sponge as a O2 permeable membrane. After 22-24 h of culture, cells were fractionated into, inclusion bodies (IB), cytoplasm (C), membrane (MB) and periplasm fraction (P) using EDTA/ lysozyme/ cold osmotic shock for periplasm extraction. **A:** WB for fractionation of cells expressing TorA-Brazzein in the absence/presence of CyDisCo. **B:** WB for fractionation of cells expressing TorA-Dulaglutide in the absence and presence of CyDisCo. **C:** WB for fractionation of cells expressing TorA-Romiplostim in the absence and presence of CyDisCo. In all cases samples were immunoblotted to the C-terminal 6x histidine-tag. Molecular weight ladder (L) is shown on the left in kDa and WB positive control is labelled as "+ control". All the cells' fractions have been diluted in the same volume. The reproducibility of the experiments has not been ascertained.

2.3.2. Two-plasmid based format for Tat dependent export of disulphide bonded proteins

After numerous attempts to export our POIs and knowing that 70-90% of folding could be achieved when expressing them in the cytoplasm in the presence of CyDisCo, we focused on expressing enough CyDisCo components and/or making it functional.

Our collaborator in Oulu shared with us their standard operating procedure (SOP) for cytoplasmic protein expression involving the usage of a two-plasmid system. In this system, the POI is expressed from a pET23-based plasmid with a pTac promoter and CyDisCo components i.e. Erv1p and PDI are co-expressed from the second plasmid, pLysS-based vector, under a pTac promoter and with a p15A origin of replication (Gąciarz et al., 2017) (Figure 11A).

pET23-based plasmid is a high copy number plasmid (carrying the POI), and pLysS-based vector lies between low to medium copy number (carrying CyDisCo components). In previous work, overexpressing the POI caused the overloading of Tat pathway, preventing the export to the periplasm due to blockage of the system (Sutherland et al., 2018). For this reason, experiments in two-plasmid based system were carried out with the aim of successfully exporting TorA-Brazzein, TorA-Dulaglutide and TorA- Romiplostim to the periplasm, but instead of adopting their POI plasmid, that could overload Tat pathway, we expressed the POIs in the previously discussed pEXT22 (section 2.3.1) without CyDisCo (Figure 11A). For CyDisCo components, their pLysS-based vector was used (*tac* promoter).

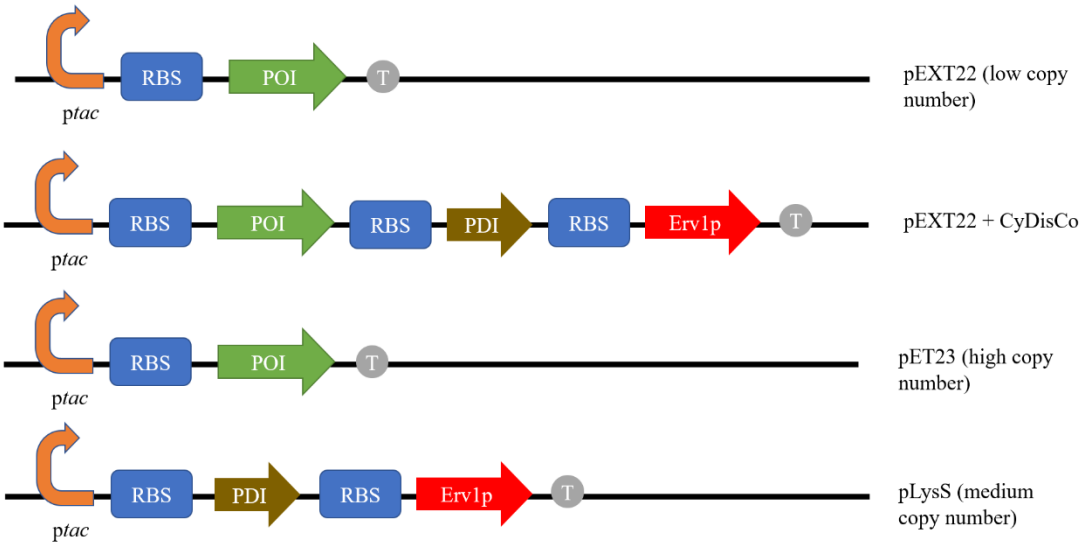
The co-expression was easily setup by a co-transformation of the recombinant plasmids into expression strain *E. coli* BL21. The cells were pre-cultured in LB media supplemented with 2 g/L of glucose and corresponding antibiotics for 6 hours and transferred to rich AIM for cell expression. After 22-24 h in this media at 30˚C at 250 rpm, the cells were fractionated in cytoplasm (C), inclusion bodies (IB), periplasm (P) and insoluble fraction (I). This time, two controls were set up apart from the actual experiment. i) the POI with CyDisCo (two-plasmid system) without any SP as a positive control and ii) the TorA-POI without CyDisCo (one plasmid system, section 2.3.1) as a negative control.

The POI without any SP and with CyDisCo (POI + CyDisCo) in the two-plasmid system will confirm the CyDisCo functionality in our laboratory (positive control). These proteins without SP should mainly remain soluble in the cytoplasm only if we are making enough CyDisCo and only if CyDisCo is functional. In case of finding any of these three POIs in inclusion bodies, a misfunction/mishandling or other problem related to CyDisCo should be considered.

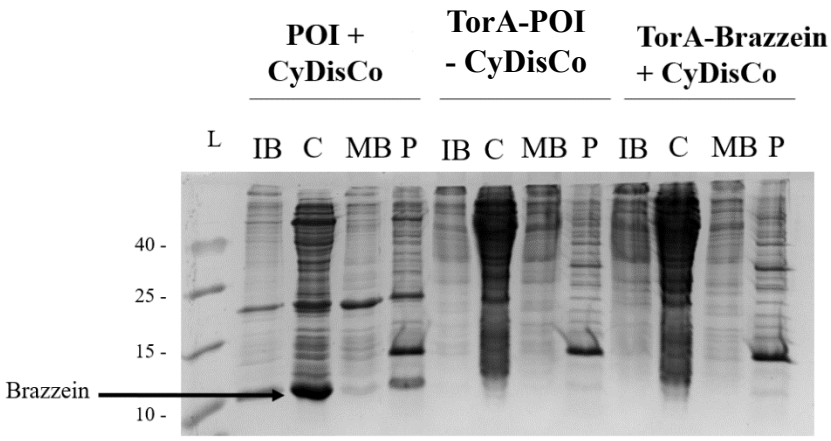
The TorA-POI expressed in pEXT22 system without CyDisCo (TorA-POI – CyDisCo) (negative control) would not be exported, as in the absence of CyDisCo, DSBs will not be catalysed, preventing Tat mediated export.

In Figure 11A TorA-Brazzein export assessment in the aforementioned conditions is shown. The positive control (in the figure as "POI + CyDisCo"), shows the fractionation of His-Brazzein in the two plasmid-system in the presence of CyDisCo. As our collaborators at the University of Oulu reported, and when CyDisCo stoichiometry matches what the POI needs and is on its active form, we identified Brazzein in the cytoplasm of the positive control. In this fractionation, periplasm fraction (P) was found to be contaminated with cytoplasmic proteins (C), as numerous cytoplasmic proteins can be identified in the P fraction. This is probably because in this case, Brazzein is expressed at a higher level than in the previous experiments shown in section 2.3.1, as it was cloned in a high copy number plasmid (pET23) by our collaborators and this causes stress to the cells, making them prone to lysis and more delicate when fractionating. The negative control (in the figure as "TorA-POI - CyDisCo") shows the fractionation of TorA-Brazzein in pEXT22 one plasmid system in the absence of CyDisCo. In this case, and in line with results in section 2.3.1, no POI was identified. Finally, when TorA-Brazzein was tried for export by Tat pathway in the two-plasmid system (CyDisCo components in one plasmid + low copy number plasmid with the POI, in the Figure 11A as "TorA-Brazzein + CyDisCo"), Brazzein could not been identified in any of the fractions of the cells. Unfortunately, periplasmic fractions in both negative control and TorA-Brazzein + CyDisCo were contaminated with cytoplasmic fraction, as it is deduced by the similarity of the two fractions. This is probably due to an inefficient practice of the periplasmic extraction protocol.

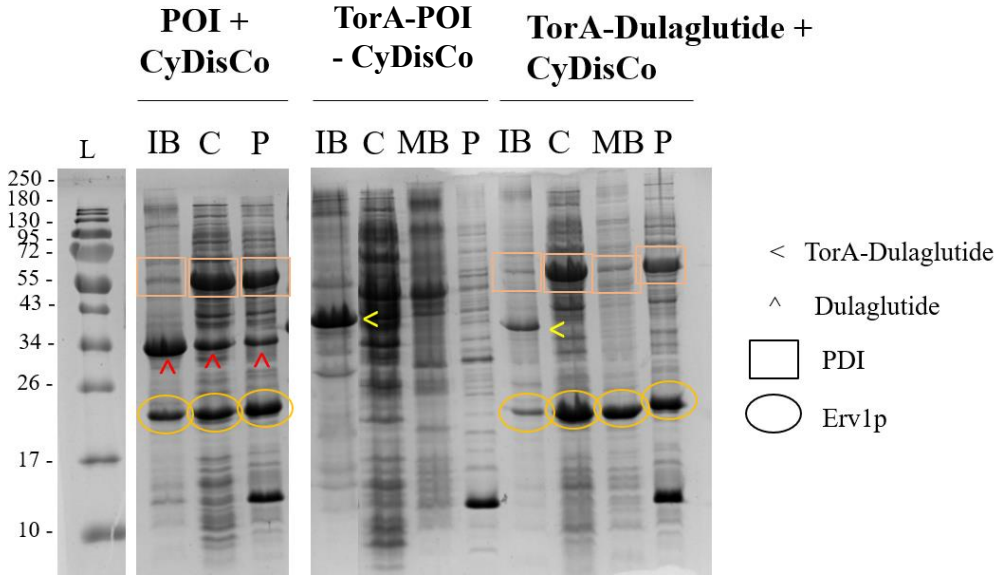
A


ptac
RBS
POI
T
pEXT22 (low copy number)
RBS
POI
RBS
PDI
RBS
Erv1p
T
pEXT22 + CyDisCo
RBS
POI
T
pET23 (high copy number)
RBS
PDI
RBS
Erv1p
T
pLysS (medium copy number)

B


POI + CyDisCo
TorA-POI - CyDisCo
TorA-Brazzein + CyDisCo
L
IB C MB P IB C MB P IB C MB P
40 -
25 -
15 -
Brazzein
10 -

C


POI + CyDisCo
TorA-POI - CyDisCo
TorA-Dulaglutide + CyDisCo
L
IB C P
IB C MB P IB C MB P
250 -
180 -
130 -
95 -
72 -
55 -
43 -
34 -
26 -
17 -
10 -
< TorA-Dulaglutide
^ Dulaglutide
PDI
Erv1p

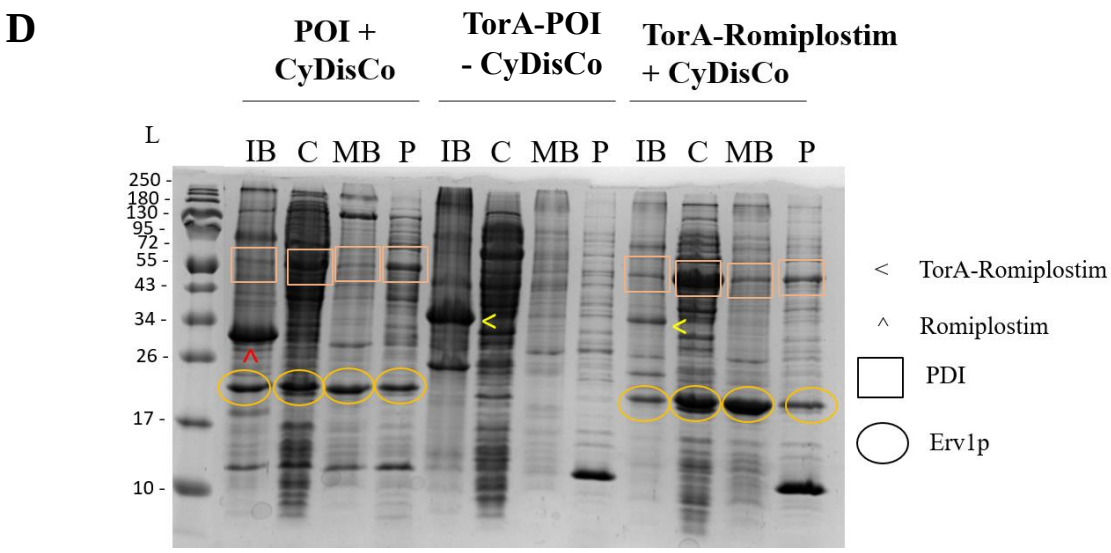


Figure 11. Expression of TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim in a pEXT22 plasmid in the presence of CyDisCo in rich AIM at 30˚C in two-plasmid system. A: Schematic representation of the plasmids used in this study (Chapter 2). tac promoter (p*tac*) is shown as an orange arrow, ribosome binding site (RBS) is shown in blue, protein of interest (POI) in green and terminator (T) in grey. CyDisCo components, sulfhydryl oxidase (Erv1p) and disulphide isomerase (PDI) are shown in brown and red respectively. N-terminal His-tagged Brazzein, Dulaglutide and Romiplostim (POI + CyDisCo, positive control) in a two-plasmid system with CyDisCo; C-terminal His-tagged TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim were expressed in absence of CyDisCo in pEXT22 one plasmid system (TorA-POI – CyDisCo, negative control); C-terminal His-tagged TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim were expressed in presence of CyDisCo in pEXT22 two-plasmid system in BL21 strain in 100 mL shake flask at 250 rpm with a sponge as a O2 permeable membrane. After 22-24 h of culture, cells were fractionated into, inclusion bodies (IB), cytoplasm (C), membrane (MB) and periplasm fraction (P) using EDTA/ lysozyme/ cold osmotic shock for periplasm extraction. **B:** Coomassie-stained gel of TorA-Brazzein experiment highlighting by an arrow the POI without the SP in the cytoplasmic fraction of the positive control (POI + CyDisCo). **C:** Coomassie-stained gel of TorA-Dulaglutide experiment. **D:** Coomassie-stained gel of TorA-Romiplostim experiment. CyDisCo components, PDI and Erv1p, are marked by a square and a circle, respectively. "<" marks the POI with TorA SP attached and "^" the POI without the SP. Molecular weight ladder (L) is shown on the left in kDa. All the cells' fractions have been diluted in the same volume. The reproducibility of the experiments has not been ascertained.

In Figure 11B, TorA-Dulaglutide export assessment in this newly implemented two-plasmid system was carried out. The positive control (in the figure as "POI + CyDisCo"), shows the fractionation of His-Dulaglutide in the two plasmid-system in the presence of CyDisCo. As for Brazzein, our collaborators at the University of Oulu reported that Dulaglutide is 70%-90% soluble in the cytoplasm of the positive control. In our experiment, we can visually determine that

we could only produce around a half of the total production in a soluble state (divided between cytoplasm where it ideally should be located-, and periplasm due to cross-contamination of fractions). The rest of the produced protein stays as inclusion bodies and in the membrane. For the first time, Erv1p and PDI were easily identifiable in the SDS-PAGE (mainly in the cytoplasm, but again, due to cross-contamination of the fractions, they are also visible in other fractions). The negative control (in the figure as "TorA-POI – CyDisCo") shows the fractionation of TorA-Dulaglutide in the pEXT22 plasmid in the absence of CyDisCo. In this case, and as we expected, the POI can be identified mostly as inclusion bodies (IB) in the cell. Finally, when two-plasmid system was tried for the export by Tat pathway of TorA-Dulaglutide (CyDisCo components in one plasmid + pEXT22 with the POI, in the Figure 11B as "TorA-Dulaglutide + CyDisCo"), TorA-Dulaglutide is identified in inclusion bodies. In the positive control, CyDisCo components can also be identified. The periplasmic fractions in both "POI + CyDisCo" and "TorA-Dulaglutide + CyDisCo" were contaminated with cytoplasmic fraction, as it is deduced by the similarity of the two fractions. This is probably due to an inefficient practice of the periplasmic extraction protocol or overexpression, and consequent cell stress and weakness, of CyDisCo and the POI in the two-plasmid system.

In Figure 11C TorA-Romiplostim export assessment in this newly implemented two-plasmid system was carried out. The positive control (in the figure as "POI + CyDisCo"), shows the fractionation of His-Romiplostim in the two plasmid-system in the presence of CyDisCo. As for Brazzein and Dulaglutide, our collaborators at the University of Oulu reported that Dulaglutide is between 70% to 90% soluble in the cytoplasm of the positive control. Our results show romiplostim in inclusion bodies only with no or very low production of this protein in the cytoplasm. As before, Erv1p and PDI are easily identifiable in the SDS-PAGE (mainly in the cytoplasm, but also in the rest of the cell fractions due to cross-contamination of the fractions). The negative control (in the figure as "TorA-POI - CyDisCo") shows the fractionation of TorA-Romiplostim in pEXT22 one plasmid system in the absence of CyDisCo. In this case, and as we hypothesized, the POI can be identified mostly as IBs in the cell. Finally, when two-plasmid system was tried for the export by Tat pathway of TorA-Romiplostim (CyDisCo components in

one plasmid + pEXT22 with the POI, in the Figure 11C as "TorA-Romiplostim + CyDisCo"), TorA-Romiplostim is identified in a low amount in inclusion bodies. Also, as in the positive control, CyDisCo components are identified. Unfortunately, periplasmic fractions in both positive control and TorA-Romiplostim + CyDisCo were contaminated with cytoplasmic fraction, as it is deduced by the similarity of the two fractions. Once again, we hypothesise that this is probably due to an inefficient practice of the periplasmic extraction protocol or overexpression of CyDisCo and the POI in the two-plasmid system. The overexpression of these proteins might be causing cell stress and consequently, cells become weak for fractionation, giving as a result a cross-contaminated cell fractions.

Even though the ORI of the CyDisCo plasmid and our low copy number plasmid (pEXT22) are different, therefore compatible, the export to the periplasm of the three POIs was unsuccessful. For the first time, PDI and Erv1p were visible in the Coomassie gels, although not fully functional (see Figure 11C as an example). CyDisCo plasmid might be depleting the resources of the cells, as this is a low to medium copy number plasmid in comparison with the pEXT22 where the POIs have been expressed. If this is the case, the POI will barely be produced, and the yields will always be very low.

To overcome this potential issue, and to maintain a high level of CyDisCo expression as the same time as a high production of the target POI, the complete University of Oulu protein expression protocol was adopted. For this, TorA-POIs were cloned in pET23 previously used to produce Brazzein, Dulaglutide and Romiplostim without a SP in Figure 11 and were co-expressed with CyDisCo plasmid. Unlike in previous experiment, the BL21 TatExpress (TE) cells were used. TE cells, over-expresses the *tatABC* genes (encoding the Tat system) from the chromosome and will potentially avoid the saturation of the pathway when producing a large amount of protein to be exported (Browning et al., 2017).

The co-expression was setup by co-transformation of the recombinant plasmids into expression strain *E. coli* BL21 TE. The cells were pre-cultured in LB media supplemented with 2 g/L of glucose and corresponding antibiotics for 6 hours and transferred to rich AIM for cell expression. After 22-24 h in this media at 30˚C at 250 rpm, the cells were fractionated in inclusion bodies

(IB), cytoplasm (C), periplasm (P) and membrane fraction (MB). This time, TorA-hGH was also expressed as a positive control to ensure that BL21 TE was growing and expressing in this media as expected. *E. coli* TatExpress cells have been shown to export high levels of hGH but while this protein contains two disulphide bonds, they are not required for proper folding and the Tat system can efficiently export the protein in its reduced state without CyDisCo system. For unknown reasons, the TorA SP is prematurely cleaved in the cytoplasm, which prevents Tat from exporting all the protein produced to the periplasm (Guerrero Montero et al., 2019). Overall, the hGH was found to be exported in high levels to the periplasm by TE cells in autoinduction media (Figure 12D).

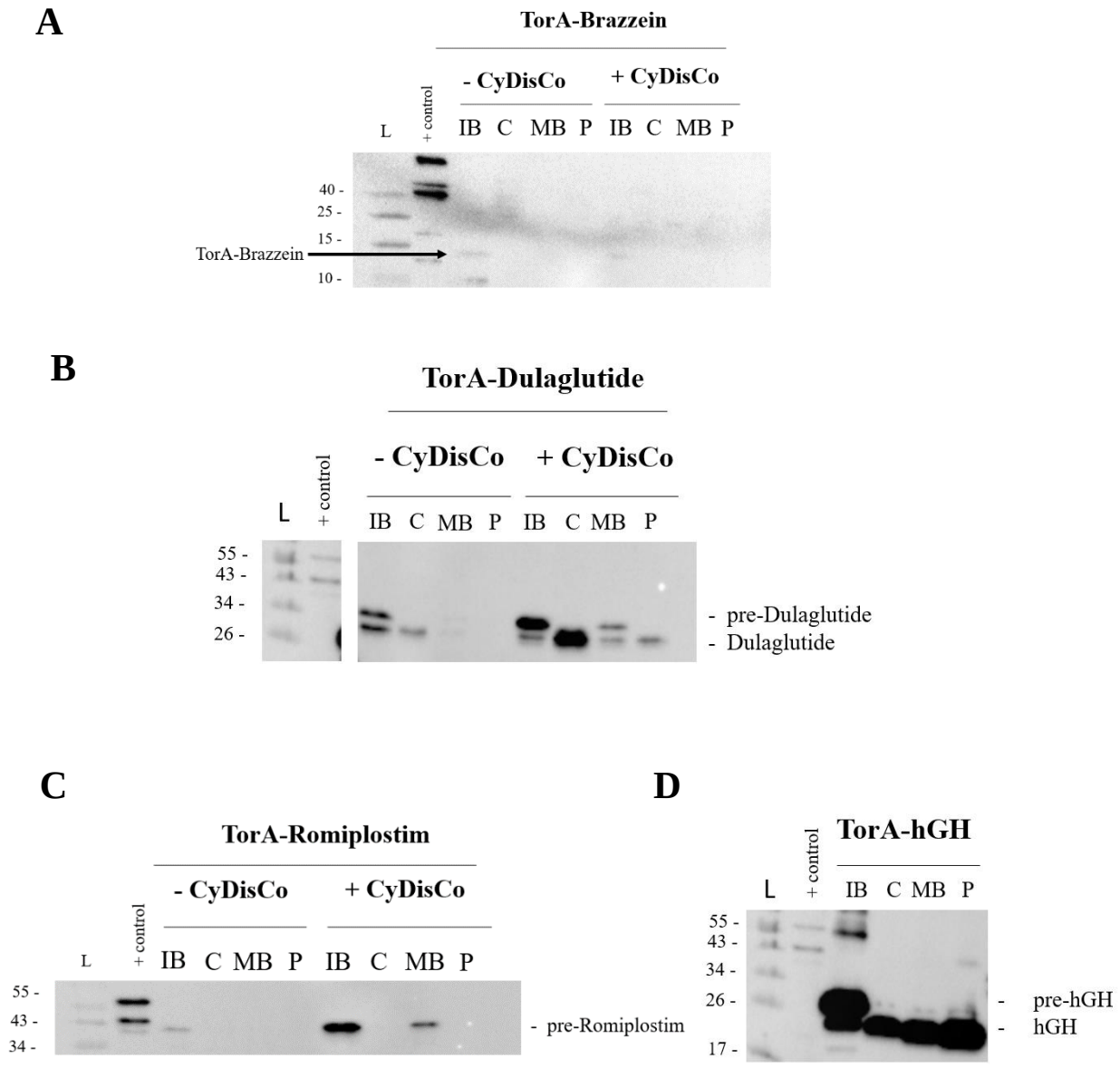


Figure 12. Expression of TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim in a high copy number plasmid (pET23) in the presence and absence of CyDisCo in rich AIM at 30˚C in two-plasmid system in BL21 TatExpress cells. TorA-Brazzein, TorA-Dulaglutide and TorA-Romiplostim were expressed in the presence (+) and absence (-) of CyDisCo in BL21 TE strain in 100 mL shake flask at 250 rpm with a sponge as a O_2 permeable membrane. After 22-24 h of culture, cells were fractionated into, inclusion bodies (IB), cytoplasm (C), membrane (MB) and periplasm fraction (P) using EDTA/ lysozyme/ cold osmotic shock for periplasm extraction. TorA-hGH was also expressed in the same conditions as a positive control of the media for TE cells. **A:** WB for fractionation of cells expressing TorA-Brazzein in the absence/presence of CyDisCo. **B:** WB for fractionation of cells expressing TorA-Dulaglutide in the absence and presence of CyDisCo. **C:** WB for fractionation of cells expressing TorA-Romiplostim in the absence and presence of CyDisCo. **D:** WB for fractionation of cells expressing TorA-hGH in a polycistronic low copy number plasmid. In all cases samples were immunoblotted to the C-terminal 6x histidine-tag. Molecular weight ladder (L) is shown on the left in kDa and WB positive control is labelled as "+ control". All the cells' fractions have been diluted in the same volume. The reproducibility of the experiments has not been ascertained.

Figure 12A shows a WB of the fractionation of TorA-Brazzein in this newly adopted two-plasmid system in the absence and presence of CyDisCo in the conditions described above. As we saw previously in the expression of this POI in a pEXT22 plasmid, the TorA SP does not appear to support Brazzein export, and a very low amount of protein could be visualised on the WB in the IB fraction in the absence and presence of CyDisCo when using this two-plasmid system.

Figure 12B shows a WB of the fractionation of TorA-Dulaglutide also in a this newly adopted two-plasmid system in the absence and presence of CyDisCo. TorA-Dulaglutide, that appeared to be a promising Tat apparatus substrate, underperformed in these conditions. In the absence of CyDisCo, a duplet band corresponding to the uncleaved POI (pre-Dulaglutide) and to the mature protein (Dulaglutide, without the SP) were visible in the IB fraction of the cell. Also, in the absence of CyDisCo, some mature protein was detected in the C fraction of the cell. In the presence of CyDisCo, a duplet band corresponding to the uncleaved POI (pre-Dulaglutide) and to the mature protein (Dulaglutide, without the SP) were visible in the IB and MB fraction of the cell. In the cytoplasmic fraction (C), a very high amount of mature Dulaglutide can be detected. In the periplasmic (P) fraction some mature protein can be detected, most probably due to a contamination with the C fraction. Ultimately, TorA-Dulaglutide was not exported or was very inefficiently exported to the periplasm by Tat pathway in the presence and absence of CyDisCo when using the original two-plasmid system. Although Dulaglutide is highly soluble in this system in the presence of CyDisCo, for unknown reasons the TorA SP is prematurely cleaved in the cytoplasm, which may prevent Tat from exporting the protein to the periplasm.

TorA-Romiplostim is only detected when co-expressed with CyDisCo in the form of IBs. Some protein can also be identified in the MB fraction, presumably on its way to the periplasm in the presence of CyDisCo. In conclusion, TorA-Romiplostim cannot be exported by Tat in the presence or absence of CyDisCo (Figure 12C).

2.4. Discussion

This study set out to analyse the capacity of Tat pathway to export disulphide-bonded proteins that have been folded with the help of the CyDisCo system. For that, three disulphide-bonded proteins were tested: Brazzein, Dulaglutide and Romiplostim. The chosen POIs are a very high-value proteins and warrant the development of new, more efficient platform to reduce costs. Therefore, we thought that the export of these POIs to the periplasm by Tat pathway in *E. coli* can be beneficial. *E. coli* is known for its rapid growth, high product yield, cost effective production and easy scale-up processes. At the same time, periplasm targeting facilitates protein isolation from the relatively small periplasmic proteome, the control of the nature of the N-terminus of the mature protein and minimizes the exposure to cytoplasmic proteases. When choosing the Tat pathway as a system to export the protein to the periplasm, its proofreading capability could allow for a more homogeneous product to be produced in the periplasm (Fisher et al., 2008; Kleiner-Grote et al., 2018). In addition, the recent development of TatExpress strains boosts the industrial potential of this pathway. Moreover, these POIs are between 70% to 90% solubly produced by CyDisCo (data not published, information provided by University of Oulu), which makes them a very good target for the export by Tat, as Tat can only transport correctly folded proteins.

When the export of these three POIs was initially assessed, they were expressed in a polycistronic low-copy number plasmid (pEXT22) in LB media at 30˚C in the presence and absence of CyDisCo, following previous protocols where protein export was obtained by Tat pathway (Guerrero Montero et al., 2019). Low copy number plasmid (pEXT22) would be expected to prevent the Tat pathway from being overloaded (Sutherland et al., 2018). Unfortunately, none of the proteins were successfully exported to the periplasm after 3 h post-induction. TorA-Brazzein was not expressed at all, or it was degraded immediately. TorA SP even though it is a very well-studied Tat specific SP, has been reported to induce degradation of the protein of interest, for unknown reasons (Alanen et al., 2015). TorA-Dulaglutide and TorA-Romiplostim were expressed insolubly in these conditions. We also reproduced the same experiment, lowering expression temperatures to facilitate the production of folded, soluble protein (Francis & Page,

2010), and therefore the translocation of the protein from the cytoplasm to the periplasm. In the 25 ˚C expression experiments, samples were taken 3 hours post induction, and after o/n expression in order to monitor the production of the POI at different time points. Overall, lowering the temperature did not help in the export of these disulphide-bonded proteins by Tat pathway and an o/n sample did not show any difference for our purpose to the earliest time-point sample. A follow-up experiment switching to rich AIM was meant to provide desirable conditions for the export of our POIs as it permits a more similar bioreactor-like conditions and aeration is not stopped at any time in the incubators (Rosano et al., 2019), assuring that Erv1p gets the oxygen needed for its optimal performance. Also, we expected that a longer culture time would permit CyDisCo components to be active and catalyse the formation of DSBs of our POIs. At the same time, the shaking speed of our cultures was increased and the culture volume was reduced to ensure efficient aeration. Unfortunately, AIM in these conditions did not help in the export of our POIs by Tat pathway, and they were insolubly produced in BL21. Positively, TorA-Brazzein was detected in the presence of CyDisCo, nurturing the idea of AIM superiority for these experiments, thus suggesting that TorA-Brazzein can be expressed and that TorA SP might be targeting for degradation the protein in previous experiments.

When transiting to the two-plasmid system, we also realized that CyDisCo performance/production was not the only issue for the export of the proposed POIs by the Tat pathway. Dulaglutide and Romiplostim without a SP, in our hands, were only half soluble (70%-90% solubility has been achieved in similar conditions), strongly suggesting that AirOtop as an O_2 transferable membrane (not used in our experiments but used by our collaborators) plays a very big role in the production of these. Although the amount of CyDisCo components expressed should be enough to get some export to the periplasm especially in the case of Brazzein, the simpler and smaller protein tried in this research. Brazzein was almost undetectable when fused to the TorA-SP, TorA-Romiplostim was insoluble and the SP in TorA-Dulaglutide was found to be cleaved before the export of the protein to the periplasm by Tat pathway. These observations reinforce the hypothesis that the TorA SP could be one of the main obstacles for efficient export of these POIs.

Thanks to this study, we found very important bottlenecks to consider when employing CyDisCo for the export of recombinant proteins via the Tat pathway. Even though we did not successfully export any of these three proteins to the periplasm, we optimized a protocol for the production of high-value proteins that has been successfully employed and proved in Chapter 4. Our findings suggest that CyDisCo must be employed in a two-plasmid system (CyDisCo components in one plasmid and POI in a high copy number plasmid) when the proteins require a lot of Erv1p and PDI for the correct formation of their DSBs. Also, we discovered that the production (or detection) of CyDisCo components in the Coomassie gel does not always assure the correctly operation of CyDisCo when folding the POIs: aeration seems to be crucial for the correct function of CyDisCo (Gąciarz et al., 2017). Using O_2 transferable membrane like AirOtop, less media volume in the flasks and high incubators orbital shaking will provide a better CyDisCo performance. At the same time, LB does not seem to be the ideal media for the production of these proteins. Rich autoinduction media instead, provides time enough for CyDisCo to be translated and function. In addition, stopping the culture for IPTG addition can be avoided as it might be affecting the aeration and dissolved oxygen levels of the culture medium.

As expected, this two-plasmid system also poses some drawbacks. When using this system, too much POI is produced and can potentially block Tat system, preventing an efficient export of the POI to the periplasm (Sutherland et al., 2018), even when using TE cells. Moreover, cells are stressed due to the overproduction of the POI and tend to lyse while fractionating, resulting in cross-contamination of cell fractions. To overcome this overproduction of the POI while still using this promising two-plasmid system, point mutations in tac promoter is proposed (Browning et al., 2019). These mutations could be a solution to lower the production of the POI in order to avoid the saturation of Tat system and reduce the stress on the cells, but still assuring good yields. Furthermore, when using the two-plasmid system it gets more and more important to have a good periplasmic extraction protocol. When utilizing one-plasmid system, the periplasmic fraction seemed to be comparatively less contaminated with cytoplasmic proteins than when using two-plasmid system as this system can stress the cells and make them more delicate and prone to lysis while fractionating. Therefore, a gentle and efficient periplasmic extraction protocol was

developed by Malherbe and colleagues (2019), termed as PureFrac and will be applied for the fractionation of the cells in the following experiments.

To conclude, we have laid the groundwork for the employment of CyDisCo in Tat pathway export. While the findings presented in this chapter are fundamental for the future disulphide-bonded protein expression, a further understanding of the limitations of Tat as a fully capable protein export system is essential. The chosen POI were very interesting targets, but complicated ones. To the best of our knowledge, Tat performance for export of proteins with more than four DSBs has not been assessed. Therefore, the expression of a known Tat-dependent biopharmaceutical could be the ideal model protein to find what is holding Tat back to be the preferable pathway to export difficult-to-express proteins.

CHAPTER 3: Equal first author publication. Yields and product comparison between *Escherichia coli* BL21 and W3110 in industrially relevant conditions: anti-c-Met scFv as a case study

3.1. Contribution

For this project, I contributed to concept design, worked at Boehringer Ingelheim together with Luisa Buscajoni to carry out the work, analysed the data, reviewed the literature available on differences between B and K-12 strains in yields, growth/metabolism behaviour and product differences, made the figures and wrote the manuscript in collaboration with Luisa Buscajoni. The manuscript was revised by all the authors.

In this project, we carried out a comparative study of the two most widely used *E. coli* strains in the biopharmaceutical industry: BL21 and W3110. We compared and analysed the yield differences and the product structure and heterogeneity by a wide variety of techniques. All the experiments for this manuscript were performed at Boehringer-Ingelheim (Austria) at Process Science Downstream Development and Process Science Upstream Development laboratories, except for NMR experiments that were performed at University of Kent (UK).

The original research project workflow aiming to compare Tat and Sec based periplasmic export of scFvM was designed by Luisa Buscajoni and me with the help of our supervisors. I carried out sample analysis for soluble and insoluble protein quantification by SDS-PAGE of the POI when exported by Tat and Sec. I also fractionated 5 L fermenter samples to verify the export to the periplasm and carried out the data evaluation. The following manuscript idea was also conceived by us. Luisa Buscajoni's major input was the downstream section, where I also took part in purifying and dialysing the product (under her guidance) as well as analysing and correlating the data. On the other hand, my major input was in the upstream section: designing experimental culture conditions in multifermenters and 5 L bioreactors, preparation of samples obtained from

fermentation for SDS-PAGE analysis and quantification, software data analysis, problem solving and data interpretation.

Finally, mass spectrometry data was obtained by Laura Niederstaetter; Nicole Weiner performed the DSC technique and Martin Voigtmann group performed the immunoassays (all of them acknowledged at the end of the manuscript). NMR data was collected by Gary Thompson and Karin Koch guided us and provided her help on the scientific steps of this collaboration. The sample preparation and interpretation of the data obtained by the above-mentioned techniques was carried out by Luisa Buscajoni and me.

3.2. Preface

Many high-value proteins are produced in *Escherichia coli*, and a favoured strategy is to export the protein to the periplasm, usually by the well-characterized "Sec" pathway (Rosano & Ceccarelli, 2014). However, the Sec pathway cannot transport some heterologous proteins (Guerrero Montero et al., 2019). Alternatively, the Tat pathway transports fully folded proteins by a completely different pathway (reviewed by Natale et al., 2008). Previous studies showed that Tat was capable of exporting 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 could export several biotherapeutics, including human growth hormone (hGH) in fermenters (Guerrero-Montero et al., 2019), single chain antibody fragments, and interferons at shake-flask level (DeLisa et al., 2003; Tullman-Ercek et al., 2007; Matos et al., 2014; Alanen et al., 2015; Browning et al., 2017). Even though Tat pathway seems a very promising system to export high-value proteins, Sec is still the preferred pathway and Tat has rarely been used for high scale production of biopharmaceuticals. The novelty of comparing Sec and Tat pathway when producing the same biopharmaceutical in bioreactors and characterization of the product obtained can elucidate the drawbacks of Tat and give a real insight into the differences in product yields and quality between these two pathways.

Jones et al. (2016) showed that the Tat pathway could export an scFv (named as scFvM in their paper and in this chapter) in *E. coli* BL21 (DE3) when the TorA SP was attached. While a different scFv has previously been shown to be efficiently exported by Tat in both the presence and absence of CyDisCo components (Alanen et al., 2015), they found the export of scFvM to be almost totally dependent on the presence of CyDisCo. This scFv (technically known as anti-c-Met scFv) is a two disulphide-bonded protein expected to be useful in the clinical treatment or imaging of many cancer cells. c-Met is a potential target in the development of therapeutic reagents of cancer as well as in targeted therapies because of its biological roles in the proliferation of malignant tumour cells (Qamsari et al., 2017).

Overall, scFvM seems to be a very interesting POI to understand the potential of the Tat pathway in industrially relevant conditions. This is a simpler protein (two DSBs) in comparison to the ones in Chapter 2, and its efficient Tat export has already been reported at the shake-flask scale. In addition, it is a CyDisCo dependent and industrially relevant protein, so scale-up for its abundant production is necessary. Furthermore, comparison with the periplasmic export of the same POI through the Sec pathway allows a very realistic comparison of what each pathway can provide to the biopharmaceutical industry and, to our knowledge, this has never been done before. The initial aim of this study was to compare Sec and Tat pathway based periplasmic export of an scFv in two widely employed industrial *E. coli* strains, namely BL21 (B strain) and W3110 (K-12 strain) in 5 L fermenters. At the same time, expression in a miniaturised fermentation platform of 10 mL (multifermenters, MF) was carried out to further understand Tat and Sec optimization and scalability.

Tat underperformed and less than 0.1 g/L of scFvM was obtained in 5 L fermenters at the end of fermentation (EoF) in both BL21 and W3110, while Sec exported 2.41 g/L and 1.01 g/L respectively (all the conditions and methodology has been explained in Materials and Methods section 3.3.). This means that the Sec pathway exports around 25 times and 10 times more scFvM than the Tat pathway in *E. coli* BL21 and W3110 respectively (Figure 13), even though the end of fermentation (EoF) OD_{550} is higher in experiments employing the Tat pathway in the B strain (Sec pathway BL21: 212; Tat pathway BL21: 247). This underperformance of Tat pathway made

the comparison of the two pathways futile and the comparison of product quality impossible. However, BL21 and W3110 performed very differently when the Sec pathway was studied as the means of export to the periplasm of the scFvM, providing interesting data and an opportunity to understand novel aspects of industrially-relevant production (manuscript in section 3.3.).

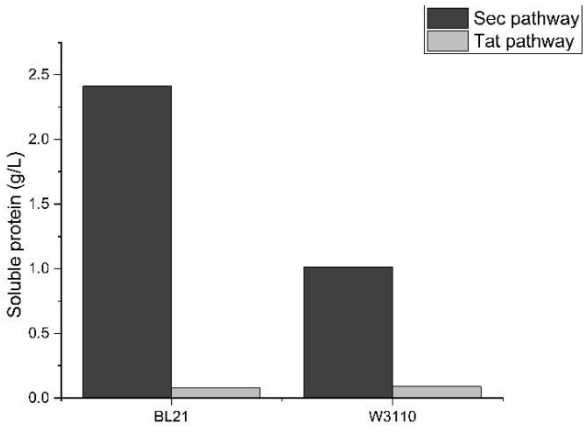


Figure 13. Soluble scFvM production in BL21 and W3110 in 5 L fermentations at the end of fermentation (10-hours post-induction). Sec pathway is represented in black colour and Tat pathway in grey. Soluble scFvM product was determined by automated immunoassay (Gyros) from suspension samples (see Materials and Methods in section 3.3.). Representative data from duplicates is shown.

The Tat pathway seems to underperform more considerably in 5 L fermenters than in 10 mL fermenters. To eliminate a possible impact of the different OD levels at both scales, the specific soluble product titer at 7 hours post-induction (end of fermentation point in MF) was divided by the OD_{550} at that specific time for both strains and scales (Figure 14). BL21 strain shows a specific soluble product of 5.01 mg/OD in 10 mL fermenters and 0.43 mg/OD in benchmark 5 L fermenters. The W3110 strain, on the other hand, shows a specific soluble product of 1.82 mg/OD in 10 mL fermenters and 0.43 mg/OD in 5 L fermenters. Janzen et al. (2019) describes a slightly higher specific soluble product in small- scale cultivations than in 5 L fermenters. In this case, we hypothesize a scalability issue of Tat when scaling up from 10 mL fermenters to 5 L fermenters.

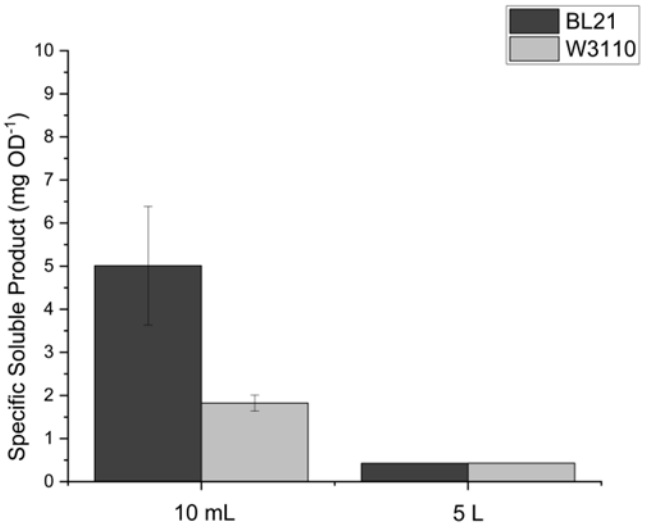


Figure 14. Specific soluble product formation of scFvM in BL21 and W3110 in 10 mL and 5 L fermentations. Specific soluble product is expressed in mg OD^{-1}. To calculate these values, the 7 h post-induction soluble scFvM titer from suspension samples is divided by the OD_{550} of the culture at that time in both scale bioreactors and analysed by automated immunoassay (Gyros). MFs specific soluble product formation values are the average of four times replicates; error bars have been added to the figure. BL21 strain is represented in black colour and W3110 in grey. All the Materials and Methods employed for the MF and 5 L fermenters runs can be found in 3.3. section.

Despite these unanticipated results, we further analysed these results with the aim of understanding why Tat failed in exporting high yields of this already studied Tat substrate. A loss in the plasmid copy number (PCN) was suggested, but when assessing so, this possibility was discarded. The PCN was found to be stable throughout the fermentations in both BL21 and W3110 (PCN: 7-8 copies/cell) (refer to Material and Methods in Section 3.3. "Plasmid copy number (PCN) estimation"). Also, the plasmid vector used (pYU49-ta-scFv-M-c6His-CyDisCo from Jones et al., 2016) was sequenced and no mutations were found. CyDisCo poor performance was also considered, however, CyDisCo has been found to work in a variety of media (Jones et al., 2016; Gąciarz et al., 2017; Sohail et al., 2020) and the dissolved oxygen (DO) level was always maintained to ≥35%, ensuring the correct folding and function of CyDisCo components. Tat machinery could not be overwhelmed, as it is able to export more than 1 g/L of POI (Guerrero-Montero et al., 2019), and no uncleaved protein could be detected in the insoluble fraction of the cell. Therefore, we suggest that TorA SP might be causing the degradation of the POI. In the

previous chapter we strongly suspected that TorA SP was degrading some of the POIs and it seems that this scFv is also one of them. This degradation is enhanced when scaling-up to 5 L fermenters.

This last statement reinforces the idea that TorA SP is most probably one of the main reasons why Tat is not an industrially relevant pathway to export the POIs in biopharmaceutical industry. TorA SP is potentially causing the degradation of many target POIs, therefore low yields can be expected. This idea opens Chapter 4, that tries to find an alternative to this Tat specific SP in order to maximize the industrial relevance of the Tat pathway.

As explained before, BL21 and W3110 strain performance and product differences when using Sec pathway permitted a further analysis of all the data obtained of these fermentations and is presented as a manuscript in 3.3. section. In this section also, all the protocols carried out for small scale and 5 L scale bioreactor cultivations, as well as analytical methods and PCN estimation methodology are described.

3.3. Publication

The paper is in preparation to be submitted to Microbial Cell Factories as follows:

Yields and product comparison between *Escherichia coli* BL21 and W3110 in industrially relevant conditions: anti-c-Met scFv as a case study

Klaudia Arauzo-Aguilera[1]*, Luisa Buscajoni[2]*, Karin Koch[3], Gary Thompson[4], Colin Robinson[1], Matthias Berkemeyer[2]

[1]School of Biosciences, University of Kent, Canterbury CT2 7NJ, United Kingdom

[2]Boehringer-Ingelheim RCV GmbH & Co KG, Biopharma Austria, Process Science Downstream Development, Dr. Boehringer-Gasse 5- 11, 1120 Vienna, Austria

[3]Boehringer-Ingelheim RCV GmbH & Co KG, Biopharma Austria, Process Science Upstream Development, Dr. Boehringer-Gasse 5- 11, 1120 Vienna, Austria

[4]Wellcome Biological NMR Facility, School of Biosciences, University of Kent, Canterbury CT2 7NJ, United Kingdom

k.arauzo@kent.ac.uk (KAA)

luisa.buscajoni@boehringer-ingelheim.com (LB)

karin.koch@boehringer-ingelheim.com (KK)

g.s.thompson@kent.ac.uk (GT)

c.robinson-504@kent.ac.uk (CR)

matthias.berkemeyer@boehringer-ingelheim.com (MB)

* Luisa Buscajoni and Klaudia Arauzo-Aguilera contributed equally to this work

Corresponding Author:

Matthias Berkemeyer, matthias.berkemeyer@boehringer-ingelheim.com, tel. +43 (1) 80105-5043

Abstract

Introduction: In the biopharmaceutical industry, Escherichia coli is one of the preferred expression hosts for large-scale production of therapeutic proteins. Although increasing the product yield is important, product quality is a major factor in this industry because greatest productivity does not always correspond with the highest quality of the produced protein. While some post-translational modifications, such as disulphide bonds, are required to achieve the biologically active conformation, others may have a negative impact on the product's activity, effectiveness, and/or safety. Therefore, they are classified as product impurities, and they represent a crucial quality parameter for regulatory authorities. In this study, we aim to compare the yields and product quality of a single-chain variable fragment (scFv) in fermentation conditions of two widely employed industrial E. coli strains, BL21 and W3110.

Results: In this study, fermentation conditions of two widely employed industrial E. coli strains, BL21 and W3110 are compared for recombinant protein production of a single-chain variable fragment (scFv) in an industrial setting. We found that the BL21 strain produces more soluble scFv than the W3110 strain, even though W3110 produces more recombinant protein in total. A quality assessment on the scFv recovered from the supernatant was then performed. Unexpectedly, even when our scFv is correctly disulphide bonded and cleaved from its signal peptide in both strains, the protein shows charge heterogeneity with up to seven distinguishable variants on cation exchange chromatography. Biophysical characterization confirmed the presence of altered conformations of the two main charged variants.

Conclusions: The findings indicated that BL21 is more productive for this specific scFv than W3110. When assessing product quality, a distinctive profile of the protein was found which was independent of the E. coli strain. This suggests that alterations are present in the recovered product although the exact nature of them could not be determined. This similarity between the two strains' generated products also serves as a sign of their interchangeability. This study encourages the development of innovative, fast, and inexpensive techniques for the detection of heterogeneity

while also provoking a debate about whether intact mass spectrometry-based analysis of the protein of interest is sufficient to detect heterogeneity in a product.

Keywords: Sec pathway, fermentation, *Escherichia coli* BL21 and W3110, disulphide bond, product heterogeneity, protein purification

Introduction

Escherichia coli is one of the expression hosts of choice in the biopharmaceutical industry for large-scale production of therapeutic proteins because of its rapid growth, high product yield, cost effective production and easy scale-up processes (Rettenbacher et al. 2021). If the protein of interest (POI) contains disulphide bonds (DSBs), as in the case of antibody fragments, periplasmic expression *via* the Sec pathway is often preferred over cytoplasm expression (Sandomenico et al. 2020). Thereby the POI is transported in an unfolded state to the periplasm by fusing a signal peptide (SP) to the N-terminus of the POI. Once in the periplasm, correct DSB formation is achieved (Kipriyanov 2009). Furthermore, product translocation into the medium can be enforced which also simplifies downstream processing (Zhang et al. 2006).

Traditional upstream bioprocess development involves the use of shaken bioreactor systems (usually shake flask). Cultivation in shake flasks is normally performed in a batch manner, provides very limited variable monitoring, and produces low cell densities and product yields. Furthermore, they rely on uncontrolled surface aeration leading to limited oxygen transfer rates and low batch-to-batch reproducibility (Ali et al. 2012). Therefore, cultivation conditions that are used during shaken culture bioprocess development may be changed or completely discarded once they are optimized at pilot scale (Panula-Perälä et al. 2008). To overcome the limitations described above, there has been a concerted effort to develop fully automated high-throughput cultivation systems to significantly accelerate the identification of the optimal expression systems and process conditions (Baumann and Hubbuch 2017; Janzen et al. 2019). A good understanding of fermentation parameters and their impact on *E. coli* cell growth and final product yield is critical in defining biopharmaceutical production processes. Until today, many process

adaptations to maximise product yield are based on optimizations of temperature, dissolved oxygen (DO) levels, pH, media composition, etc. (Tripathi and Shrivastava 2019). However, maximum productivity does not always coincide with the highest quality of the recombinantly expressed protein (Huleani et al. 2021).

Correct folding and *in vivo* stability of the recombinant protein are two crucial factors that must be controlled while optimising the cultivation conditions. Correct folding includes both the acquisition of the correct 3D structure as well as addition of post-translational modifications (PTMs), such as DSB formation. In *E. coli,* PTMs that occur during and after protein synthesis can represent a limitation when compared to other microorganisms. In the past decades several approaches have been explored in *E. coli* to overcome this drawback, reviewed by Rettenbacher et al. (Rettenbacher et al. 2021).

Some PTMs and physiochemical transformations of recombinant proteins can also originate from non-enzymatic reactions at all steps of the production process from cell culture to purification and storage (Beyer et al. 2018). In this case, PTMs are caused by chemical reactions occurring between the amino acid side chain and reagents present in either culture media or buffers in specific conditions of pH, temperature and oxygenation level. Some of these modifications can negatively affect the activity, efficacy and safety of the desired product by altering the product stability and its biological active conformation. Therefore, the percentage of product harbouring these modifications, within the heterogenous product pool generated, is identified as product related impurities and represents a crucial quality parameter for regulatory authorities (Rudge and Nims 2017). Common PTMs are methionine oxidation, asparagine and glutamine deamidation, and aspartate isomerization. The importance of such unwanted modifications has been evaluated and ranked for recombinant monoclonal antibodies (mAbs) (Liu et al. 2014). While N-terminal pyroglutamate, for example, is not considered as a critical quality attribute, modifications occurring in the complementary determining regions are of high importance as they could affect the antigen recognition capacity (Liu et al. 2014).

To investigate product heterogeneity, ion exchange chromatography (Moorhouse et al. 1997; Lee et al. 2018; Nascimento et al. 2018) coupled with enzymatic digestion followed by mass

spectrometry (MS) analysis (peptide mapping) (Khawli et al. 2010; Singh et al. 2021) are often the methods of choice. The former allows the separation of the protein heterogeneity based on the charge properties while the latter allows the identification of mass changes and the exact position of the modification within the protein expressed. However, since peptide mapping requires extensive work and can also generate artefactual modifications, the research for the improvement of this method is ongoing (Gaza-Bulseco et al. 2008; Ren et al. 2009).

In this study, fermentation conditions of two widely employed industrial E. coli strains, namely BL21 (B strain) and W3110 (K-12 strain) are compared for recombinant protein production of a single-chain variable fragment (scFv) in an industrial setting. Rather than analysing the well-known and studied performance and behavioural differences (Shiloach et al. 1996; Noronha et al. 2000; Shiloach and Rinas 2009; Marisch et al. 2013), we focused on yields, differences in product structure and heterogeneity between strains grown in 5 L fed-batch bioreactors using a number of different downstream and analytical techniques. The results reveal surprising difference in protein quantity and quality between the two strains, and equally surprising heterogeneity in the final preparations of this relatively simple biopharmaceutical.

Results and discussion

BL21 strain produces more soluble scFvM than W3110 strain

One of the major aims of this study was to directly compare the production of a biopharmaceutical product under industrial conditions in the two extensively used *E. coli* strains: BL21 and W3110. The viable and cost-effective production of a POI using *E. coli* varies enormously depending on many different factors. POI related factors and upstream process parameters such as pH, temperature, media composition, strain type and others influence the recombinant expression (Tripathi 2016; Koopaei et al. 2018).

In our research, a screening of different conditions for the optimization of soluble yields of the scFvM was carried out for both strains listed above combining different temperatures, pH and inducer concentrations (described in Materials and Methods). However, a further optimization of

media composition and induction time could not be carried out since the implementation of these changes would require a complete revaluation of this automated protocol for 10 mL fermentation (Janzen et al., 2019).

In Figure 1A, the specific soluble product (soluble target protein) formation was calculated for BL21 and W3110 in 10 mL and 5 L bioreactors following the standard protocol at 7 h post-induction (T7) (refer to Material and Methods). To eliminate a possible impact of the different OD levels at both scales, the specific soluble product titer was determined by dividing soluble product titer by OD_{550} for both strains and scales at time point T7 (end of fermentation in 10 mL scale) (Figure 1A). The BL21 strain shows a specific soluble product titer of 23.5 mg/OD in 10 mL fermenters and 12.3 mg/OD in benchmark 5 L fermenters. Janzen et al. (Janzen et al. 2019) also described a higher specific soluble product titer formation in small-scale cultivations than in 5 L fermenters when employing a B strain. It has been suggested that BL21 may suffer from DO limitations when scaling up to 5 L fermenters (Kang et al. 2002). The W3110 strain, on the other hand, shows a very similar specific soluble product titers in both scales: 5.7 mg/OD in 10 mL fermenters and 4.9 mg/OD in 5 L fermenters. These results validate the robustness and reproducibility that this strain provides in industry (Kang et al. 2002; Yoon et al. 2012). However, with respect to the expression strains used, BL21 showed significantly higher titers in all direct comparisons (Figure 1A). Specific soluble product titer comparison between scales was also carried out with the optimized conditions after the screening experiments in 10 mL fermenters (refer to Material and Methods). Since a very similar pattern of soluble protein was obtained when comparing scales and strains, the data set is not shown because it had a comparable trend.

Looking more closely at the benchmark process, Figure 1B shows the soluble production of scFvM in BL21 and W3110 in 5 L fermenters and standard conditions at different time points: 0 h (T0), 4 h (T4), 7 h (T7) and 10 h (T10) post-induction. Soluble scFvM was quantified by an immunoassay from suspension samples. Overall, BL21 shows a higher soluble protein content (two-fold) during the entire induction period in the 5 L fermentation system compared to W3110 (Figure 1B). In BL21, the peak production of soluble protein is achieved 4 h after induction (2.61 g/L) and it remains stable until T10 (end of fermentation, 2.41 g/L). BL21 shows a tighter

regulation of the tac promoter under non-induced conditions (0 h post-induction): it leaks 0.33 g/L. On the other hand, W3110's peak production of soluble protein is achieved 7 h after induction (1.16 g/L) and it also remains stable until T10 (end of fermentation, 1.01 g/L). Unlike BL21, W3110's tac promoter is more leaky and produces more than half of the total soluble scFvM before induction (0.72 g/L). This leaky expression in W3110 could be linked to plasmid instability, which many times explains a poor yield of target protein (Rosano and Ceccarelli 2014). However, in this case, differences in yield between the chosen strains are not connected to plasmid loss or instability, as plasmid copy number (PCN) remains stable and comparable between them throughout the whole fermentation process (observed: ≈ 12-18 copies/cell; expected: 15-20 copies/cell). When plasmid instability is discarded, BL21 and W3110 critical genome differences for recombinant protein production should be considered to understand these yield differences. Even though BL21 and W3110 are both widely used in recombinant protein production, B strains are deficient in the Lon protease, which degrades many recombinant proteins. The B strain also lacks the outer membrane protease OmpT, whose function is to degrade extracellular proteins (Rosano and Ceccarelli 2014). These genetic differences between strains may explain the higher yields obtained with BL21. In addition, it should be noted that BL21 reaches a lower OD_{550} value than the W3110 strain (BL21: 212 and W3110: 272) at the end of fermentation. These OD differences between the compared strains might correlate with the metabolic burden caused by the continuous export to the periplasm by Sec pathway (Horga et al. 2018) and/or the lethal outer membrane punctures occurred as a result of limited periplasmic capacity (Tong et al. 2000; Schofield et al. 2016) in BL21. A second run of bioreactors with parameters optimised for BL21 (refer to Material and Methods) was carried out and similar patterns for yields and OD_{550} values were obtained compared to standard conditions (data set is not shown because it had a comparable trend).

Even when the focus is on soluble production, an additional inherent part of disulphide bonded protein production in *E. coli* cannot be dismissed: inclusion body (IB) formation. The Coomassie blue-stained gel in Figure 2 shows the lysates of BL21 and W3110 from 5 L fermenters when expressing OmpA SP-scFvM at T0, T4, T7 and T10 time-points in standard conditions. Cell

suspension was analysed and fractionated in total titer (TT, comprising soluble and insoluble proteins), total soluble (TS, comprising intracellular and extracellular soluble proteins) and supernatant samples (SN, only proteins located in the extracellular medium). The insoluble POI production is remarkably different between the compared strains (Figure 2). When looking at the Coomassie blue-stained gel, it is important to notice that total production of the POI (TT) is visually higher in W3110 than in BL21 at all time points. This result suggests that the majority of the protein is produced as IBs and only a small part is translocated to the periplasm and extracellular medium and is therefore soluble (TS). As explained before in Figure 1B, and as it can be noticed in the Coomassie blue-stained gel in Figure 2 (see and visually compare T0 scFvM production in both strains), W3110 pre-induction leakiness is higher than BL21's. We hypothesise that due to this early high-level expression in W3110, hydrophobic stretches in the polypeptide are present at high concentrations very early in the cell and are available for interaction with similar regions. This may lead to protein instability and aggregation (IB formation) (Carrió and Villaverde 2002; Rosano and Ceccarelli 2014). Over time, and in both strains, but more remarkably in BL21, the POI starts to be detectable in the supernatant due to the leakiness of the outer membrane (Tong et al. 2000; Schofield et al. 2016), active export (Xia et al. 2008) and/or lysis of the cells (Kleiner-Grote et al. 2018).

The purified scFvM shows multiple charged variants on a CEX

The second objective of this research was to determine whether the protein expressed by the two strains, after translocation in the periplasm, was similarly folded and contained the same charge heterogeneity. The produced scFvM was purified from periplasmic extract and from culture supernatant by nickel IMAC affinity chromatography. Analysis *via* Coomassie blue-stained gel (Figure 3A) shows that the scFvM was obtained with high purity, independently of the expression host and purified compartment. For BL21 460 µg from the periplasmic extract and 45.9 mg from the supernatant were obtained, while for W3110, the yields were 350 µg and 26.3 mg, respectively. It should be noted that the intracellular soluble titer (periplasmic fraction) was higher than the extracellular one in both strains (BL21: Intracellular 2.12 g/L Extracellular 0.47 g/L; W3110: Intracellular 0.91 g/L Extracellular 0.20 g/L). However, this was not reflected in the

product titer after purification. This difference depends on the low amount of periplasm that could be extracted due to setting constraints in the maximal cell pellet that can be processed (see Materials and Methods).

To further investigate the presence of possible heterogeneity of the expressed scFvM, the IMAC purified material from culture supernatants was separated by cation exchange chromatography (CEX) *via* gradient elution, since the scFvM has a basic isoelectric point of 7.8 (Expasy Protparam). CEX is typically considered a gold standard technique to separate and purify charge variants (Wagner-Rousset et al. 2017; Yüce et al. 2021). However, the results from this technique can be strongly influenced by differences in operational parameters such as column type, particle size and flow rate (Fekete et al. 2015; Jing et al. 2020). Since previous studies demonstrated the importance of the diameter of the resin particles and flow rate on the separation performance (Jing et al. 2020; Yüce et al. 2021), a small resin particle (10 µm diameter resin) coupled with a slow flow rate (0.5 mL/min) was selected in our case. The results indicated a high separation performance. The chromatograms show the presence of two main peaks: one more acidic (Main A) and another one more basic (Main B), both coupled with some minor subforms (Figure 3B). Between the two strains, the elution pattern is maintained, however, the relativity of the peaks slightly changes BL21 being richer in acidic variants while W3110 in basic ones (Figure 3C). To verify, firstly, if the multiple peaks showed different masses and impurities, a non-reducing gel was assessed. However, no differences could be detected (Figure 3D).

scFvM is correctly disulphide bonded and cleaved from its SP in BL21 and W3110

In this study an offline approach was applied, consisting in the isolation of the separated forms from CEX followed by individual analysis for better understanding of possible modifications. The workflow involved coupling size-exclusion chromatography (SEC) directly to mass spectrometry (LC-MS). This was done to verify that the scFvM was correctly folded and that different peaks on CEX were not caused by a pool of species with free thiols or uncleaved SP. IMAC and CEX purified samples (Main A and Main B) from both periplasm and supernatant samples from BL21 and W3110 were analysed by LC-MS. This analysis confirmed that all scFvM samples had the expected molecular weight, consistent with the cleavage of the SP when the POI

is exported from the cytoplasm to the periplasm and its four cysteines engaged in two DSBs (Table 1). The main component in all these samples is the unmodified scFvM molecule. In addition to it and as second species, both BL21 and W3110 samples show comparable reduction of 17 Da, probably due to N-terminal pyroglutamate modification (pyroQ), while gluconoylation is only seen in BL21 samples (+178 Da). PyroQ modification is generated after a non-enzymatic cyclization of N-terminal glutamine whose rate of formation can be affected by various environmental factors during purification and storage (Beyer et al. 2018; Beck and Liu 2019). In previous studies, CEX has been reported as a method of choice for the separation of pyroQ modifications since the loss of a primary amine causes an acidity shift of the antibody (Brorson and Jia 2014; Beyer et al. 2018). However, in this study the use of a strong cation exchange did not show the same results. In fact, LC-MS analysis run on each of the CEX peaks showed the presence of a -17 Da modifications, ranging from 12-25 % of the total protein, in each sample (Table 1). The strain selectivity of the non-enzymatic gluconoylation modification agrees with previous literature. Because B strains accumulate 6-phosphogluconolactone due to a deficiency of 6-phosphogluconolactonase (Meier et al. 2012), which promotes gluconoylation, it is not surprising that this strain generates some gluconoylated proteins. Although this modification can adversely affect protein quality, the gluconoylation is not very stable and can transform back into unmodified protein and gluconate *via* a hydrolysis reaction (Martos-Maldonado et al. 2018).

In addition, possible mismatches of the DSBs were also analysed among the multiple peaks in CEX (purified from BL21) by MS. In this case the purified protein from each CEX peak was digested. However, no differences in the size of the peptides generated were identified both in native conditions and after reduction, confirming that DSB shuffling is not essentially the reason for the heterogeneous pattern in CEX (Table 1). Disulfide bond shuffling occurs when an S-S bond between two cysteine residues is rearranged such that the residues are pairing with unpredicted partners.

Table 1*:* **Analysis of free thiol content, SP cleavage and secondary modifications based on MS**.

Strain	Technique	Location	N° of Cysteine	$M_{ox\ Theor}$[a]	M_{Exp}[b]	Δ mass	+178 Da (5-11 %)	-17/18 Da (12-25 %)
BL21	IMAC	Periplasm	4	27478.27	27478.27	0	+[c]	+
		Supernatant		27478.27	27478.27	0	+	+
	CEX	Main A		27478.27	27478.27	0	+	+
		Main B		27478.27	27478.27	0	+	+
W3110	IMAC	Periplasm		27478.27	27478.27	0	-	+
		Supernatant		27478.27	27478.27	0	-	+
	CEX	Main A		27478.27	27478.27	0	-	+
		Main B		27478.27	27478.27	0	-	+

a) Calculated theoretical oxidised molecular weight ($M_{OxTheor}$)

b) Experimental molecular weight (M_{Exp})

c) The sign "+" indicates the presence of the modification (+178 Da or -17 Da) while "-" indicates the absence of it.

The stability of the two main peaks excludes a handling artefact

A reversibility analysis on the two main peaks was then performed to verify if these two main CEX forms were not caused by downstream operation. The eluted single peaks were therefore pooled from different purifications in Main A and B respectively, dialyzed against the equilibration buffer and each pool was loaded again on the CEX. Figure 4A shows the comparison of the elution profiles from each pool. In both cases the purification revealed a perfect reproducibility of the peaks that seem to be only in a very slow reversible equilibrium with the counterpart. Moreover, since with this experiment the heterogeneous pattern observed during the CEX purification of the purified material (Figure 3B) was not visible, an artefact of the column caused by overloading of the samples could be excluded. The two peaks therefore represent scFvM isoforms that are stable under these conditions.

Biophysical characterization confirmed altered conformations of the two main charged variants

To further investigate potential conformational differences between the two main CEX forms, and the exact comparability among the two strains, a biophysical assessment was established *via* differential scanning calorimetry (DSC) and one-dimensional proton nuclear magnetic resonance

(1D-^{1}H-NMR) spectroscopy. DSC is commonly employed to assess the thermal and conformational stability of a protein under specific buffer conditions (Johnson 2013). The two separated main peaks (Main A and B) were evaluated by DSC (Figure 4B). One single transition, corresponding to the unfolding of the scFvM, was observed in all samples. The melting temperatures of the single Main Peaks A and B (from both BL21 and W3110) were 68.7 °C and 69.1 °C, respectively and corresponds to a temperature difference of 0.4 °C. These results proved a high comparability, in terms of thermal stability, of the POI produced by the two *E. coli* strains.

In addition to this thermal stability difference, a modification in the state of the protein from the two main peaks from the CEX was detected *via* 1D-^{1}H-NMR spectroscopy. This technique shows signals for each hydrogen atom in the protein that is covalently bound or exchanging slowly with water (for example amide signals will be present but those from -OH and NH_3 groups will be missing). These signals resonate at different frequencies (chemical shifts in ppm; parts per million of the main field) and with different intensities based on the 3D structure, ligand binding state and dynamics of the protein all of which affect local magnetic fields in the protein. The position of peaks in the 1D-^{1}H-NMR depend at first order on the chemistry of the atoms (Kwan et al. 2011), so for example CH_2 and CH_3 groups from different amino acid types (e.g., Val vs Leu) appear at different positions and also have small differences due to the primary sequence. On top of these chemical effects from residue types and the primary sequence the spectrum is also extremely sensitive to the 3D structure of the protein and very small changes in local environment and dynamics (see for example Chen et al. 2013) can be detected, so 1D-^{1}H-NMR can be used as a fingerprint of the proteins 3D structure and to monitor small changes in the state of the protein. In this case the spectra suggest both samples contain proteins that are well-folded as indicated by the well resolved and dispersed signals in the amide region (Figure 5) and a series of well-resolved methyl peaks at < 0 ppm which are indicative of stable methyl aromatic packing in the protein's core (boxes in Figure 5). On top of this, the spectra are similar enough to conclude that there have been no major changes in 3D structure and that the overall 3D fold is the same. However, while the methyl peaks at < 0 ppm are well dispersed, they also show variations between the samples with the peaks that resonate at -0.736 ppm and -0.981 ppm showing shifts of +0.04 ppm and -

0.02 ppm and ~ 10 % weaker peak intensities (Figure 5A). As the samples were very thoroughly dialysed these differences would indicate a change in conformation within the core of the protein or the presence of a yet unidentified strongly bound ligand.

All these results combined suggested that the two strains produced a target protein with the same overall 3D fold and thermal stability. However, the difference between the two main peaks of the CEX seems to be generated by a conformational change of the protein that leaves the mass unchanged as observed by MS analysis. Two main reasons could be behind this result: the presence of a ligand tightly bound to the protein or a modification that leaves the mass unchanged. With regards to the ligand, the reversibility analysis on the two main peaks suggests that, if present, the ligand binds with high affinity to one of the two forms since it could not be removed during the dialysis process. Moreover, at least, the presence of a high molecular mass molecule as ligand could be excluded since, prior LC-MS analysis, the samples were desalted by size exclusion chromatography (SEC) and no additional peaks were observed (data not available). Concerning the second hypothesis, various types of modification can occur during protein expression, manufacturing, and storage (Beyer et al. 2018; Beck and Liu 2019). Some of these modifications can lead to mass shifts while some others can result in protein modifications leaving the mass unchanged. Examples of the latter are DSB mismatch and aspartate isomerization which can further generate aspartic acid racemization (Wagner-Rousset et al. 2017). The presence of mismatched DSBs was excluded as a possible reason for multiple peaks in CEX, as described above. Aspartate (Asp) isomerization is a non-enzymatic modification that can cause conformational changes of the protein since it introduces an additional methyl group in the protein backbone (Beyer et al. 2018; Beck and Liu 2019). Furthermore, the specific structural outcome can lead to two isomeric products (L-isoAsp and D-isoAsp) where the D-amino acid can affect the peptide function (Riggs et al. 2017). This reaction occurs at an optimal pH of 5, produces succinimide (-18 Da specie) as a reaction intermediate and it is favoured on aspartate residues that are followed by a glycine (Ouellette et al. 2010; Nowak et al. 2017; Gupta et al. 2021). Moreover, it was shown in a previous study that antibody variants containing isoaspartate elute later in CEX (Harris et al. 2001). The current analysis was carried out at pH 5.5, a second and

more acid peak was present, in the protein sequence an aspartate close to a glycine is present and a species with -17/18 Da was detected. Therefore, the presence of aspartate isomerization as a possible modification cannot be excluded completely. There are some other methods to identify this modification such as LC-MS peptide mapping and 2D NMR analysis. LC-MS peptide mapping is becoming an important method for the characterization of primary sequences and PTMs in antibody products. On the other hand, this technique is labour-intensive, time-consuming and can introduce artificial PTMs, resulting in an overestimation of the target protein modifications. During sample preparation, long digestions can generate unnecessary reactions interfering with the quantitation of the peaks of interest, causing low reproducibility of the results. Shortening digestion-time can cause incomplete peptide cleavages, thus low sequence coverage and poor repeatability (Jiang et al. 2020). In this case study, where BL21 and W3110 protein product comparison was the aim, this technique had to be discarded due to reproducibility issues. On the other hand, 2D NMR analysis typically utilises labelled isotopes during the fermentation process for proteins of this mass, and therefore, this approach was not available due to experimental constraints.

Conclusion

In this case-study, we report that *E. coli* strain differences may have an influence on the final product yield but not necessarily on its heterogeneity pattern. The results showed that BL21 and W3110 have a very different productivity profile in the conditions employed, with BL21 being more industrially relevant to produce this specific scFv in terms of yield. In terms of quality, except for the B strain characteristic gluconoylation, the expressed scFvM displays a similar heterogeneity profile. This resemblance of the produced product represents an indication of the interchangeability between the two strains, a characteristic that presents an important perspective for biosimilar and biobetter production. In our paper it is further shown that when scFvM product quality was assessed by different analytical methods, a distinctive profile of the product was obtained, suggesting that alterations in the recovered product are present independently of the host strain. These alterations appear to be stable and significant, since the two forms elute at very different salt concentrations during CEX. However, the exact identity of the cause of product

heterogeneity could not be appropriately confirmed. This opens a discussion on whether MS based intact mass analysis of the POI is enough to spot the heterogeneity in a product. At the same time, we want to encourage the discoveries of new, fast and affordable methods for analysis of heterogeneity other than 2D NMR and LC-MS based peptide mapping for the identification of protein heterogeneity in biopharma, which are time-consuming.

Materials and Methods

All chemicals, reagents and enzymes were of highest quality and were obtained from Sigma-Aldrich, Roth or Thermo Fisher Scientific, unless otherwise noted.

anti-c-Met scFv (scFvM) expression strain generation

E. coli DH5α (Invitrogen) was used for genetic manipulations. The anti-c-Met scFv (sequence taken from Edwardraja et al. (Edwardraja et al. 2010)), with an N-terminal wild-type OmpA SP and a C-terminal 6x His-tag was commercially synthesised (GeneArt). The synthesis construct was sub-cloned into the pFLAG-CTC vector (Sigma Aldrich) under the control of a tac promoter using NdeI and EcoRI restriction sites. This construct will be termed as OmpA-scFvM in this work. Individual clones were sequenced before transforming the expression plasmid into the expression strains *E. coli* BL21 (Novagen) and W3110 (DSMZ). In this work, the protein anti-c-Met scFv is referred to as scFvM to further correlate with Edwardraja et al. and Jones et al. (Edwardraja et al. 2010; Jones et al. 2016) which investigated the same protein in a different setup.

Expression in a miniaturised fermentation platform (Multifermenter, MF) and a 5 L fermentation system

The fully automated cultivation at 10 mL scale in the MF was performed as described in Janzen et al. (Janzen et al. 2019) with the exception that the temperature in all reactors was set to 37 °C and lowered to the corresponding experiment temperature prior to induction. For the screening of conditions, a range of different temperatures (25-33.5 °C), pH values (6.3-7.3) and isopropyl β-D-1-thioglactopyranoside (IPTG) inducer concentrations (0.5-1 mM) were tested with a DoE

setting in 32 MF vessels. While for BL21 a slight increase in yields and OD was observed in a screened condition set-up (refer as optimised conditions), with W3110 no optimisation could be achieved. Therefore, solely the standard conditions developed by Janzen et al. (Janzen et al. 2019) and optimised conditions developed for BL21 after MFs run were tested in quadruplicates in 10 mL bioreactors for BL21 and W3110 to maintain a better comparability within the study. In the standard set-up, temperature was set to 30 °C prior to induction and the pH was constantly maintained at 6.8. Cultures in this case were induced with 1 mM IPTG (0.024 mL from 75 mM stock). In the optimised BL21 set-up, temperature was set to 32 °C prior to induction and the pH was constantly maintained at 7.3. Cultures in this set-up were induced with 0.5 mM IPTG (0.012 mL from 75 mM stock). In all cases, the DO level was maintained at ≥ 35 %. In the case of the W3110 strain, a pre-culture in shake-flask was performed, since a significantly prolonged batch phase (e.g. lag phase of the cells) interfered with the fermentation protocol in the MF. The pre-culture was performed at 37 °C and 250 rpm until the culture reached an OD_{550} value of 2. The OD_{550} measurement of the pre-culture was manually performed (Genesys 10S UV-Vis; Thermo Fisher Scientific). In the case of BL21, bioreactors were directly inoculated from the cell bank.

Benchmark fed-batch cultivations were performed in fully controlled 5 L standard stirred-tank bioreactor systems (BIOSTAT Cplus; Sartorius Stedim) and the manufacturer provided PCS (MFCS-Win; Sartorius Stedim). Calibration and cultivation conditions and the used material and equipment are described in Janzen et al. (Janzen et al. 2019). As in MF experiments, the two set-ups (standard conditions, data shown in Results and Discussion section and optimised for BL21, data set not shown as it had a comparable trend as standard conditions) of the experiments were carried out with slight differences in the temperature, pH and induction concentration parameters (more details above). In all cases, the DO level was maintained at ≥ 35 % and the pH was kept constant at the set pH ± 0.2 using 25 % ammonia and 3 M phosphoric acid. Cultures were induced either with 0.5 mM or 1 mM IPTG (11.8 mL or 23.6 mL from 211.9 mM stock). Samples for product quantification, plasmid copy number (PCN) estimation and OD_{550} determination were manually withdrawn before induction (T0) and 4 h (T4), 7 h (T7) and 10 h (T10, end of cultivation) after the IPTG pulse. OD_{550} measurements were directly performed (Genesys 10S

UV-Vis; Thermo Fisher Scientific). Samples for product quantification and PCN estimation were stored in reaction tubes at – 20 °C until further use.

The synthetic media used for all cultivations (both 10 mL and 5 L systems) was prepared as described in Striedner et al. (Striedner et al. 2010). The medium was supplemented with 1 mL/L antifoam agent (PPG 2000; Dow Chemical Co.) and was autoclaved for SIP prior cultivation in case of the 5 L system. In the MFs, each block was equipped with bioreactors which were aseptically filled with sterile medium supplemented with 1 mL/L antifoam agent (PPG 2000; Dow Chemical Co.) in a laminar flow hood.

Fermentation sample preparation and quantification

Quantitative analysis was performed by automated immunoassay (Gyros). The sample preparation was conducted with a liquid handling system (Tecan) in 96-well format. For fermentation suspension samples, high viscosity due to leaked nucleic acids caused by fermentation condition and sample freeze and thawing was encountered. High viscosity causes imprecise pipetting by the liquid handling robot. Therefore, as an initial step a nucleic acid hydrolysis with Benzonase® (Merck) (0.5 U/µL for ≥ 10 min, 450 rpm at room temperature (RT)) was performed. Cell lysis was performed by incubation with 1/10 v/v Lysonase (Merck) in FastBreak cell lysis reagent (Promega) for 30 min at RT with shaking at 450 rpm. The soluble fraction was analysed by the immunoassay. For this purpose, the digested cells were centrifuged at 2900 g for 10 min (Tecan centrifuge Hettich Rotanta) and the supernatant was further used. Samples were diluted in the analysis buffer (RexxipA (Gyros)) for quantification.

In the case of supernatant samples from fermentation, straight dilution in the analysis buffer (RexxipA (Gyros)) was performed.

Content quantification was performed using a Gyrolab xPlore by an automated immunoassay with an scFv-specific antibody (109-066-097 (Jackson ImmunoResearch), biotinylated for immobilization within the Gyros CD-microstructure) and an his-tag specific antibody (34670 (Qiagen), Alexa647 fluorescence labelled for detection). The Gyros protocol (200-3W-002-A) was performed according to the manufacturer's instructions. The standard curve was analysed

with the Gyros Evaluator SW using a five-parameter fit. Quantification was performed in the linear range of the standard curve (15 ng/mL to 1000 ng/mL).

Plasmid copy number (PCN) estimation

For PCN estimation, 5/OD_{550} fermentation pellet samples were used. Fully automated plasmid extraction was performed using the QIAprep Spin Miniprep Kit (Qiagen) and the QIAcube Connect (Qiagen). DNA concentrations were measured using a NanoDropTM (Thermo Fisher Scientific). The following equation was used to estimate the PCN:

$$PCN = \frac{Plasmid\ concentration\ \left[\frac{ng}{\mu L}\right] * 30\ \mu L * 1.32\ (correction\ factor\ for\ plasmid\ loss)}{nucleotides\ of\ plasmid * 308.95\ \frac{Da}{nucleotide} * \left(1.66 * 10^{-15}\frac{ng}{Da}\right) * (7.4 * 10^{10}\frac{cells}{pellet})}$$

The weight per single plasmid in ng was calculated from the number of nucleotides and the conversion factors 308.95 (mean weight per nucleotide in Da) and $1.66 \cdot 10^{-15}$ (conversion from Da to ng). DNA loss during plasmid preparation (determined by spiking experiments) and the average number of cells in 5/OD_{550} pellet samples ($7.4 \cdot 10^{10}$, determined in a cell counting chamber (Marienfeld Superior)) were considered to calculate the number of plasmids per single cell.

Isolation of scFvM from periplasm and supernatant and content quantification

The expressed scFvM was purified from supernatant and periplasm. The periplasmic extraction was achieved following the pureFrac protocol as described elsewhere (Malherbe et al. 2019). Since a larger volume was necessary an upscaling factor of 130 was applied during all steps. The periplasmic fractions and the supernatant from BL21 and W3110 were then applied on a Ni SepharoseTM 6 Fast Flow (Cytiva) column on the ÄKTA Avant chromatography system (GE Healthcare) at RT and purified *via* immobilised metal affinity chromatography (IMAC). At a flow rate of 1 mL/min low binding impurities were washed from the column with 10 column volumes of equilibration buffer (20 mM Phosphate, 500 mM NaCl, 20 M Imidazole, pH 7.4). The protein was then eluted with 5 column volumes of 100 % elution buffer (20 mM Phosphate, 500 mM NaCl, 500 mM Imidazole, pH 7.4). The protein content was monitored online by absorbance at 280 nm. After each purification the column was stripped, washed, and recharged to avoid

contaminations between the different strains and purified compartments. The eluted protein was dialyzed against 20 mM Phosphate, 20 mM NaCl, at pH 5.5 using 3.5 MWCO dialysis cassettes (Thermo Scientific) and stored at -20 °C between purification steps.

With the eluted and rebuffered fractions (20 mM Phosphate, 20 mM NaCl, pH 5.5) from the supernatant a cation exchange chromatography (CEX) with a MonoSTM 5/50 GL (Cytiva) column was performed on an ÄKTA Purifier chromatography system (GE Healthcare) at RT. The used buffers were 20 mM Phosphate, 20 mM NaCl, at pH 5.5 (equilibration buffer) and 20 mM Phosphate, 500 mM NaCl, at pH 5.5 (elution buffer). At a flow rate of 0.5 mL/min, a gradient from 0-100 % elution was applied for 30 column volumes followed by 10 column volumes at 100 % elution buffer. The protein content was monitored online by absorbance at 280 nm. The eluted protein was dialyzed against 20 mM Phosphate, 20 mM NaCl, at pH 5.5 using 3.5 MWCO dialysis cassettes (Thermo Scientific) and concentrated with 3 kDa MWCO Pierce™ Protein Concentrators PES (ThermoFisher). The fractions were stored at -20 °C until analysed.

Purified protein concentration was determined by measuring the absorbance at 280 nm with the Nanodrop 2000 (ThermoFisher) and a calculated molar extinction coefficient of 58'580 M^{-1} (Expasy Protparam). Affinity purified scFvM and its charged variants (CEX samples) were denatured for 5 min at 80 °C and run on an SDS-PAGE gel under reducing and non-reducing conditions.

DSC analysis

The DSC experiments were performed using a MicroCal VP-DSC system (Malvern). All samples were dialyzed against the same buffer (20 mM Phosphate, 20 mM NaCl, pH 5.5) prior analysis using 3.5 MWCO dialysis cassettes (Thermo Scientific). The reference cell was filled with a buffer corresponding to the sample buffer. The samples were heated from 10 °C to 95 °C at a heating rate of 60 °C/h. The pre-scan was 3 min, the filtering period was 10 s, and the feedback mode/gain was set to passive. The midpoint of thermal transition temperature (Tm) was obtained by analysing the data using OriginTM 7 software. All experiments were performed at a protein concentration of 1 mg/mL.

Mass spectrometry analysis

For intact mass analysis, samples were injected without prior sample preparation into the UltiMate™ 3000 UHPLC system coupled to the Orbitrap Eclipse™ Tribrid™ mass spectrometer (all Thermo Fisher Scientific). Proteins were loaded on an ACQUITY UPLC BEH200 SEC, 1.7 µm, 4.6 x 150 mm, applying a 10 min isocratic method (20 % B), with a flow rate of 0.2 mL/min and 0.1 % Formic acid (FA) (Fisher Chemical, LC-MS grade) in water as mobile phase A and 0.1 % FA in Acetonitrile (ACN) (Merck, Hypergrade for LC-MS) as mobile phase B. Electrospray ionisation was performed in positive ionisation mode and molecules analysed in the Orbitrap with a scan range of 500-2000 m/z and a resolution set to 240 000 (at 200 m/z) for full scan.

For DSB analysis, proteins were precipitated with $CHCl_3$/Methanol, dried at RT and subsequently dissolved in lysis buffer (7.6 M urea/50 mM Tris-HCl, pH 8), diluted with 50 mM Tris-HCl, at pH 8 and digested with Trypsin/Lys-C Mix (Promega).

Peptides were analysed on the same LC-MS system and mobile phases as for intact mass analysis. Peptides were separated on a ACQUITY UPLC Peptide CSH C18, 130 Å, 1.7 µm, 2.1x150 mm applying a 25 min gradient from 5-30 % B, increasing further to 95 % B within 5 min, resulting in total run time of 44 min, with a flow rate of 0.25 mL/min. Electrospray ionization was performed in positive ionization mode, the resolution was set to 120 000 (at 200 m/z), with a scan range of 200-2000 m/z for MS1 analysis. A Top N method was applied for fragmentation with CID Assisted Collision and resulting fragments analysed in the Orbitrap at a resolution of 30 000 (at 200 m/z).

The raw data files were subjected to the Byos software (v 4.2) from Protein Metrics Inc. for data processing and reporting. For intact mass evaluation, peaks found in the total ion chromatogram were integrated and full mass spectra were deconvoluted. For DSB analysis, the Byos DSB workflow was used, searching against a built-in database based on the sequence of the POI.

1D-^{1}H-NMR analysis

Samples were dissolved in a buffer containing 20 mM Phosphate, 20 mM NaCl, at pH 5.5 and were analysed by 1D-^{1}H-NMR after extensive dialysis in a common pool of buffer to reduce

effects due to any systematic differences in sample preparation. Before NMR spectra were collected 5 % D_2O was added as a lock solvent. Spectra were acquired at 25 °C using a double pulse field gradient spin echo sequence (DPFGSE) to suppress water (Hwang and Shaka 1995) on a Bruker Advance III spectrometer operating at 600 MHz with a He cooled QCI-P cryogenic probe using Topspin 3.6.1. Spectra were measured using 32k complex data points over a sweep width of 9615 Hz using 1024 scans with an inter-scan relaxation delay of 1 s and 4 dummy scans for equilibration. These data were processed and analysed using Topspin and an exponential window function of 3 Hz was applied to improve signal to noise.

List of abbreviations

CEX: Cation exchange chromatography

DO: Dissolved oxygen

DoE: Design of experiment

DSB: Disulphide bond

DSC: Differential scanning calorimetry

IB: Inclusion body

IMAC: Immobilised metal affinity chromatography

IPTG: isopropyl β-D-1-thioglactopyranoside

MF: Multifermenter

MS: Mass spectrometry

MWCO: Molecular weight cut-off

NMR: Nuclear magnetic resonance

OD: Optical density

PCN: Plasmid copy number

POI: Protein of interest

PTM: Post-translational modification

scFv: Single-chain variable fragment

SP: Signal peptide

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data used and /or analysed during the current study are available from the corresponding authors on reasonable requests.

Competing interests

All authors declare that they have no competing interests.

Funding

This work was funded by the People Programme (Marie Skłodowska-Curie Actions) of the European Union's Horizon 2020 Programme under REA grant agreement no. 813979 (SECRETERS).

Author's contributions

KAA, LB, KK performed the research. GT performed NMR experiments and analysis. MB and CR conceived and coordinated the study. All authors were involved in experimental planning and have read and approved the final manuscript.

Acknowledgments

The authors acknowledge Dr. Laura Niederstaetter, Dr. Martin Voigtmann, Patricia Cantarell-Bargueño, and Dr. Nicole Weiner for carrying out, respectively, the mass spectrometry, immunoassay, and DSC analyses. Additionally, the authors acknowledge Alma Medimorec, Gabriela Bachleda-Kubanska and Michael Deimel for their technical support during fermentation runs.

Additional information

Luisa Buscajoni and Klaudia Arauzo-Aguilera contributed equally to this work.

Figure Legends

Figure 1**. Soluble scFvM production in BL21 and W3110 in 10 mL and 5 L scale fermentations.** A) Specific soluble product titer in BL21 and W3110 in 10 mL and 5 L bioreactors expressed in mg/OD. To calculate these values, the 7 h post-induction soluble scFvM titer is divided by the OD_{550} of the culture at that time in both scale bioreactors. 10 mL fermentations specific soluble product formation values are the average of four replicates, error bars have been added to the figure. B) Soluble titer variation of the scFvM in BL21 and W3110 strains at different time points (0 h (T0), 4 h (T4), 7 h (T7) and 10 h (T10) post-induction) in 5 L fermentations expressed in g/L. In both graphs, BL21 strain is represented in black and W3110 in grey. Soluble scFvM product was determined by immunoassay from suspension samples. 5 L bioreactors were run two times for each strain with slight variations in temperature, inductor concentration and pH resulting in comparable profiles for OD_{550} and titer.

Figure 2. **Lysates of BL21 and W3110 from 5 L fermenters when expressing OmpA-scFvM.** Suspension and supernatant samples of BL21 and W3110 expressing OmpA-scFvM were recovered at different time points: T0, T4, T7 and T10 for SDS-PAGE analysis. Representative Coomassie blue-stained gel of the total titer (TT, comprising soluble and insoluble proteins), total soluble (TS, comprising intracellular and extracellular soluble proteins) and supernatant samples

(SN, only proteins located in the supernatant) and scFvM reference protein. Ladder (Mark12™ Unstained Standard) on the left in kDa.

Figure 3. **scFvM purification *via* two steps chromatography**. A) Representative Coomassie blue-stained gel of the total protein sample loaded (L), column flow-though (FT) and eluate (E) from either the periplasm or the culture supernatant in non-reducing condition. Ladder (Mark12™ Unstained Standard) on the left in kDa. B) Representative MonoS normalised chromatograms of scFvM purified from the culture supernatant. Smaller peaks that elute around the Main peak A are referred to as acidic species, and peaks that elute around Main peak B are referred to as basic species. C) Relative abundance of the CEX peaks in BL21 and W3110. BL21 strain is represented in black and W3110 in grey. D) Representative Coomassie blue-stained gel of a CEX run in non-reducing condition. Main A elution fractions are in the continued box while Main B ones are in the dotted box.

Figure 4. **Stability analysis of the two main peaks from the CEX.** A) Overlaid MonoS chromatograms of the two main peaks that were collected, dialyzed, and reloaded separately to demonstrate the stability and purity of the two forms. B) DSC profiles of the peaks Main A and Main B obtained from CEX purification of scFvM expressed by BL21 and W3110. The obtained Tms are: Main A 68.7 °C and Main B 69.1 °C. Straight line represents Main A and dotted line Main B. BL21 strain is represented in black and W3110 in grey.

Figure 5. **^{1}H-NMR spectra of Main A and Main B conformers measured at 600 MHz.** Spectra from basic, acidic species and control samples show the conformational and temporal stability of the species. The samples were many times co-dialysed before analysis. Samples conditions: pH 5.5 and 25 °C, sample concentration ~ 50 µM. A) Spectra from BL21 and W3110 showing the clear difference in the fingerprint of Main A (blue) and Main B (purple) conformers. The box represents the expansion of the high field methyl aromatic fingerprint region. B) Spectra from two separate batches of protein (blue and purple) from BL21 representing analysis of the same species showing reproducible conformational state as a control. The box represents the expansion of fingerprint region from shielded methyl groups.

References

Ali S, Perez-Pardo MA, Aucamp JP, et al (2012) Characterization and feasibility of a miniaturized stirred tank bioreactor to perform E. coli high cell density fed-batch fermentations. Biotechnol Progr 28:66–75. https://doi.org/10.1002/btpr.708

Baumann P, Hubbuch J (2017) Downstream process development strategies for effective bioprocesses: Trends, progress, and combinatorial approaches. Eng Life Sci 17:1142–1158. https://doi.org/10.1002/elsc.201600033

Beck A, Liu H (2019) Macro- and Micro-Heterogeneity of Natural and Recombinant IgG Antibodies. Antibodies 8:18. https://doi.org/10.3390/antib8010018

Beyer B, Schuster M, Jungbauer A, Lingg N (2018) Microheterogeneity of Recombinant Antibodies: Analytics and Functional Impact. Biotechnol J 13:1700476. https://doi.org/10.1002/biot.201700476

Brorson K, Jia AY (2014) Therapeutic monoclonal antibodies and consistent ends: terminal heterogeneity, detection, and impact on quality. Curr Opin Biotech 30:140–146. https://doi.org/10.1016/j.copbio.2014.06.012

Carrió MM, Villaverde A (2002) Construction and deconstruction of bacterial inclusion bodies. J Biotechnol 96:3–12. https://doi.org/10.1016/s0168-1656(02)00032-9

Chen H, Huang Z, Dutta K, et al (2013) Cracking the Molecular Origin of Intrinsic Tyrosine Kinase Activity through Analysis of Pathogenic Gain-of-Function Mutations. Cell Reports 4:376–384. https://doi.org/10.1016/j.celrep.2013.06.025

Edwardraja S, Neelamegam R, Ramadoss V, et al (2010) Redesigning of anti-c-Met single chain Fv antibody for the cytoplasmic folding and its structural analysis. Biotechnol Bioeng 106:367–375. https://doi.org/10.1002/bit.22702

Fekete S, Beck A, Guillarme D (2015) Characterization of cation exchanger stationary phases applied for the separations of therapeutic monoclonal antibodies. J Pharmaceut Biomed 111:169–176. https://doi.org/10.1016/j.jpba.2015.03.041

Galloway CA, Sowden MP, Smith HC (2003) Increasing the Yield of Soluble Recombinant Protein Expressed in E. coli by Induction during Late Log Phase. Biotechniques 34:524–530. https://doi.org/10.2144/03343st04

Gaza-Bulseco G, Li B, Bulseco A, Liu H (2008) Method to Differentiate Asn Deamidation That Occurred Prior to and during Sample Preparation of a Monoclonal Antibody. Anal Chem 80:9491–9498. https://doi.org/10.1021/ac801617u

Gupta S, Jiskoot W, Schöneich C, Rathore AS (2021) Oxidation and Deamidation of Monoclonal Antibody Products: Potential Impact on Stability, Biological Activity, and Efficacy. J Pharm Sci 111:903–918. https://doi.org/10.1016/j.xphs.2021.11.024

Harris RJ, Kabakoff B, Macchi FD, et al (2001) Identification of multiple sources of charge heterogeneity in a recombinant antibody. J Chromatogr B Biomed Sci Appl 752:233–245. https://doi.org/10.1016/s0378-4347(00)00548-x

Horga LG, Halliwell S, Castiñeiras TS, et al (2018) Tuning recombinant protein expression to match secretion capacity. Microb Cell Fact 17:199. https://doi.org/10.1186/s12934-018-1047-z

Huleani S, Roberts MR, Beales L, Papaioannou EH (2021) Escherichia coli as an antibody expression host for the production of diagnostic proteins: significance and expression. Crit Rev Biotechnol 42:756–773. https://doi.org/10.1080/07388551.2021.1967871

Humphreys DP, Vetterlein OM, Chapman AP, et al (1998) F(ab′)2 molecules made from Escherichia coli produced Fab′ with hinge sequences conferring increased serum survival in an animal model. J Immunol Methods 217:1–10. https://doi.org/10.1016/s0022-1759(98)00061-1

Hwang TL, Shaka AJ (1995) Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. J Magnetic Reson Ser 112:275–279. https://doi.org/10.1006/jmra.1995.1047

Janzen NH, Striedner G, Jarmer J, et al (2019) Implementation of a Fully Automated Microbial Cultivation Platform for Strain and Process Screening. Biotechnol J 14:1800625. https://doi.org/10.1002/biot.201800625

Jiang P, Li F, Ding J (2020) Development of an efficient LC-MS peptide mapping method using accelerated sample preparation for monoclonal antibodies. J Chromatogr B 1137:121895. https://doi.org/10.1016/j.jchromb.2019.121895

Jing S-Y, Gou J-X, Gao D, et al (2020) Separation of monoclonal antibody charge variants using cation exchange chromatography: Resins and separation conditions optimization. Sep Purif Technol 235:116136. https://doi.org/10.1016/j.seppur.2019.116136

Johnson CM (2013) Differential scanning calorimetry as a tool for protein folding and stability. Arch Biochem Biophys 531:100–109. https://doi.org/10.1016/j.abb.2012.09.008

Jones AS, Austerberry JI, Dajani R, 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 - Mol Cell Res 1863:3116–3124. https://doi.org/10.1016/j.bbamcr.2016.09.006

Kang D, Kim Y, Cha H (2002) Comparison of green fluorescent protein expression in two industrial Escherichia coli strains, BL21 and W3110, under co-expression of bacterial hemoglobin. Appl Microbiol Biot 59:523–528. https://doi.org/10.1007/s00253-002-1043-3

Khawli LA, Goswami S, Hutchinson R, et al (2010) Charge variants in IgG1. Mabs 2:613–624. https://doi.org/10.4161/mabs.2.6.13333

Kipriyanov SM (2009) Antibody Phage Display, Methods and Protocols. Methods Mol Biology 562:205–214. https://doi.org/10.1007/978-1-60327-302-2_16

Kleiner-Grote GRM, Risse JM, Friehs K (2018) Secretion of recombinant proteins from E. coli. Eng Life Sci 18:532–550. https://doi.org/10.1002/elsc.201700200

Koopaei NN, Khadiv-Parsi P, Khoshayand MR, et al (2018) Optimization of rPDT fusion protein expression by Escherichia coli in pilot scale fermentation: a statistical experimental design approach. Amb Express 8:135. https://doi.org/10.1186/s13568-018-0667-3

Kwan AH, Mobli M, Gooley PR, et al (2011) Macromolecular NMR spectroscopy for the non-spectroscopist. Febs J 278:687–703. https://doi.org/10.1111/j.1742-4658.2011.08004.x

Lee HJ, Lee CM, Kim K, et al (2018) Purification of antibody fragments for the reduction of charge variants using cation exchange chromatography. J Chromatogr B 1080:20–26. https://doi.org/10.1016/j.jchromb.2018.01.030

Liu H, Ponniah G, Zhang H-M, et al (2014) In vitro and in vivo modifications of recombinant and human IgG antibodies. Mabs 6:1145–1154. https://doi.org/10.4161/mabs.29883

Malherbe G, Humphreys DP, Dav E (2019) A robust fractionation method for protein subcellular localization studies in Escherichia coli. Biotechniques 66:171–178. https://doi.org/10.2144/btn-2018-0135

Marisch K, Bayer K, Scharl T, et al (2013) A Comparative Analysis of Industrial Escherichia coli K–12 and B Strains in High-Glucose Batch Cultivations on Process-, Transcriptome- and Proteome Level. Plos One 8:e70516. https://doi.org/10.1371/journal.pone.0070516

Martos-Maldonado MC, Hjuler CT, Sørensen KK, et al (2018) Selective N-terminal acylation of peptides and proteins with a Gly-His tag sequence. Nat Commun 9:3307. https://doi.org/10.1038/s41467-018-05695-3

Meier S, Jensen PR, Duus JØ (2012) Direct Observation of Metabolic Differences in Living Escherichia Coli Strains K-12 and BL21. Chembiochem 13:308–310. https://doi.org/10.1002/cbic.201100654

Moorhouse KG, Nashabeh W, Deveney J, et al (1997) Validation of an HPLC method for the analysis of the charge heterogeneity of the recombinant monoclonal antibody IDEC-C2B8

after papain digestion1Presented at the Well Characterized Biotechnology Pharmaceuticals Meeting in San Francisco, 6–8 January 1997.1. J Pharmaceut Biomed 16:593–603. https://doi.org/10.1016/s0731-7085(97)00178-7

Nascimento A, Pinto IF, Chu V, et al (2018) Studies on the purification of antibody fragments. Sep Purif Technol 195:388–397. https://doi.org/10.1016/j.seppur.2017.12.033

Noronha SB, Yeh HJC, Spande TF, Shiloach J (2000) Investigation of the TCA cycle and the glyoxylate shunt in Escherichia coli BL21 and JM109 using 13C-NMR/MS. Biotechnol Bioeng 68:316–327. https://doi.org/10.1002/(sici)1097-0290(20000505)68:3<316::aid-bit10>3.0.co;2-2

Nowak C, Ponniah G, Neill A, Liu H (2017) Characterization of succinimide stability during trypsin digestion for LC-MS analysis. Anal Biochem 526:1–8. https://doi.org/10.1016/j.ab.2017.03.005

Ouellette D, Alessandri L, Chin A, et al (2010) Studies in serum support rapid formation of disulfide bond between unpaired cysteine residues in the VH domain of an immunoglobulin G1 molecule. Anal Biochem 397:37–47. https://doi.org/10.1016/j.ab.2009.09.027

Panula-Perälä J, Šiurkus J, Vasala A, et al (2008) Enzyme controlled glucose auto-delivery for high cell density cultivations in microplates and shake flasks. Microb Cell Fact 7:31–31. https://doi.org/10.1186/1475-2859-7-31

Ren D, Pipes GD, Liu D, et al (2009) An improved trypsin digestion method minimizes digestion-induced modifications on proteins. Anal Biochem 392:12–21. https://doi.org/10.1016/j.ab.2009.05.018

Rettenbacher LA, Arauzo-Aguilera K, Buscajoni L, et al (2021) Microbial protein cell factories fight back? Trends Biotechnol 40:576–590. https://doi.org/10.1016/j.tibtech.2021.10.003

Riggs DL, Gomez SV, Julian RR (2017) Sequence and Solution Effects on the Prevalence of d-Isomers Produced by Deamidation. Acs Chem Biol 12:2875–2882. https://doi.org/10.1021/acschembio.7b00686

Rosano GL, Ceccarelli EA (2014) Recombinant protein expression in Escherichia coli: advances and challenges. Front Microbiol 5:172. https://doi.org/10.3389/fmicb.2014.00172

Rudge SR, Nims RW (2017) ICH Quality Guidelines: An Implementation Guide. 467–486. https://doi.org/10.1002/9781118971147.ch17

Sandomenico A, Sivaccumar JP, Ruvo M (2020) Evolution of Escherichia coli Expression System in Producing Antibody Recombinant Fragments. Int J Mol Sci 21:6324. https://doi.org/10.3390/ijms21176324

Schofield DM, Templar A, Newton J, Nesbeth DN (2016) Promoter engineering to optimize recombinant periplasmic Fab′ fragment production in Escherichia coli. Biotechnol Progr 32:840–847. https://doi.org/10.1002/btpr.2273

Shiloach J, Kaufman J, Guillard AS, Fass R (1996) Escherichia coli BL21 (hDE3) and Escherichia coli JM109. Biotechnology and Bioengineering 49:421–428. https://doi.org/10.1002/(sici)1097-0290(19960220)49:4<421::aid-bit9>3.0.co;2-r

Shiloach J, Rinas U (2009) Glucose and Acetate Metabolism in E. coli – System Level Analysis and Biotechnological Applications in Protein Production Processes. Systems Biology and Biotechnology of Escherichia coli. https://doi.org/10.1007/978-1-4020-9394-4_18

Singh SK, Kumar D, Malani H, Rathore AS (2021) LC–MS based case-by-case analysis of the impact of acidic and basic charge variants of bevacizumab on stability and biological activity. Sci Rep-uk 11:2487. https://doi.org/10.1038/s41598-020-79541-2

Striedner G, Pfaffenzeller I, Markus L, et al (2010) Plasmid-free T7-based Escherichia coli expression systems. Biotechnol Bioeng 105:786–794. https://doi.org/10.1002/bit.22598

Tong L, Lin Q, Wong WKR, et al (2000) Extracellular Expression, Purification, and Characterization of a Winter Flounder Antifreeze Polypeptide from Escherichia coli. Protein Expres Purif 18:175–181. https://doi.org/10.1006/prep.1999.1176

Tripathi NK (2016) Production and Purification of Recombinant Proteins from Escherichia coli. Chembioeng Rev 3:116–133. https://doi.org/10.1002/cben.201600002

Tripathi NK, Shrivastava A (2019) Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. Frontiers Bioeng Biotechnology 7:420. https://doi.org/10.3389/fbioe.2019.00420

Wagner-Rousset E, Fekete S, Morel-Chevillet L, et al (2017) Development of a fast workflow to screen the charge variants of therapeutic antibodies. J Chromatogr A 1498:147–154. https://doi.org/10.1016/j.chroma.2017.02.065

Xia X, Han M, Lee SY, Yoo J (2008) Comparison of the extracellular proteomes of Escherichia coli B and K-12 strains during high cell density cultivation. Proteomics 8:2089–2103. https://doi.org/10.1002/pmic.200700826

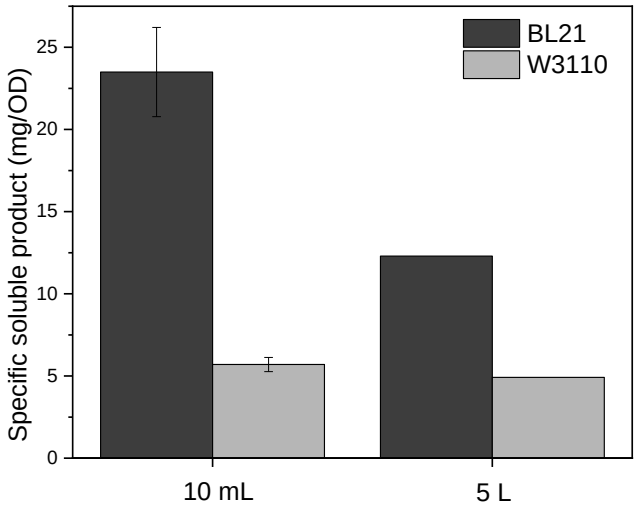
Yoon SH, Han M-J, Jeong H, et al (2012) Comparative multi-omics systems analysis of Escherichia coli strains B and K-12. Genome Biol 13:R37–R37. https://doi.org/10.1186/gb-2012-13-5-r37

Yüce M, Sert F, Torabfam M, et al (2021) Fractionated Charge Variants of Biosimilars: A Review of Separation Methods, Structural and Functional Analysis. Anal Chim Acta 1152:238189. https://doi.org/10.1016/j.aca.2020.12.064

Zhang G, Brokx S, Weiner JH (2006) Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in Escherichia coli. Nat Biotechnol 24:100–104. https://doi.org/10.1038/nbt1174

FIGURE 1

A)


BL21
W3110
Specific soluble product (mg/OD)
0
5
10
15
20
25
10 mL
5 L

B)

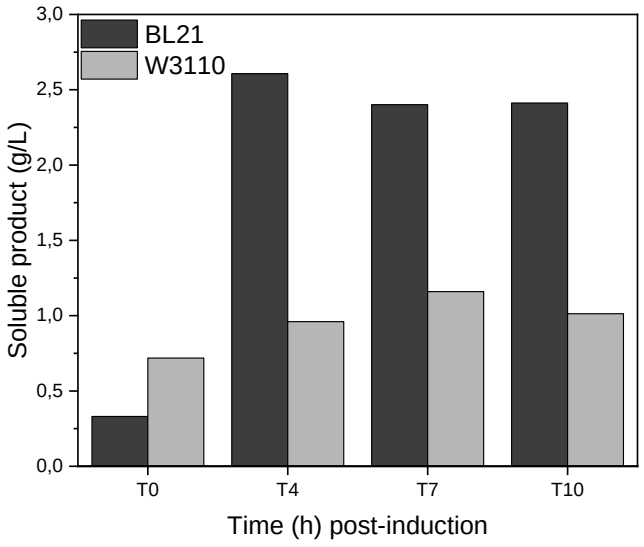

BL21
W3110
Soluble product (g/L)
0,0
0,5
1,0
1,5
2,0
2,5
3,0
T0
T4
T7
T10
Time (h) post-induction

FIGURE 2

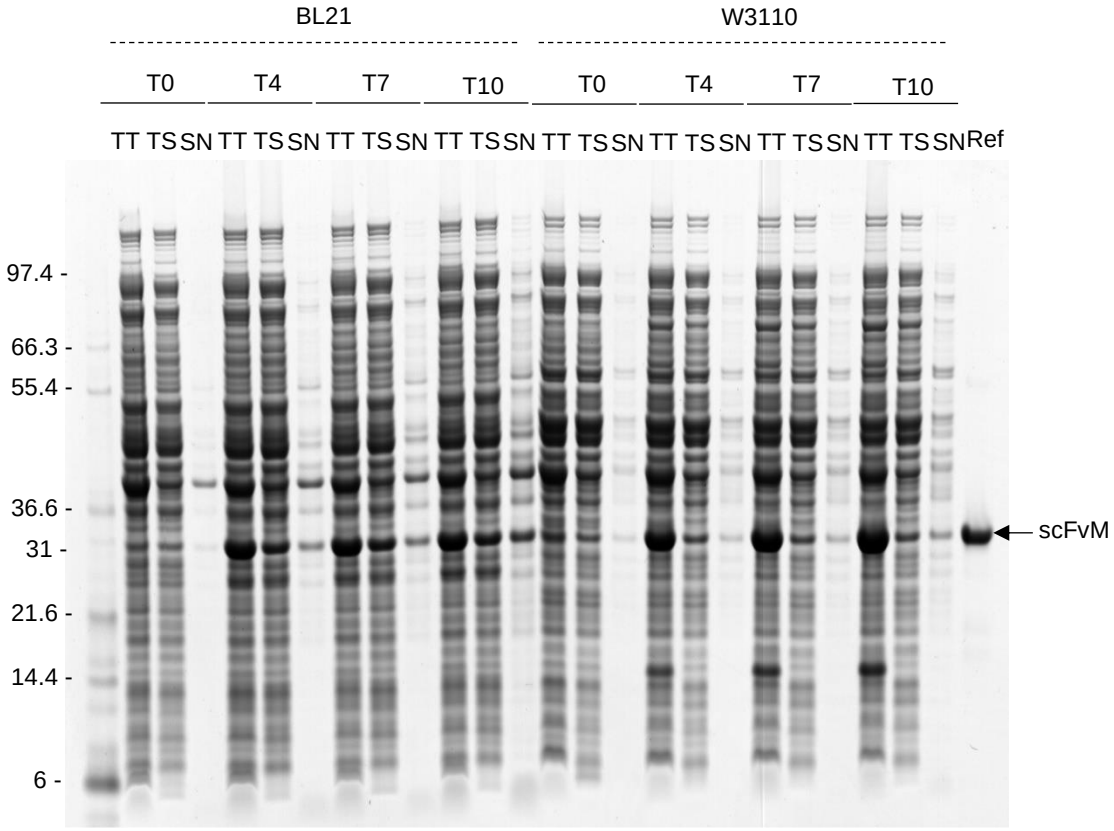

BL21
W3110
T0
T4
T7
T10
T0
T4
T7
T10
TT TS SN TT TS SN TT TS SN TT TS SN TT TS SN TT TS SN TT TS SN Ref
97.4 -
66.3 -
55.4 -
36.6 -
31 -
21.6 -
14.4 -
6 -
scFvM

FIGURE 3

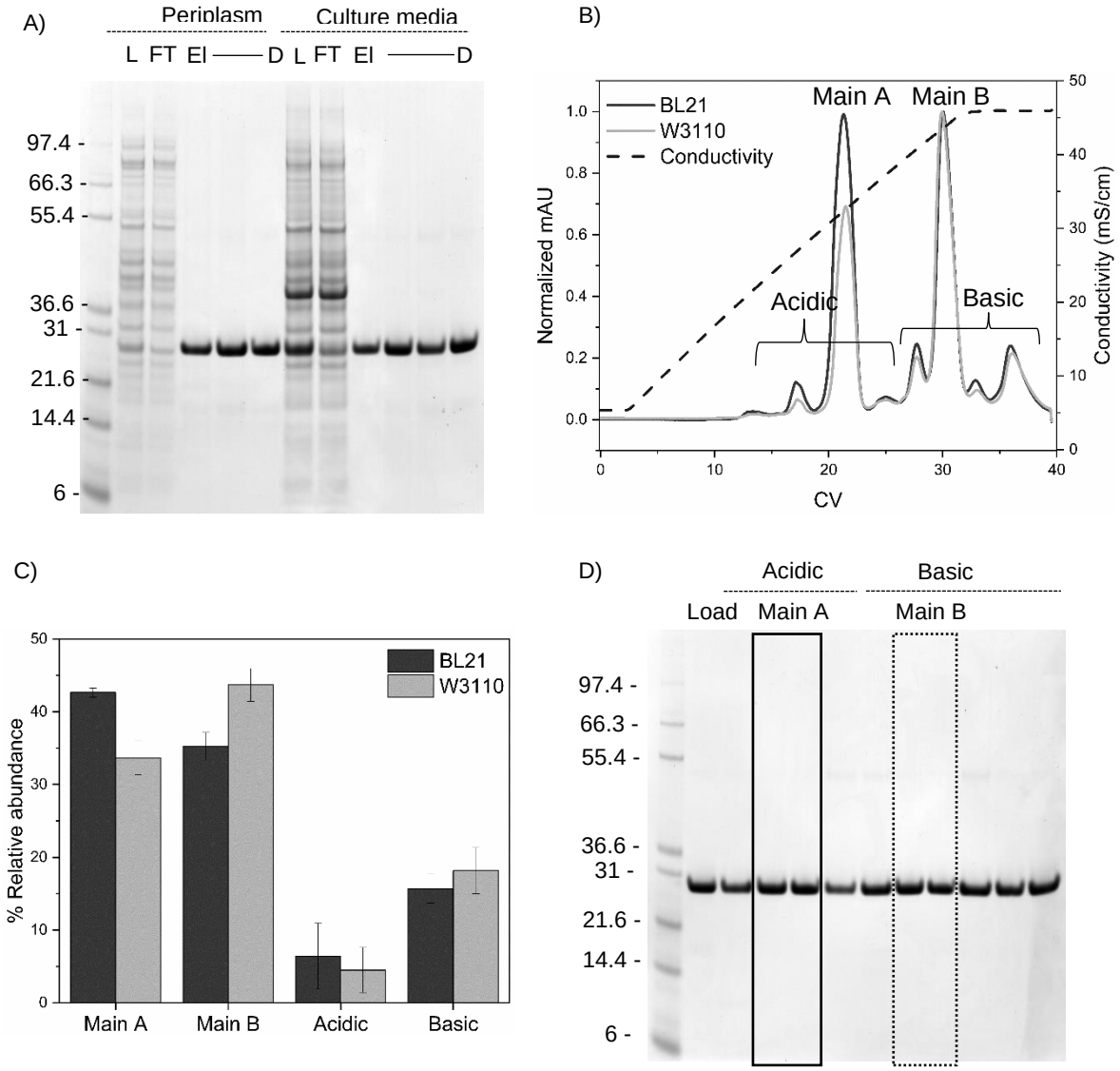

A)
Periplasm
Culture media
L FT El D L FT El D
97.4
66.3
55.4
36.6
31
21.6
14.4
6
B)
BL21
W3110
Conductivity
Main A
Main B
Acidic
Basic
Normalized mAU
Conductivity (mS/cm)
CV
C)
BL21
W3110
% Relative abundance
Main A
Main B
Acidic
Basic
D)
Acidic
Basic
Load
Main A
Main B
97.4
66.3
55.4
36.6
31
21.6
14.4
6

FIGURE 4

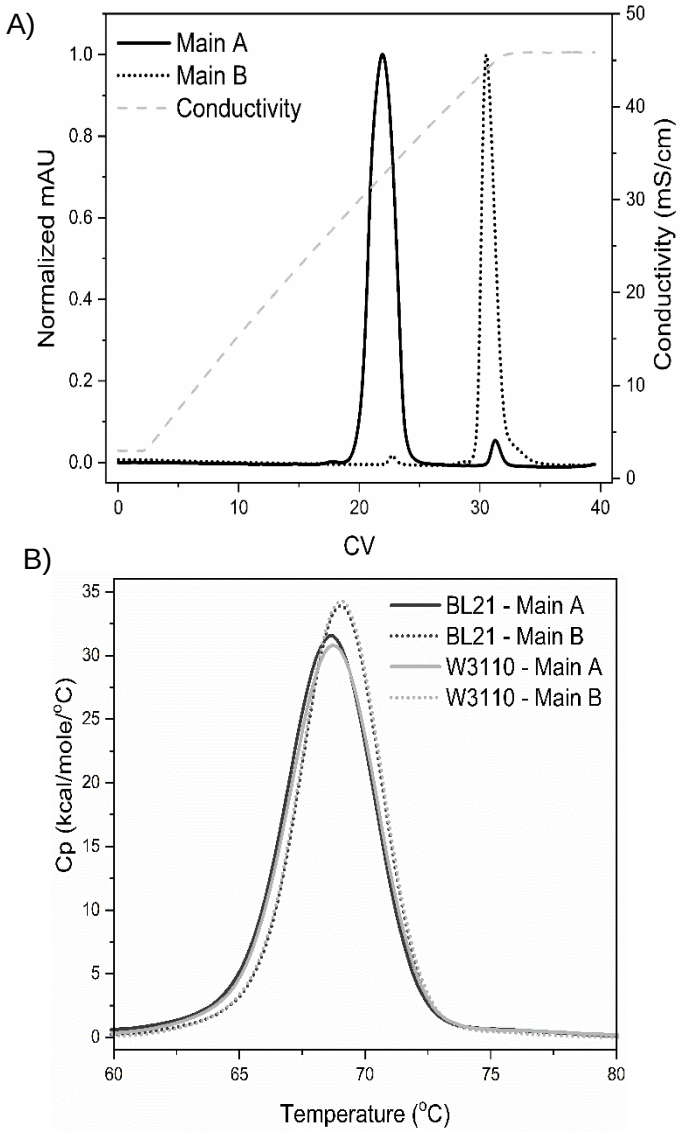
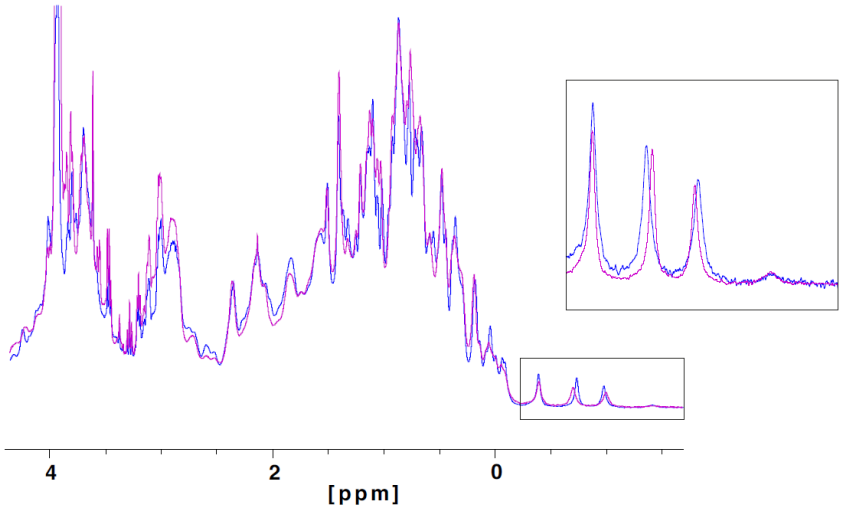
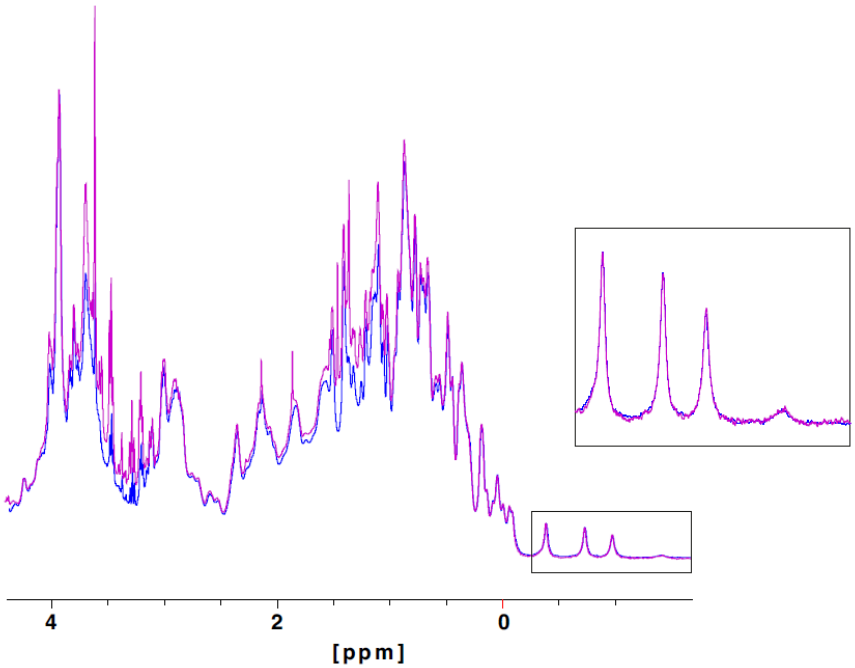

A)
Main A
Main B
Conductivity
Normalized mAU
Conductivity (mS/cm)
CV
B)
BL21 - Main A
BL21 - Main B
W3110 - Main A
W3110 - Main B
Cp (kcal/mole/°C)
Temperature (°C)

FIGURE 5

A)



B)



CHAPTER 4: First author publication. Highly efficient export of a disulphide bonded protein to the periplasm and medium by the Tat pathway using CyDisCo in *Escherichia coli*

4.1. Contribution

For this project, I heavily contributed to the conceptualization and experimental design. I also performed the experiments, analyzed the data, reviewed the literature available, made the figures and wrote the manuscript. The manuscript was revised by all the authors.

The objective of the project was to find the best Tat-dependent SP for a range of disulphide-bonded proteins. To investigate the same, I cloned different proteins and expressed them in 24 deep well plates (DWP) in the presence and absence of CyDisCo. The POIs were purified from the soluble lysate fraction through Immobilized Metal Affinity Chromatography (IMAC) for the comparison between +/- CyDisCo expression tests.

After finding the most desirable protein for our purpose, I scaled-up the expression from 24 DWP to shake-flask level. I fractionated the cells and recovered the media to find out whether this POI attached to the promiscuous SPs was actually being exported by the Tat pathway. This was verified by the comparison of the purified periplasmic fraction between BL21 wild-type and BL21 TatExpress performed in triplicates. These experiments were carried out at the University of Oulu (Finland). Furthermore, back at the University of Kent (UK), I performed the experiments with an *E. coli* ΔTat strain to confirm Tat-dependent export.

Finally, the DSB formation of YebF (our POI) was confirmed by mass spectrometry carried out by the Biocenter Oulu core facility, and data analysis and interpretation were carried out by Mirva J Saaranen and me.

4.2. Publication

The paper as was sent for publication in MicrobiologyOpen as follows:

Highly efficient export of a disulphide bonded protein to the periplasm and medium by the Tat pathway using CyDisCo in *Escherichia coli*

Klaudia Arauzo-Aguilera[1], Mirva J Saaranen[2], Colin Robinson[1] and Lloyd W Ruddock[2]*

[1]School of Biosciences, University of Kent, Canterbury CT2 7NJ, United Kingdom

[2]Faculty of Biochemistry and Molecular Medicine, University of Oulu, 90220 Oulu, Finland.

k.arauzo@kent.ac.uk (K.A.A.)

mirva.saaranen@oulu.fi (M.J.S.)

c.robinson-504@kent.ac.uk (C.R.)

lloyd.ruddock@oulu.fi (L.W.R)

* Correspondence:

Lloyd W. Ruddock, lloyd.ruddock@oulu.fi (L.W.R)

Abstract

High-value heterologous proteins produced in *Escherichia coli* that contain disulphide bonds are almost invariably targeted to the periplasm *via* the Sec pathway as it, among other advantages, enables disulphide bond formation and simplifies downstream processing. However, the Sec system cannot transport complex or rapidly folding proteins, as it only transports proteins in an unfolded state. The Tat system also transports proteins to the periplasm, and it has significant potential as an alternative means of recombinant protein production because it transports fully folded proteins. Most of the studies related to Tat secretion have used the well-studied Tat specific TorA signal peptide, but this signal peptide also tends to induce degradation of the protein of interest, resulting in lower yields. This makes it difficult to use Tat in industry. In this study, we show that YebF can be exported to the periplasm and media at a very high level by the Tat pathway in an almost completely CyDisCo-dependent manner by other two putative Tat SPs: those of MdoD and AmiC. The TorA SP, however, exports YebF at a lower level.

Keywords: Tat pathway, CyDisCo, disulphide bond, signal peptides, periplasm, *Escherichia coli*, TatExpress

Introduction

High-value heterologous proteins produced in *Escherichia coli* that contain disulphide bonds (DSB) are almost invariably targeted to the periplasm *via* the general secretory (Sec) pathway by means of a cleavable N-terminal signal peptide (SP) (Mirzadeh et al., 2020). This guides newly synthesised proteins through the SecYEG membrane channel in an unfolded state. Once across the membrane, the SP is cleaved and proteins fold in the periplasm, acquiring DSBs where appropriate (Kleiner-Grote et al., 2018). This protein export approach offers several advantages to produce therapeutic proteins, such as i) it enables disulphide bond formation, which only occurs in the periplasm in wild type cells, ii) it facilitates protein isolation from the relatively small periplasmic proteome, iii) allows control of the nature of the N-terminus of the mature protein and iv) minimizes exposure to cytoplasmic proteases (Karyolaimos & de Gier, 2021).

Some proteins, however, fold too rapidly for the Sec system to handle, or require co-factor insertion in the cytoplasm, thereby precluding translocation *via* the Sec system. The twin arginine

translocation (Tat) pathway offers a potential alternative method of localising proteins to the periplasm and, unlike Sec, this system transports fully folded proteins. As with Sec substrates, Tat substrates are synthesised with N-terminal SPs, but these contain Tat-specific determinants including the presence of a highly conserved twin-arginine motif (reviewed by Natale et al., 2008).

An additional, and indeed unique, feature of the Tat pathway is its in-built proofreading mechanism that can detect structurally incorrect substrates and reject them for export. This proofreading capability could allow for a more homogeneous product to be produced in the periplasm (reviewed by Frain et al., 2019). Furthermore, the recent development of TatExpress (TE) strains, that over-express the *tatABC* genes (encoding the Tat system) from the chromosome boosts the industrial potential of this pathway (Browning et al., 2017). However, this quality control poses problems for the export of disulphide-bonded proteins, because such proteins often only obtain a native conformation after the formation of DSB. CyDisCo strains (cytoplasmic disulphide bond formation in *E. coli*), express a catalyst of disulphide bond formation, usually the sulfhydryl oxidase Erv1p, and a catalyst of disulphide isomerization, usually human protein disulphide isomerase (PDI). This system facilitates DSB formation in the cytoplasm of wild-type *E. coli* and may therefore allow the efficient production of DSB-containing proteins in the cytoplasm prior to their export *via* the Tat pathway (Matos et al., 2014).

The efficiency of protein secretion varies depending on the host strain, signal sequence, and the type of protein to be secreted (Freudl, 2018). There are at least 29 SPs in the *E. coli* genome that contain a twin arginine motif characteristic of proteins exported *via* the Tat pathway. However, many of these SPs are not completely Tat-specific and can lead to secretion of the POI *via* Sec, Tat or both depending on the POI. They are sometimes termed promiscuous SPs (Tullman-Ercek et al., 2007; Bendtsen et al., 2005). In contrast, the TorA SP, an *E. coli* Tat SP derived from pre-trimethylamine *N*-oxide (TMAO) reductase (TorA), is a very well-studied Tat specific SP (Blaudeck et al., 2001; Lee, et al., 2006; Jack et al., 2004). However, the use of the TorA SP also tends to induce degradation of the protein of interest, for unknown reasons. This results in low yields of some POI, as seen in the absence of precursor forms in many export studies (e.g. Alanen et al., 2015).

While it has been shown that Tat can export some disulphide-bonded proteins at high levels in fed-batch fermentation studies (for example hGH; Guerrero Montero et al., 2019) and shake-flask studies (Alanen et al., 2015; Browning et al., 2017; DeLisa et al., 2003; Matos et al., 2014; Tullman-Ercek et al., 2007), it has yet to be determined whether it can efficiently export a protein that requires DSB formation for correct folding. In this study we used YebF, a 10.8 kDa *E.coli* protein of unknown function which contains a single DSB. It has been previously reported that recombinant YebF is secreted by laboratory strains of *E. coli* into the extracellular medium after first being translocated into the periplasm by the Sec-system (Zhang et al., 2006). A variety of proteins ranging from 15 to 48 kDa and more recently N-glycosylated protein domains have been shown to be readily secreted into the growth medium *via* fusion with YebF (Fisher et al., 2011). Here, we report that YebF can be exported to the periplasm and media by the Tat pathway in an almost completely CyDisCo-dependent manner. The use of TorA-SP results in low yields, but we show that two other Tat SPs, namely MdoD and AmiC, direct very high levels of protein export. While both are capable of directing export by Sec, expression in TatExpress strains results in a major increase in export flux which indicates that export is largely carried out by Tat.

Materials and Methods

All chemicals used in this study were supplied by Fisher Scientific (Thermo Fisher Scientific Inc, USA), Sigma (Sigma-Aldrich, USA) or Formedium (UK) unless otherwise stated.

Cloning

The desired gene fragments were obtained as synthetic genes from GeneArt or by PCR from *E.coli* genomic DNA and cloned with restriction digestion and ligation into a modified pET23-based vector with a pTac promoter replacing the T7 promoter (Gąciarz et al., 2016). The vector design allows for incorporation of C-terminal hexa-histidine tag (-Leu-Glu-6xHis) to the expressed protein. For YebF without the signal sequence, a vector incorporating both N- (Met-6xHis-) and C-terminal hexa-histidine tags was used. The gene inserts in the plasmids made were fully sequenced prior to use (see Table 1 for plasmid names and details).

Table 1. Strains and constructs used in this study.

Strain/Plasmid	**Description**	**Source/Reference**
BL21	*E. coli* B F^{-} *dcm ompT lon hsdS*($r_B^- m_B^-$) *gal*	Agilent
BL21 TatExpress	BL21 carrying a pTac promotor upstream of tatABCD	Browning et al., 2017
ΔTatABCDE (ΔTat)	MC4100 strain (*Ara^R, F2 araD139 DlacU169 rpsL150 relA1 flB5301 deoC1 ptsF25 rbs^R*) lacking tat*ABCDE* genes, *Ara^R*	Wexler et al., 2000
pMJS289	YebF (A22-R118) [1)]	This study
pMJS285	AmiC (M1-Q35)-YebF (A22-R118) [2)]	This study
pMJS284	MdoD (M1-D36)-YebF (A22-R118) [2)]	This study
pMJS288	TorA (M1-A43)-YebF (A22-R118) [2)]	This study
pMJS205	CyDisCo: Erv1p and PDI	Gaciarz et al., 2016
pAG82	empty	Gaciarz et al., 2016

[1)] with N-terminal Met-6xHis and C-terminal -Leu-Glu-6xHis -tags

[2)] with C-terminal -Leu-Glu-6xHis -tag

Expression

Plasmids with the gene of interest together with the plasmid containing the CyDisCo components (pMJS205) or empty plasmid (pAG82) were co-transformed into chemically competent *E. coli* cells and spread onto Lysogeny Broth (LB) agar plates supplemented with 35 µg/mL of chloramphenicol and 100 µg/mL of ampicillin for selection. After overnight incubation at 37°C these were used to inoculate 2-5 mL of LB media supplemented with 2 g/L of glucose, 35 µg/mL of chloramphenicol and 100 µg/mL of ampicillin. These starter cultures were grown 6 hours, or overnight for ΔTat experiments, at 30°C ◦C, 250 rpm (2.5 cm radius of gyration) and were used to seed the cultures in 1:100 ratio.

Expression tests to screen for optimal SP were carried out for the constructs in 24 deep well plates (DWP). The constructs were expressed alone or co-expressed with CyDisCo components in *E.*

coli BL21 in 3 mL per well of terrific broth autoinduction media (AIM – Terrific Broth Base including Trace elements, Formedium) supplemented with 0.8% glycerol, 35 µg/mL of chloramphenicol and 100 µg/mL of ampicillin. The DWP was covered with air permeable membrane (Thomson) and the cultures were grown at 30°C, 250 rpm, and harvested after 24 hours. The cells were collected by centrifugation at 3,220 × g for 20 minutes at 4°C and resuspended in 3 mL of 50 mM sodium phosphate pH 7.4, 20 µg/ml DNase, 0.1 mg/ml egg white lysozyme. After 10 min incubation the resuspended cultures were frozen at - 20°C. Cells were lysed by freeze-thawing.

Main cultures with selected constructs in either BL21 or BL21 TatExpress cells were grown in 100 mL flasks with 10 mL culture in each flask. The flasks were covered with oxygen permeable membranes (AirOtop, Thomson) to ensure proper oxygenation of the cultures, grown at 30°C, 250 rpm for 24 hours and harvested for fractionation.

For ΔTat experiments, cultures were grown in LB media at 30°C, 250 rpm until the OD_{600} of the cultures reached approximately 0.5 and were then induced with 50 µM IPTG for 2 hours before harvesting.

Fractionation of the cells

For the fractionation of the cells, PureFrac fractionation protocol was used (Malherbe et al., 2019). For purification, EDTA was not added in any of the buffers. Apart from the periplasm and cytoplasm, medium samples were also recovered (same volume in all cultures). In ΔTat experiments, the 1x PBS wash of the cells and the separation of the cytoplasm and insoluble fraction was not carried out to avoid extra manipulation of this cell line due to its fragility. Samples were prepared for SDS-PAGE analysis in reducing conditions.

Purification of cytoplasmic, periplasmic and medium samples and SDS-PAGE analysis

Purification of hexa-histidine tagged proteins was performed by standard immobilized metal affinity chromatography using HisPur Cobalt resin (Thermo Scientific) under native conditions. For 3 mL cultures from 24 DWP, IMAC was performed using 0.2 mL resin in small gravity feed columns. The resin was washed with 2 x 2 mL of water and equilibrated with 2 x 2 mL of 50 mM

phosphate buffer (pH 7.4). Cell lysates on 24 DWP were cleared by centrifugation (3,220 x g, 20 min, 4 °C) and loaded on the columns. The columns were rinsed with 2 mL of 50 mM phosphate buffer (pH 7.4), washed with 4 x 2 mL of wash buffer (50 mM sodium phosphate, 10 mM imidazole, 300 mM sodium chloride; pH 7.4) and then rinsed with 2 mL of 50 mM sodium phosphate (pH 7.4) before elution with 3 x 0.2 mL of 50 mM sodium phosphate, 50 mM EDTA (pH 7.4). For 10 mL cultures the same protocol was used with the following changes: medium samples were 1:2 diluted (total volume 10 mL), periplasmic and cytoplasmic fractions were diluted in 2.5 mL of 200 mM sodium phosphate buffer and made up to 10ml with water to reduce the salt concentration. Samples were prepared for SDS-PAGE analysis and 10 µL were loaded in 4–20% Criterion™ TGX™ Precast Midi Protein Gel, 26 well (BioRad).

Mass Spectrometry

The theoretical oxidized monoisotopic molecular weight ($M_{theorOx}$) of the His-tag YebF constructs in dalton (Da) were calculated using ExPaSy Compute pI/Mw -tool (Gasteiger et al., 2005) (Table 2). The molecular weights of purified protein samples were measured by electrospray ionization mass spectrometry combined with liquid chromatography (LC-ESI-MS) using a Q Exactive Plus Mass Spectrometer. The protein samples were mixed with trifluoroacetic acid (TFA) to a final concentration of 0.5% prior to analysis. For N-ethylmaleimide (NEM)-trapped samples the protein was incubated with 20 mM NEM in 50 mM phosphate buffer pH 7.3 with 6 M guanidine-HCl for 10 minutes and quenched with 0.5% TFA prior to analysis.

Results

Folding of YebF is CyDisCo dependent without its native SP and with the AmiC, MdoD and TorA SPs

In this study, the first aim was to demonstrate the potential use of other Tat SPs in place of the well-known but non-ideal TorA SP (Alanen et al., 2015). We also sought to test whether a CyDisCo-dependent protein could be exported at high rates by the Tat system; *E. coli* TatExpress cells have been shown to export high levels of human growth hormone (hGH) but while this

protein contains 2 disulfide bonds, they are not essential for proper folding and the Tat system can efficiently export the protein in its reduced state (Alanen *et al.*, 2015 and references therein). Tat can only transport fully folded proteins from the cytoplasm to the periplasm, so we hypothesized that a disulphide containing protein that required disulphide formation to reach a native state e.g. one that was dependent on CyDisCo in our system, would be a good model protein to test the true capabilities of the Tat system. Based on its reported use as a transport mechanism for secretion to the medium, we initially chose a small disulphide bonded *E. coli* protein as a test protein – YebF.

In the absence of a SP (Figure 1, 'YebF no SP'), the folding of YebF showed a strong CyDisCo dependence. In this experiment, YebF was expressed in the presence and absence of CyDisCo and the protein was then purified by IMAC. The gel shows that soluble (and therefore presumably folded) YebF is far more abundant in the + CyDisCo cells, strongly indicating that CyDisCo improves folding to a significant extent. Some soluble protein is, however, present in the absence of CyDisCo indicating that the protein can fold to a low extent without CyDisCo. YebF was then tested with the widely used TorA Tat-dependent SP (construct denoted TorA-YebF) along with AmiC and MdoD SPs as alternative Tat SPs (referred to as AmiC-YebF and MdoD-YebF, respectively). In all cases YebF was expressed with a 6xHis tag on the C-terminus.

For both the AmiC-YebF and MdoD YebF constructs there was a strong dependency on CyDisCo to produce soluble protein (Figure 1), indicating that the protein is probably secreted *via* the Tat pathway rather than *via* the Sec pathway. If the proteins were exported by Sec, they would be unfolded until they reached the periplasm where they would rapidly acquire their DSBs, and CyDisCo would not influence their folding. An extra band in the gel (labelled SP-YebF) can be seen, which may be uncleaved SP-YebF (the POI with the SP still attached). In contrast to the results with AmiC-YebF and MdoD-YebF, the TorA-YebF construct produces very low levels of soluble protein in the presence of CyDisCo and no visible protein in the absence of CyDisCo. This may reflect inefficient transport by Tat, with the protein being degraded in the cytoplasm as observed with other constructs bearing the TorA signal peptide (Alanen et al., 2015).

TatExpress cells export much more YebF to the periplasm than standard BL21 cells with AmiC and MdoD SPs

We next assessed the export of the four constructs by fractionating cell samples into cytoplasm (C), periplasm (P) and medium (M) to discover where and how in the cell or medium this protein was targeted in the presence of CyDisCo. The experiments were carried out in both a standard BL21 strain and the BL21 TatExpress strain which has been engineered for higher levels of Tat-dependent export (Browning et al., 2017). We reasoned that if the AmiC-YebF and MdoD-YebF constructs were exported primarily by Tat, we would observe higher levels of export in the TatExpres cells, as observed with hGH (Browning et al., 2017; Guerrero Montero et al., 2019).

The control experiment shows that YebF, when expressed without a SP, remains in the cytoplasm when expressed in both BL21 and BL21 TatExpress (Figure 2; lanes 'C' denote cytoplasmic fraction; 'X' denotes TatExpress cells in this and subsequent Figures). No YebF is visible in the periplasm (Figure 2 lanes P, PX). IMAC purification of the protein from fractions confirmed the cytoplasmic localization (lanes denoted “Purification”), with a small amount of YebF being purified from the medium, most likely due to cell lysis.

Since YebF without a signal sequence could be efficiently folded by CyDisCo but was retained in the cytoplasm, we then examined if YebF could be exported to the periplasm and, if so, whether this export was increased in TatExpress cells when using the AmiC-, MdoD- and TorA SPs (Figure 3). The constructs were expressed in BL21 and BL21 TatExpress and the cells fractionated into cytoplasm, periplasm and medium (C, P, M) with YebF purified by IMAC in the 'Purification' panels.

In both constructs and strains, the periplasmic fraction is relatively clean of cytoplasm cross-contaminants as judged by the low levels or absence of major cytoplasmic proteins. Importantly, YebF is the most abundant periplasmic protein after expression of AmiC-YebF and MdoD-YebF, even in standard BL21 cells, and its abundance increases dramatically in TatExpress cells (Figure 3A, B lanes PX). IMAC purification of YebF confirms that export of both proteins is particularly efficient in TatExpress cells. Purification from the cytoplasmic fractions shows a very faint duplex band for AmiC-YebF and a slightly more prominent duplex for the MdoD-YebF construct, which

probably represent the SP- and mature forms of YebF. In purifications from periplasm fractions, only the mature form is seen for both constructs.

In order to compare the efficiency of the export of these two new proposed Tat specific SPs to the well characterized TorA SP, TorA-YebF was also expressed in BL21 and BL21 TatExpress and the cells fractionated into cytoplasm, periplasm and medium (C, P, M) with YebF again purified by IMAC in the 'Purification' panels. YebF is also exported to the periplasm (Figure 4) but in this case the export efficiencies are fairly similar in BL21 and BL21 TatExpress cells. Again, however, it is notable that the mature YebF protein is the most abundant protein in the periplasm, indicating efficient Tat-dependent export, though it is considerably less than for the MdoD- or AmiC-Sp constructs (Figure 1 and compare Figure 3 vs Figure 4). Since the TorA SP is highly Tat-specific (Tullman-Ercek et al., 2007) this export must be by the Tat pathway, which in turn means that the disulphide-bonded protein is being exported.

Together, the results indicate a very efficient export by Tat pathway to the periplasm, which is higher for AmiC-YebF and MdoD-YebF compared with TorA-YebF. In addition, purified yields of YebF from the medium fractions are high. This may come partially from lysis, but probably arises mainly from the translocation of YebF from the periplasm to the medium by an unknown mechanism (Zhang et al., 2006).

The MdoD SP is the the most efficient Tat SP for export of YebF

The CyDisCo dependency for AmiC-, MdoD- and TorA-YebF folding and export (Figure 1) and the enhancement of export to the periplasm in TatExpress cells (Figure 3), strongly indicates that the predominant pathway used for all three constructs must be Tat, as Sec will only transport proteins in an unfolded state.

To examine this further, we expressed our constructs in a strain (ΔTat) that lacks the *tatABCDE* genes that encode the Tat apparatus. In this strain any periplasmic export must be *via* Sec. ΔTat is a relatively fragile strain that tends to lyse more easily than wild-type strains in many growth conditions, and it is thus difficult to subfractionate (Harrison et al., 2005; Sargent et al., 1998).

ΔTat cells expressing YebF constructs with and without CyDisCo co-expression were fractionated into periplasm (P) and spheroplast (Sph), and medium samples (M) were also collected to monitor lysis and possible export. Unlike in previous experiments in BL21 and BL21 TatExpress, periplasmic samples were contaminated with cytoplasm, due to the fragility of this engineered strain. This can be observed as a periplasmic protein profile that is more similar to the cytoplasmic protein fraction, in contrast to the results in previous fractionations (Figure 3). However, medium samples suggest very low levels lysis in these growth conditions (Figure 5).

When YebF was expressed without a SP we observed no clear band in the periplasm, as expected. A possible band representing YebF is found in the spheroplast fraction in the presence of CyDisCo (arrowed), but we were not able to purify the protein in significant amounts and it appears that it may be susceptible to degradation in the cytoplasm of ΔTat cells. AmiC-YebF is clearly exported to the periplasm in both the presence and absence of CyDisCo; a prominent band of the correct size is detected in these fractions but not when YebF is expressed without an SP. The protein is presumably exported by the Sec pathway, which is consistent with the known ability of the AmiC SP to function as a Sec SP (Tullman-Ercek et al., 2007). The band is significantly stronger in the absence of CyDisCo.

These results suggest that some Sec dependent secretion occurs in the ΔTat strain for both AmiC and MdoD SPs and this is inhibited when CyDisCo is present - presumably because YebF folds to a native state in the cytoplasm which blocks export of these molecules by Sec.

Sec dependent secretion is much lower for MdoD-YebF than for AmiC-YebF, and we again observe that export is much more efficient in the absence of CyDisCo, which reinforces the suggestion that CyDisCo is able to fold the protein before it can be exported by Sec. In the ΔTat strain no protein could be observed in any fraction when TorA-YebF is expressed, demonstrating, once more, the degradation of a heterologous protein when the TorA SP is attached.

YebF control and SP-YebF constructs are folded and correctly cleaved from SPs in the periplasm and media

A final experiment to verify the folding and the correct cleavage of the SPs of the constructs in the periplasm and medium was carried out by subjecting the purified proteins to ESI-Mass Spectrometry. The theoretical monoisotopic molecular weight for oxidized ($M_{theorOx}$) YebF with no SP purified from the cytoplasm is 12937 Da and we observe an experimental molecular weight (M_{exp}) of 12985 Da, which shows a difference of 48 Da (see Table 2). The $M_{theorOx}$ of AmiC-YebF purified from the periplasm and medium is 12506 Da and we observed an M_{exp} of 12538 Da, which shows a difference of 32 Da. For MdoD-YebF purified from the periplasm and medium the $M_{theorOx}$ is the 12509 Da and we observed an M_{exp} of 12541 Da, which also shows a difference of 32 Da. For TorA-YebF purified from the periplasm and medium the $M_{theorOx}$ is the 12320 Da and we observed an M_{exp} of 12352 Da, which also shows a difference of 32 Da. Overall, analysis by ESI-MS confirmed that YebF purified from the cytoplasm (in the YebF no SP construct only), periplasm and medium had the expected molecular weight, consistent with the cleavage of the respective SP when the POI is exported from the cytoplasm to the periplasm and its two cysteines form of a disulphide bond. The presence of free cysteines in the constructs was further evaluated by treating the samples with NEM prior to mass spectrometric analysis. NEM-trapped samples would be expected to show an addition of 125 Da in the molecular weight of the protein for each free cysteine modified. None of the samples analysed showed any increase in the mass after NEM-treatment, implying that none contained free thiols (Saaranen et al., 2010; Nguyen et al., 2011). The change of mass of 48 Da (for YebF no SP) and 32 Da for the other three constructs, are due to the oxidation of methionine in the purified proteins (three and two oxygen molecules respectively). The handling, time of storage and/or repeatedly freeze-thawed, are responsible of this phenomenon (Grassi & Cabrele, 2019) (Table 2).

Table 2. Theoretical oxidized and experimental molecular weights of the proteins in this study. The theoretical monoisotopic molecular weight for oxidized ($M_{theorOx}$) His-tag YebF constructs in dalton (Da) were calculated using ExPaSy Compute pI/Mw tool (Gasteiger et al., 2005). The experimental molecular weight (M_{exp}) was determined by mass spectrometry. The same masses were obtained with NEM-treatment. The results suggest both cysteines are in a disulphide bond in the YebF constructs analysed.

Construct	Location	Number of Cysteine	$M_{theorOx}$ (Da)	M_{exp} (Da)	Δ mass
YebF no SP	Cytoplasm	2	12937	12985	48
AmiC-YebF	Periplasm		12506	12538	32
	Medium		12506	12538	32
MdoD-YebF	Periplasm		12509	12541	32
	Medium		12509	12541	32
TorA-YebF	Periplasm		12320	12352	32
	Medium		12320	12352	32

Discussion

E. coli production platforms are extensively used for the production of biotherapeutics, but current platforms have limitations in terms of the types of protein that they can handle. Typically, target proteins are refolded from inclusion bodies or targeted to the periplasm where they fold to a native state, the latter being the less time-consuming and labour-intensive approach (Bhatwa et al., 2021). Tat-based platforms offer significant advantages for the production of some molecules, but the system has not been validated using a wide range of proteins, especially proteins that require disulphide bonding to fold correctly, and there is a clear need to find an alternative to TorA as the commonly used Tat specific SP. The vast majority of export studies by the Tat pathway have been carried out using the TorA SP although it is known that it can cause POI degradation in the cytoplasm (Blaudeck et al., 2003) and inclusion body formation (Jong et al., 2017). The TorA SP has been shown to export a small number of proteins with high efficiency, but none of these examples required disulphide bonding to occur in the cytoplasm prior to translocation by Tat (Alanen et al., 2015).

Here, we set out to compare the export of YebF to the periplasm of *E. coli via* the Tat pathway with different SPs: the AmiC and MdoD SPs have been reported to be able to go through both the Sec and Tat pathway (Tullman-Ercek et al., 2007) whereas the TorA SP is reported to be Tat specific.

We examined the export to the periplasm (and medium) of YebF with different SPs in the presence and absence of CyDisCo which is required for YebF to efficiently reach a native state in the cytoplasm. In wild-type cells, YebF was exported to the periplasm and medium by the classical Tat SP TorA, but yields were relatively low – as is often reported for this SP. The use of TatExpress cells increased export to the periplasm, while no YebF could be observed in a ΔTat strain, as expected with an SP that cannot target proteins through the Sec pathway.

Generally, the results for AmiC-SP and MdoD-SP mirrored those of the TorA-SP, with two exceptions. Firstly, the periplasmic and medium yields of YebF in wild-type and TatExpress cells were far higher than for TorA i.e. they appear to be much more efficient SPs. Secondly, in ΔTat cells a periplasmic localization was seen for a small fraction of the total YebF, implying Sec-dependent export. This fraction decreased for both SPs when CyDisCo was co-expressed, as the protein could be folded to a native state in the cytoplasm and kinetic competition developed between CyDisCo-mediated (preventing export by Sec) and Sec-dependent export of the unfolded protein. These results imply that while Tat is the normal secretion pathway for the AmiC and MdoD SPs, Sec dependent secretion can occur when two conditions are met i) the Tat pathway is knocked out; ii) the attached protein is in an unfolded state. With the MdoD SP, even when both conditions are met Sec based secretion is very inefficient.

Experiments with the ΔTat mutant have limitations primarily due to the susceptibility of this strain to lysis and the difficulty of obtaining clean fractionation. Due to this the experiments were not carried out in autoinduction media with longer induction times, but in LB media with minimal induction times. As CyDisCo is co-expressed with the YebF constructs the initial levels of CyDisCo may not be optimal for folding YebF to the native state and hence the results obtained probably include a bias towards Sec secretion. The use of two different promoters, one for CyDisCo plasmid and another one for YebF, allowing pre-expression of CyDisCo would obviate this, if such conditions could be tolerated by the ΔTat strain.

The choice of the POI for this study was not arbitrary. Apart from the desired CyDisCo dependency of the protein of choice, YebF is an intriguing POI. YebF with its native SP is used as a “passenger” protein linker to export transgenic proteins to the medium by an unknown

mechanism. This new discovery gives the possibility of linking more disulphide-bonded difficult-to-express proteins to these two efficient and Tat dependent SPs for easy recovery in the extracellular medium, assuring the correct folding with CyDisCo of the target protein and a maximised yield when using TatExpress cells.

In conclusion, the AmiC and MdoD SPs appear to allow efficient secretion of a disulphide bond containing protein from the cytoplasm of *E. coli via* Tat. While these signal sequences are not completely Tat specific, they appear to result in secretion *via* Sec only when the Tat pathway is compromised, and the POI remains in an unfolded state. As they are far more efficient than the TorA SP, they are probably more suitable for large scale protein production than the more rigorously specific Tat-signal peptide.

Data availability statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions: Conceptualization L.W.R.; Methodology M.J.S., L.W.R.; Investigation and formal analysis K.A.A., M.J.S.; Visualization K.A.A.; Writing original draft preparation K.A.A.; Writing review and editing: M.J.S., L.W.R., C.R.; Supervision M.J.S., L.W.R., C.R.; Funding acquisition C.R., L.W.R. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: This work was funded by the People Programme (Marie Skłodowska-Curie Actions) of the European Union's Horizon 2020 Programme under REA grant agreement no. 813979 (SECRETERS). The use of the facilities of Biocenter Oulu core facilities, a member of Biocenter Finland, is gratefully acknowledged.

Declaration of interests: A patent for the production system used to make the protein using sulfhydryl oxidases in the cytoplasm of *E. coli* is held by the University of Oulu: Method for producing natively folded proteins in a prokaryotic host (Patent number 9238817; date of patent January 19th, 2016). Inventor: Lloyd W. Ruddock.

Ethics approval statement: Not applicable

Patient consent statement: Not applicable

Permission to reproduce material from other sources: Not applicable

Clinical trial registration: Not applicable

References

Alanen, H. I., Walker, K. L., Lourdes Velez Suberbie, M., Matos, C. F. R. O., Bönisch, S., Freedman, R. B., Keshavarz-Moore, E., Ruddock, L. W., & Robinson, C. (2015). 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. *Biochimica et Biophysica Acta, 1853*(3), 756–763. https://doi.org/10.1016/J.BBAMCR.2014.12.027

Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T., & Brunak, S. (2005). Prediction of twinarginine signal peptides. *BMC Bioinformatics*, *6*(1), 1–9. https://doi.org/10.1186/1471-2105-6-167

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X. Z., & Zhou, T. (2021). Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. *Frontiers in Bioengineering and Biotechnology*, *9*. https://doi.org/10.3389/FBIOE.2021.630551

Blaudeck, N., Sprenger, G. A., Freudl, R., & Wiegert, T. (2001). Specificity of signal peptide recognition in Tat-dependent bacterial protein translocation. *Journal of bacteriology*, *183*(2), 604-610. https://doi.org/10.1128/JB.183.2.604-610.2001

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. *Journal of Bacteriology*, *185*(9), 2811–2819. https://doi.org/10.1128/JB.185.9.2811-2819.2003

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. *Biotechnology and Bioengineering*, *114*(12), 2828–2836. https://doi.org/10.1002/bit.26434

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), 6115–6120. https://doi.org/10.1073/PNAS.0937838100

Fisher, A. C., Haitjema, C. H., Guarino, C., Çelik, E., Endicott, C. E., Reading, C. A., Merritt, J. H., Ptak, A. C., Zhang, S., & DeLisa, M. P. (2011). Production of Secretory and Extracellular N-Linked Glycoproteins in Escherichia coli. *Applied and Environmental Microbiology*, *77*(3), 871. https://doi.org/10.1128/AEM.01901-10

Frain, K. M., Robinson, C., & van Dijl, J. M. (2019). Transport of Folded Proteins by the Tat System. *The Protein Journal 2019 38:4*, *38*(4), 377–388. https://doi.org/10.1007/S10930-019-09859-Y

Freudl, R. (2018). Signal peptides for recombinant protein secretion in bacterial expression systems. *Microbial Cell Factories*, *17*(1), 1–10. https://doi.org/10.1186/S12934-018-0901-3

Gaciarz, A., Veijola, J., Uchida, Y., Saaranen, M. J., Wang, C., Hörkkö, S., & Ruddock, L. W. (2016). Systematic screening of soluble expression of antibody fragments in the cytoplasm of E. coli. *Microbial Cell Factories*, *15*(1), 22. https://doi.org/10.1186/S12934-016-0419-5

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. *The Proteomics Protocols Handbook.* Humana Press; Totowa, NJ, USA: 2005. Protein Identification and Analysis Tools on the ExPASy Server; pp. 571–607. https://doi.org/10.1385/1-59259-890-0:571

Grassi, L., & Cabrele, C. (2019). Susceptibility of protein therapeutics to spontaneous chemical modifications by oxidation, cyclization, and elimination reactions. *Amino Acids*, *51*, 1409–1431. https://doi.org/10.1007/s00726-019-02787-2

Guerrero Montero, I., Richards, K. L., Jawara, C., Browning, D. F., Peswani, A. R., Labrit, M., Allen, M., Aubry, C., Davé, E., Humphreys, D. P., Busby, S. J. W., & Robinson, C. (2019). Escherichia coli "TatExpress" strains export several g/L human growth hormone to the periplasm by the Tat pathway. *Biotechnology and Bioengineering*, *116*(12), 3282–3291. https://doi.org/10.1002/BIT.27147

Harrison, J. J., Ceri, H., Badry, E. A., Roper, N. J., Tomlin, K. L., & Turner, R. J. (2005). Effects of the twin-arginine translocase on the structure and antimicrobial susceptibility of Escherichia coli biofilms. *Canadian Journal of Microbiology*, *51*(8), 671–683. https://doi.org/10.1139/W05-048

Jack, R. L., Buchanan, G., Dubini, A., Hatzixanthis, K., Palmer, T., & Sargent, F. (2004). Coordinating assembly and export of complex bacterial proteins. *The EMBO Journal*, *23*(20), 3962–3972. https://doi.org/10.1038/SJ.EMBOJ.7600409

Jong, W. S. P., Vikström, D., Houben, D., Berg van Saparoea, H. B., Gier, J. W., & Luirink, J. (2017). Application of an E. coli signal sequence as a versatile inclusion body tag. *Microbial Cell Factories*, *16*(1), 1–13. https://doi.org/10.1186/S12934-017-0662-4

Karyolaimos, A., & de Gier, J. W. (2021). Strategies to Enhance Periplasmic Recombinant Protein Production Yields in Escherichia coli. *Frontiers in Bioengineering and Biotechnology*, *9*. https://doi.org/10.3389/FBIOE.2021.797334

Kleiner-Grote, G. R. M., Risse, J. M., & Friehs, K. (2018). Secretion of recombinant proteins from E. coli. *Engineering in Life Sciences*, *18*(8), 532. https://doi.org/10.1002/ELSC.201700200

Lee, P. A., Tullman-Ercek, D., & Georgiou, G. (2006). The bacterial twin-arginine translocation pathway. *Annual Review of Microbiology*, *60*, 373–395. https://doi.org/10.1146/ANNUREV.MICRO.60.080805.142212

Malherbe, G., Humphreys, D. P., & Davé, E. (2019). A robust fractionation method for protein subcellular localization studies in Escherichia coli. *BioTechniques*, *66*(4), 171–178. https://doi.org/10.2144/BTN-2018-0135

Matos, C. F. R. O., Robinson, C., Alanen, H. I., Prus, P., Uchida, Y., Ruddock, L. W., Freedman, R. B., & Keshavarz-Moore, E. (2014). Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in Escherichia coli CyDisCo strains. *Biotechnology Progress*, *30*(2), 281–290. https://doi.org/10.1002/btpr.1858

Mirzadeh, K., Shilling, P. J., Elfageih, R., Cumming, A. J., Cui, H. L., Rennig, M., Nørholm, M. H. H., & Daley, D. O. (2020). Increased production of periplasmic proteins in Escherichia coli by directed evolution of the translation initiation region. *Microbial Cell Factories*, *19*(1), 1–12. https://doi.org/10.1186/S12934-020-01339-8

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*, *1778*(9), 1735–1756. https://doi.org/10.1016/J.BBAMEM.2007.07.015

Nguyen, V. D., Saaranen, M. J., Karala, A. R., Lappi, A. K., Wang, L., Raykhel, I. B., Alanen, H. I., Salo, K. E. H., Wang, C. C., & Ruddock, L. W. (2011). Two endoplasmic reticulum PDI peroxidases increase the efficiency of the use of peroxide during disulfide bond formation. *Journal of Molecular Biology*, *406*(3), 503–515. https://doi.org/10.1016/J.JMB.2010.12.039

Saaranen, M. J., Karala, A. R., Lappi, A. K., & Ruddock, L. W. (2010). The role of dehydroascorbate in disulfide bond formation. *Antioxidants & Redox Signaling*, *12*(1), 15–25. https://doi.org/10.1089/ARS.2009.2674

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. *The EMBO Journal*, *17*(13), 3640. https://doi.org/10.1093/EMBOJ/17.13.3640

Tullman-Ercek, D., DeLisa, M. P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., & Georgiou, G. (2007). Export Pathway Selectivity of Escherichia coli Twin Arginine Translocation Signal Peptides. *Journal of Biological Chemistry*, *282*(11), 8309–8316. https://doi.org/10.1074/JBC.M610507200

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. *Journal of Biological Chemistry*, *275*(22), 16717–16722. https://doi.org/10.1074/JBC.M000800200

Zhang, G., Brokx, S., & Weiner, J. H. (2006). Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in Escherichia coli. *Nature Biotechnology*, *24*(1), 100–104. https://doi.org/10.1038/NBT1174

Figure legends

Figure 1. **Purification of soluble YebF protein without its native SP and with AmiC, MdoD and TorA SPs in the absence/presence of CyDisCo**. Coomassie- stained Criterion™ TGX gels showing the purified soluble YebF (≈12 kDa) in BL21 wild-type in 2 mL rich-autoinduction media at 30°C at 24DWP. All the constructs show a CyDisCo dependency for the production of YebF soluble protein (marked as "-YebF"). MdoD-YebF construct shows uncleaved YebF (shown as "SP- YebF") as well as mature protein (marked as "YebF") in the presence of CyDisCo.

Figure 2. **Expression of YebF control without SP in BL21 and BL21 TatExpress cells with CyDisCo.** Coomassie-stained Criterion™ TGX gel of cytoplasmic, periplasmic fractions and medium samples and purifications (C, P, M) from BL21 and BL21 TatExpress cells (CX, PX, MX) expressing YebF without SP (marked as "YebF"). Samples were collected after 24 hours of growth in terrific broth based autoinduction media at 30°C in shake flasks and processed immediately. Samples from the same subcellular fraction are comparable between strains, but not between different subcellular compartments. Cytoplasm was diluted in 750 µL buffer. Periplasm was diluted in 400 µL buffer. For all medium samples, 5 mL culture was recovered. Purifications of all cell fractions and medium are comparable among fractions and among strains. Some POI is visible in medium purification probably due to cell lysis. Representative gel from triplicate experiments is shown.

Figure 3. Export of AmiC- and MdoD- YebF in BL21 and BL21 TatExpress cells with CyDisCo. Coomassie-stained Criterion™ TGX gel of cytoplasmic, periplasmic fractions and medium samples and purifications (C, P, M) from BL21 and BL21 TatExpress cells (CX, PX, MX) expressing AmiC-YebF, MdoD-YebF and TorA-YebF (marked as "-YebF" and "SP-YebF"). Samples were analysed after 24 hours of growth in terrific broth based autoinduction media at 30°C in shake flasks. **A:** AmiC-YebF fractionation and purification of BL21 and BL21 TatExpress cells for comparison. **B:** MdoD-YebF fractionation and purification of BL21 and BL21 TatExpress for comparison. Samples from the same subcellular fraction are comparable between strains, but not between different subcellular compartments. Cytoplasm was diluted in 750 µL buffer. Periplasm was diluted in 400 µL buffer. Purifications of all cell fractions and medium are comparable among fractions and among strains. For all medium samples, 5 mL culture was recovered. Representative gel from triplicate experiments is shown.

Figure 4. Export of TorA- YebF in BL21 and BL21 TatExpress cells with CyDisCo. Coomassie-stained CriterionTM TGX gel of cytoplasmic, periplasmic fractions and medium samples and purifications (C, P, M) from BL21 and BL21 TatExpress cells (CX, PX, MX) TorA-YebF (marked as "-YebF"). Samples were analysed after 24 hours of growth in terrific broth based autoinduction media at 30°C in shake flasks. Samples from the same subcellular fraction are comparable between strains, but not between different subcellular compartments. Cytoplasm was diluted in 750 µL buffer. Periplasm was diluted in 400 µL buffer. Purifications of all cell fractions and medium are comparable among fractions and among strains. For all medium samples, 5 mL culture was recovered. Representative gel from triplicate experiments is shown.

Figure 5. **Expression of YebF control and YebF SPs constructs in ΔTat strain with and without CyDisCo.** Coomassie-stained 15% SDS-PAGE gel of periplasmic (P), spheroplast (Sph) fractions and medium (M) samples expressing YebF and YebF SP constructs. Samples were analysed 2 hours post-induction (50 µM IPTG) in LB media at 30°C in shake flasks. Samples from the same subcellular fraction are comparable between strains, but not between different subcellular compartments. Spheroplast was diluted in 750 µL buffer. Periplasm was diluted in 400 µL buffer. For all medium samples, 500 µL culture was recovered.

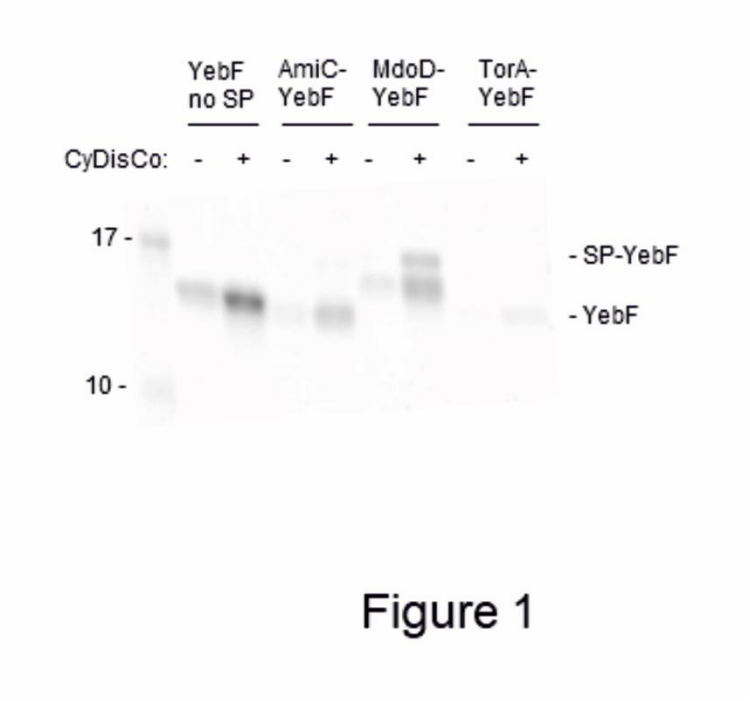

YebF no SP
AmiC-YebF
MdoD-YebF
TorA-YebF
CyDisCo: - + - + - + - +
17 -
10 -
- SP-YebF
- YebF

Figure 1

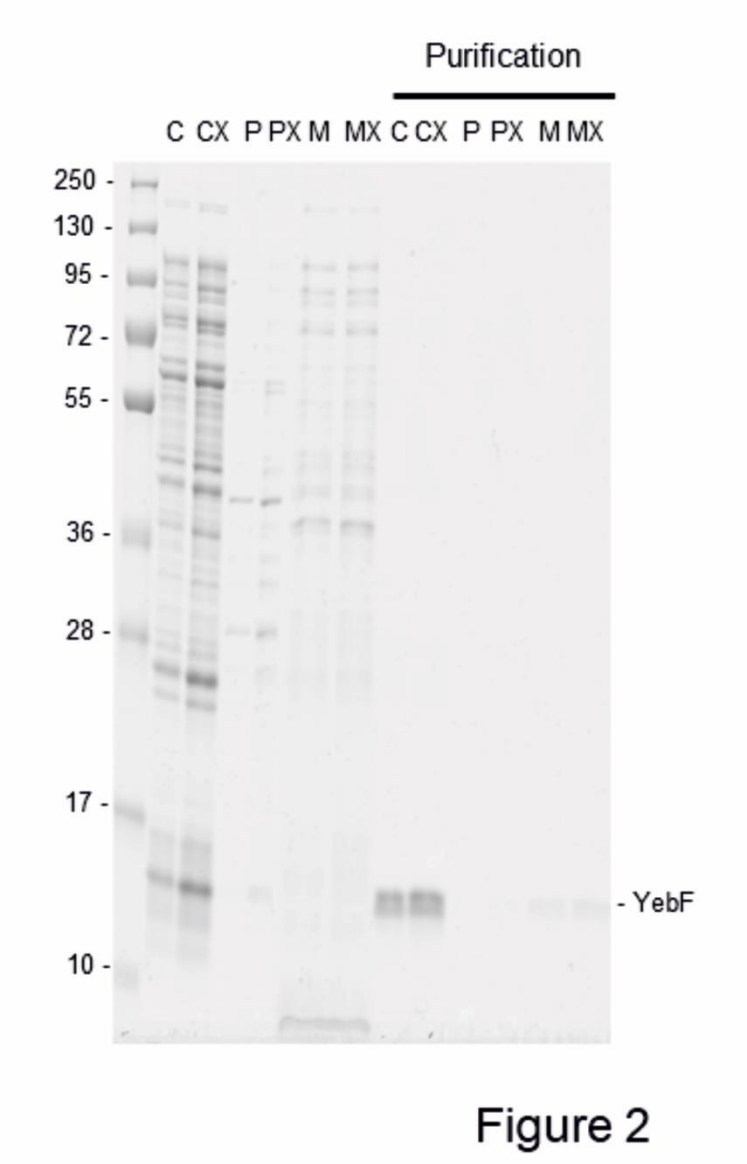

Purification
C CX P PX M MX C CX P PX M MX
250 -
130 -
95 -
72 -
55 -
36 -
28 -
17 -
10 -
- YebF

Figure 2

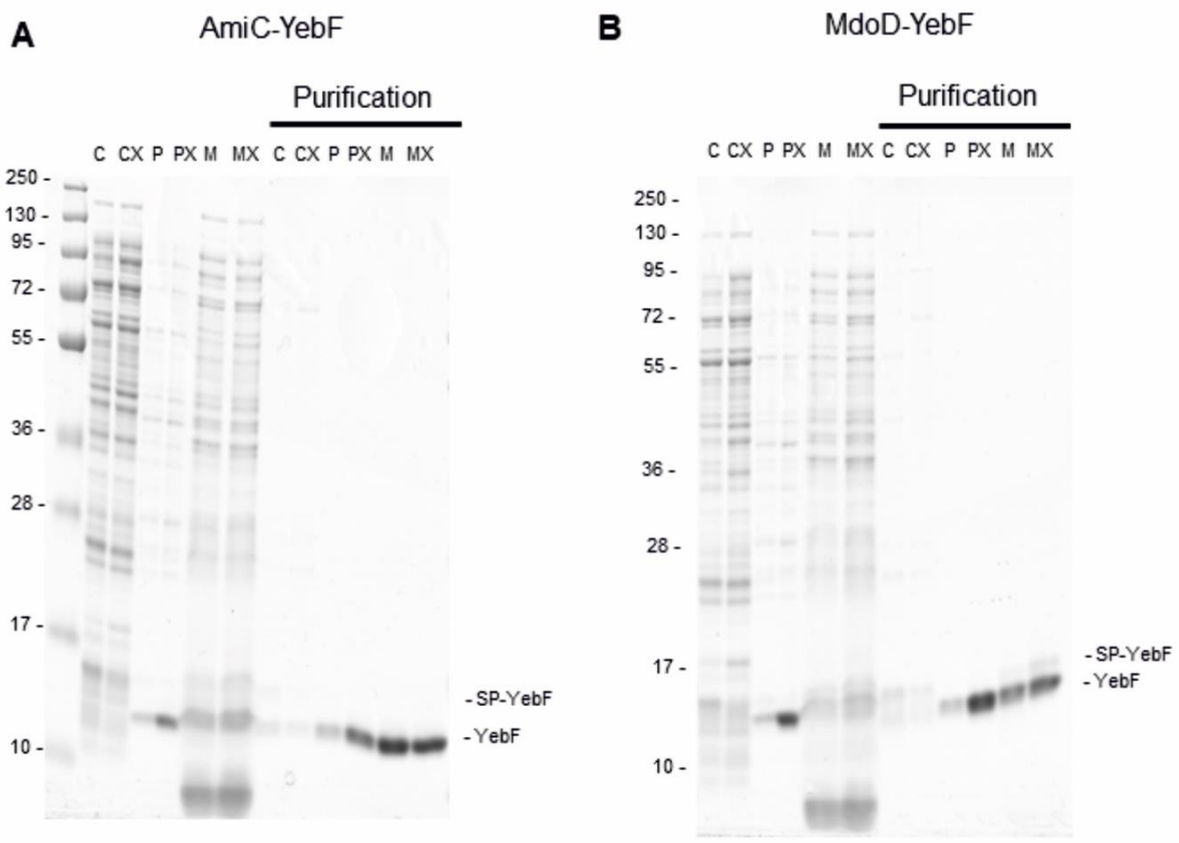

A
AmiC-YebF
Purification
C CX P PX M MX C CX P PX M MX
250 -
130 -
95 -
72 -
55 -
36 -
28 -
17 -
10 -
- SP-YebF
- YebF
B
MdoD-YebF
Purification
C CX P PX M MX C CX P PX M MX
250 -
130 -
95 -
72 -
55 -
36 -
28 -
17 -
10 -
- SP-YebF
- YebF

Figure 3

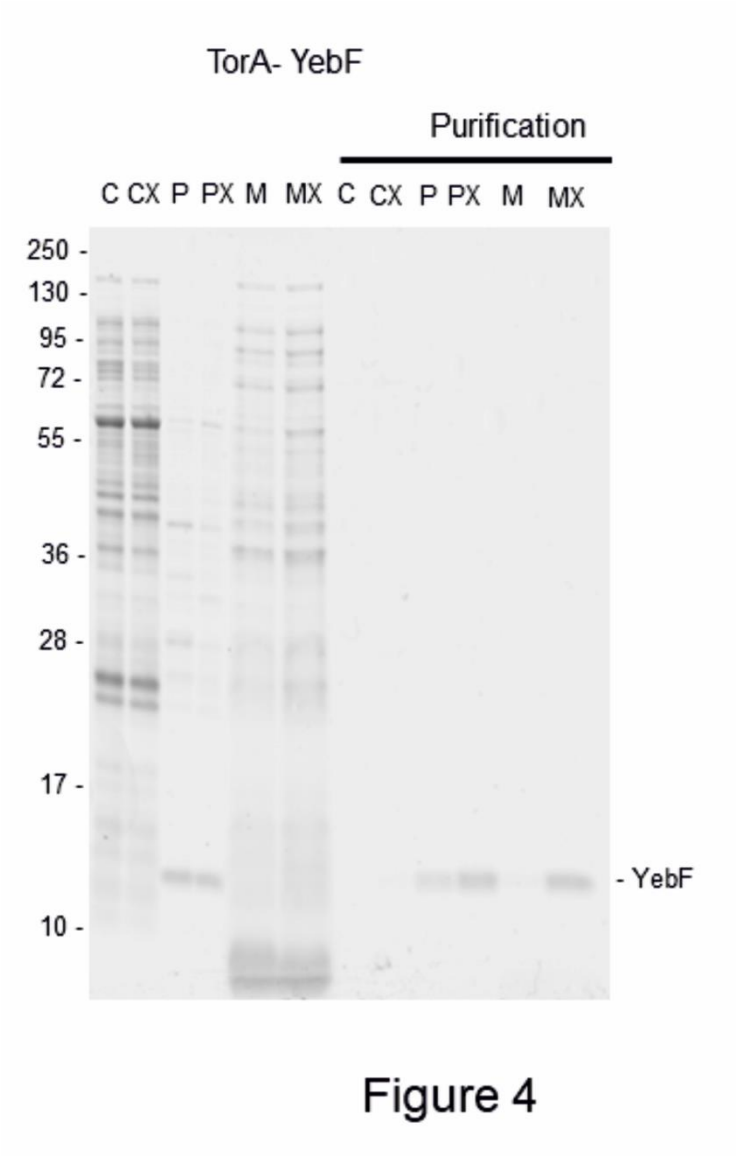

TorA- YebF
Purification
C CX P PX M MX C CX P PX M MX
250 -
130 -
95 -
72 -
55 -
36 -
28 -
17 -
10 -
- YebF

Figure 4

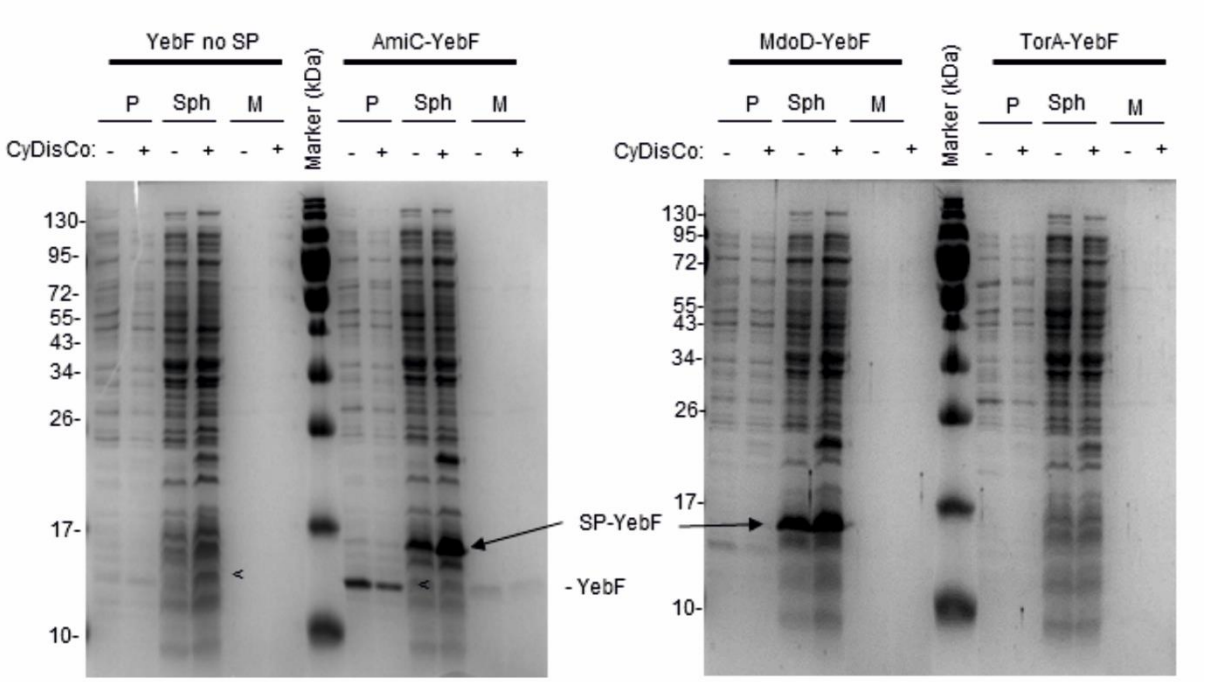

YebF no SP
AmiC-YebF
MdoD-YebF
TorA-YebF
P
Sph
M
CyDisCo: - + - + - +
Marker (kDa)
130-
95-
72-
55-
43-
34-
26-
17-
10-
SP-YebF
- YebF

Figure 5

CHAPTER 5: Final discussion

5.1. Proteins with multiple disulphide bonds as a substrate for Tat pathway

5.1.1. Context

Industrial exploitation of the Tat pathway for protein export could be a game changer for biopharmaceutical companies aiming to produce DSB containing proteins in a cost-effective way. However, our understanding of this pathway is still vague, which hampers the exploitation of this pathway. Only a few high-value proteins have been reported to be exported by Tat, and many of them have never been quantified or compared to what its counterpart Sec pathway can do. CyDisCo represents a very intriguing technology to enhance the potential of the Tat pathway. If CyDisCo folds the target protein, then Tat will possibly be able to transport it to the periplasm. At the same time, its proofreading mechanism will reject the misfolded proteins, ensuring the export of high-quality disulphide-bonded biopharmaceuticals. To date, CyDisCo enhanced the production of numerous disulphide-rich human proteins, including antibody formats, hGH and perlecan (Matos et al., 2014; Gąciarz et al., 2016; Sohail et al., 2020). More recently, functional recombinant SARS-CoV-2 spike receptor binding domain has been also produced by CyDisCo technology (Prahlad et al. 2021). Nevertheless, the application of CyDisCo to Tat pathway has only been proven effective in some cases like PhoA, AppA (model proteins) and an industrially relevant protein, a scFv (Matos et al., 2014). In some other cases, proteins like hGH, IFN α2b and other antibody fragments did not need CyDisCo for export by Tat (Alanen et al., 2015). Interestingly, all these proteins had the TorA SP attached as Tat specific SP. Therefore, there is a gap in understanding what the *E. coli* Tat pathway can actually export or cannot export with the help of CyDisCo. Finding and addressing the issues of the production of complex multiple disulphide bond containing proteins with CyDisCo for export by Tat pathway is the way to boost the industrial relevance of this tool.

5.1.2. Discussion of the work presented

In this study, we have identified some of the drawbacks of Tat mediated export when expressing CyDisCo. We realized that standard protocols, tools and plasmids for disulphide-bonded protein production are not adapted for the successful export of the proposed complex POIs with CyDisCo: Brazzein, Dulaglutide and Romiplostim. For this, an improved experimental design is proposed. Moreover, we found some general limitations of the export through Tat that should be further analysed.

5.1.2.1. Oxygen limitation: O_2 transfer specialized membrane, high shaking speeds and less culture volume helps CyDisCo performance

We realized that in our hands, CyDisCo was not performing as well as it should in comparison to what CyDisCo produces at our collaborators' lab with the same POIs. CyDisCo aeration is a crucial factor that was not considered as a major issue. Common laboratory sponges as O_2 permeable covers do not seem to permit enough O_2 transfer and specialized membranes are essential (e.g., AirOtop, Thomson). Increasing the rotation speed of the incubator to 250 rpm is a must too and lowering the volume of the culture helps in providing improved aeration. Even though in some already published papers this is not the methodology they follow, they might have been at the "limit" of oxygenation needed for CyDisCo to form the DSBs of their POIs.

5.1.2.2. Stoichiometry: low to medium amount of CyDisCo components and high amount of the POI in two plasmid-system for better results

A polycistronic low-copy number plasmid (with CyDisCo) does not produce enough CyDisCo to fold target proteins with several DSBs. Dual plasmid system with CyDisCo in a low to medium copy number plasmid and the POI in a low copy number plasmid, does not really produce enough good yields of the POI, but it allows the visualization of CyDisCo components in gels. Dual plasmid system with CyDisCo in a low to medium copy number plasmid and the POI in a high copy number plasmid seems to be the best option among all the combinations tried in Chapter 2 to produce proteins with multiple DSBs. When this is the choice, Tat overexpression via using TatExpress cells might decrease the chances of overloading the Tat system.

5.1.2.3. TorA SP is not the ideal SP for brazzein, dulaglutide and romiplostim

TorA SP is the most widely used SP for target protein export to the periplasm via Tat pathway. There are some examples where TorA SP served as a very effective SP (Matos et al., 2014; Alanen et al., 2015; Blaudeck et al., 2001; Guerrero Montero et al., 2019, Browning et al., 2017). On the other hand, in most of those cases, quantification of the total product obtained was not evaluated. Moreover, those proteins were simpler in comparison to our POIs. Based on our findings, TorA SP probably caused the degradation, IB formation and/or early cleavage of the SP of our POIs.

5.1.2.4. Media: Rich autoinduction media for CyDisCo + Tat system works better than LB media

Rich autoinduction media seems to work better for both single plasmid and the dual plasmid systems when trying to produce soluble proteins with CyDisCo. This is probably because it allows a constant aeration in a critical point when POI and CyDisCo are expressed by circumventing the need to stop the incubator for IPTG addition. Moreover, it permits higher ODs resembling batch fermentation, and it usually allows the identification of the POIs in gels without immunoblotting.

5.1.2.5. Other limitations and experimental parameters to be considered

Over the course of experiments in the thesis, a better fractionation method was implemented in order to minimize the cross-contamination of the cytoplasmic and periplasmic fraction that could lead to a misinterpretation of the obtained data. This new fractionation method is called PureFrac (Malherbe et al., 2019). Following this, in Chapter 2, cell lysis over time has not been assessed to determine how the change of culture conditions and different use of plasmids are affecting the cell integrity: the stress and metabolic burden of the cells was only based on cross-contamination of cell fractions. Additionally, it is important to note that the identification of the CyDisCo components in the gel may not be related to their functionality.

5.1.3 Future prospects

Two of the major problems we encountered when using University of Oulu's two plasmid-system with CyDisCo + Tat was the cell stress (based on cross-contamination of the cell fractions), probably due to the overexpression of three proteins or even more proteins in TatExpress cells, and the overloading of the Tat system (Sutherland et al., 2018). To investigate this, targeted mutations in the tac promoter have been proposed (Annex 4). Preliminary experiments showed that these mutations lower protein production: M1 mutation lowers target protein production more than M2, and M2 produces less target protein than the wild-type tac promoter (Figure 15). Fractionation of the cells and dual expression of CyDisCo and the POI is needed to further see the utility of such mutations in the export of disulphide bond containing proteins by Tat.

Figure 15. Purification of 6xHis-ERp29 protein model with different mutations in tac promoter. tac promoter was mutated to reduce the production of ERp29 (model protein). M1 (ptac mutation type 1); M2 (ptac mutation 2); C (control: wild type ptac). Mutants' sequences can be found in Annex 4. 6xHis-ERp29 was produced without CyDisCo in the cytoplasm of BL21 and the soluble fraction was purified by IMAC.

Testing other recipes of autoinduction media and experiments of trial and error to find better SPs while considering the risk of putative translocation are also experiments proposed for a further understanding of Tat + CyDisCo with these multi-DSB containing proteins. Finally, the impossibility of Tat to transport such a complex POI should not be discarded. The Tat translocation system is still not completely understood and hence avoiding an overestimation of Tat pathway capabilities can be the key for its industrial exploitation. The initial chosen protein targets were very valuable but very complex too, therefore the expression of a known Tat-dependent biopharmaceutical could be the ideal model protein to find what is holding Tat back to be the preferable pathway to export difficult-to-express proteins.

5.2. Yields and product comparison between *Escherichia coli* BL21 and W3110 in industrially relevant conditions: anti-c-Met scFv as a case study

5.2.1. Context

Growing *E. coli* to high cell densities in order to achieve maximum productivity has been a matter of study since the early 1970's (Shiloach & Fass, 2005). In *E. coli*, and generally in bacterial expression systems, heterologous protein secretion has almost exclusively been researched by directing the protein to the Sec pathway at fermentation scale. For the first time in 2019, Montero Guerrero et al. used fed-batch fermentation for the production of an industrially relevant protein (hGH) with the aim of testing Tat system's full potential. However, hGH does not need CyDisCo to fold and be transported via the Tat pathway. In the first chapter, we discovered some important parameters to consider when employing the CyDisCo system in combination with the Tat pathway but due to the complexity of the POIs assessed, further experiments with a simpler Tat and CyDisCo-dependent protein, an scFv (Jones et al., 2016), were proposed. Assessing the export of this biopharmaceutical at industrial scale in both 5 L bioreactors and miniaturized bioreactors (10 mL fermenters) could provide insights into the limitations and potential of the combination of these systems for the production of high-value proteins. Automated high-throughput approaches (miniaturized bioreactors) accelerate strain generation rates and enable early-stage microbial bioprocess development (Janzen et al., 2019). This can prove to be very valuable for the understanding of Tat as a tool for industrial bioproduction. Additionally, a comparison of Tat and Sec yields and product quality in the most used *E. coli* strains in industry, BL21 and W3110, was supposed to be carried out. Tat underperformed, and not enough protein could be obtained to assess and compare the quality of the protein exported via the Sec pathway. Alternatively, a compelling comparison of BL21 and W3110 yields and product heterogeneity when exporting the POI by Sec in fermenters provided a real input for product heterogeneity in industrial processes.

5.2.2. Discussion of the work presented

In this chapter, we have identified some more drawbacks of Tat mediated export when expressing an already tested simple Tat-dependent and almost fully CyDisCo-dependent biopharmaceutical. Moreover, we compared BL21 and W3110 yields and product heterogeneity when exporting a biopharmaceutical by Sec pathway in an industrial setting.

5.2.2.1. Tat pathway presents problems when scaling-up from 10 mL bioreactors to 5 L bioreactors in BL21 and W3110

Final soluble yields per OD of the proteins exported via the Tat pathway at the end of fermentations were found to be lower in 5 L fermentation than in 10 mL fermentations in both BL21 and W3110. This suggests that Tat pathway + CyDisCo exports more scFv at the smaller scale as compared to the larger scale. Although this is true, it is also very important to consider that 10 mL and 5 L fermentations are not carried out exactly the same way, as they both have different set-ups, but they are probed to be comparable (Janzen et al., 2019). Furthermore, 5 L scale is the benchmark cultivation in an industrial setting and not 10 mL cultivations.

5.2.2.2. TorA SP is not the ideal Tat SP for the chosen POI

In order to find out why Tat pathway dependent export of this scFv was so low in 5 L fermentation, we thoroughly explored the different reasons why this could have happened (see section 3.2. for further explanation). Overall and coinciding with literature and observations from Chapter 2, TorA SP might be causing the degradation of this scFv.

5.2.2.3. BL21 seems to be a better strain for Tat + CyDisCo in the setting presented in high-throughput multifermenters

Guerrero Montero et al. (2019) carried out Tat-dependent export of hGH in fed-batch fermentations in W3110 without CyDisCo. In the work presented here performed in high-throughput automated multifermenters with CyDisCo, BL21 *E. coli* strain seems to be a better option for the production of an scFv as yield/OD is higher than for W3110.

5.2.2.4. BL21 produces more soluble protein than W3110 in both 10 mL and 5 L fermentations when exporting by Sec

Following the methodology in section 3.3., BL21 seems to be a more ideal *E. coli* strain than W3110 for the production of the chosen target protein in both 10 mL and 5 L fermentations as final soluble production was found to be higher at the end of fermentation. It is necessary to remember that automatized fermentation protocols, including the media recipe, were adapted for heterologous protein production under certain conditions (Janzen et al., 2019); which suggests that different results could be obtained under alternative culture conditions.

5.2.2.5. The POI can be correctly folded in BL21 and W3110 but altered conformations of the protein might co-exist: heterogeneity in the product but similar profiles in both strains

This opens a discussion on whether mass spectrometry based intact mass analysis of the POI is enough to spot the heterogeneity in a product as even though the experimental molecular weight might coincide with the theoretical oxidized molecular weight, unexpected heterogeneity might be found from more exhaustive research on the POI structure. In Chapter 3, even though the target protein was correctly folded as confirmed by mass spectrometry, CEX showed different elution peaks of the protein in both strains in different ratios. When main peaks of each strain were run in 1D-(^{1}H)-NMR, heterogeneity in the product independent of the strain employed was seen.

5.2.2.6. BL21 and W3110 seem to be interchangeable based on their heterogeneity profile for the case study presented

Although the product heterogeneity should be further analysed using other techniques than mass spectrometry based intact mass, W3110 and BL21 seem to have very similar product heterogeneity profiles. This similarity makes them interchangeable in the production of our scFv. In this case, the choice of the host for the industrial production of this protein can be based on final product titer.

5.2.3. Future prospects

The major problem encountered in this chapter was the underperformance of Tat system in 5 L fermenters. This ended up deviating our research to another direction. We again suggest that the underperformance of Tat pathway in the described conditions is likely to be related to Tat specific TorA SP, as it might be causing the degradation of the target protein. Due to this, finding a more reliable Tat specific SP should ideally be the next step. Also, it is important to note that even though we strongly suggest that TorA SP is probably the major cause of the low yields in this case, we cannot completely ensure that CyDisCo is properly working in these conditions. Although CyDisCo has previously been shown to work in different media and in fermentation conditions (Gąciarz et al., 2017; Sohail et al., 2020), the construct employed was not tested at shake-flask level in the presence and absence of CyDisCo in the media proposed.

Finally, in order to identify the modification(s) observed in CEX and confirmed in (^{1}H)-NMR, 2D NMR and LC-MS based mapping should be carried out. These experiments are expensive, time-consuming and can further modify the product. A more exhaustive analysis and activity-based comparison might help in elucidating the modification(s) involved for later determine its implications on the biomolecule function and its industrial relevance. We would also like to encourage other scientists to further analyse and compare product heterogeneity (if any) in different *E. coli* strains for other biopharmaceuticals to help set up a correlation with our POI (an scFv).

5.3. Highly efficient export of a disulphide-bonded protein to the periplasm and medium by the Tat pathway using CyDisCo in *Escherichia coli*

5.3.1. Context

High protein titer is needed to keep up with demand and driving manufacturing costs down is essential (Zhu et al., 2017). CyDisCo technology combined with Tat pathway has already shown promising results (Matos et al., 2014): CyDisCo catalyzes the DSB formation in the POI, and Tat will export it to the periplasm. This way, thanks to the proofreading mechanism of the Tat pathway, a homogenous product can easily be recovered from the periplasm. In previous sections, we have hypothesised that TorA SP could be causing the degradation or IB formation of the high-value proteins tested. As a result, low amount of soluble protein was obtained, making the exploitation of Tat pathway very less attractive for the biopharmaceutical industry. It is then necessary to find an alternative to TorA SP in order to get higher yields and not to discard Tat pathway as an alternative to Sec based on TorA-POI titers. The *E. coli* genome encodes at least 29 putative SPs containing a twin arginine motif, characteristic of proteins exported via the Tat translocation pathway. Many of those are promiscuous SP capable of directing export of the POI via Tat or Sec or via both (Tullman-Ercek et al., 2007). A screen of SPs to obtain the best Tat dependent SP in combination with CyDisCo for each recombinant protein can diminish the promiscuity of the SP making them completely or mostly Tat-dependent for some POIs. Finding alternative SPs could minimize the disadvantages linked to the utilization of TorA SP and contribute to the relevance of the Tat pathway.

5.3.2. Discussion of the work presented

In this chapter we found out that alternative SPs to TorA SP allow very efficient secretion of a disulphide bond containing protein to the periplasm of *E. coli* via Tat in the presence of CyDisCo. As these SPs are far more efficient than the TorA SP, they are probably more suitable for large scale protein production than the more rigorously specific Tat-signal peptide.

5.3.2.1. Folding of the SP-POI must be CyDisCo dependent to later assess its Tat export dependency

Testing whether the POI attached to the SPs to be screened show a dependence on CyDisCo is essential to examine Tat-dependent export. If the solubility of the chosen recombinant protein (SP attached) increases with CyDisCo, this might be a good target for Tat. In the case of the target protein in Chapter 4, YebF, the POI without its SP and with AmiC, MdoD and TorA SPs was found to be mainly CyDisCo dependent, as more soluble protein could be detected in the presence of CyDisCo than in the absence of it.

5.3.2.2. For Tat dependency test: TatExpress cells export much more POI to the periplasm than standard BL21 cells with the screened SPs. TorA SP exports less efficiently than the screened SP alternatives

The second step to determine the SPs-POI Tat dependency is the comparison between the protein export in wild type cells and TatExpress cells. TatExpress cells have a *tac* promoter right before the original promoter of Tat in the genome, and the rest of the genome is identical to the wild type strain (Browning et al., 2017). These strains are cultured the same way and do not show any OD difference or sensitivity compared to wild type strains. If TatExpress cells export more POI to the periplasm than their wild type counterparts, the POI is probably exported via the Tat pathway. In Chapter 4, TatExpress cells exported more YebF than wild type cells with AmiC, MdoD and TorA SPs. TorA SP did not export as much target protein as AmiC and MdoD SP, suggesting that the latter two are more industrially relevant.

5.3.2.3. For Tat dependency test: *ΔTat cells*

ΔTat strain can also provide valuable information about the Tat dependency of the tested SP-POIs. The ΔTat strain lacks Tat pathway components and therefore only the Sec pathway is available for the export of the target POI (Tullman-Erceck et al., 2007). As many of the potential Tat specific and industrially relevant SPs are putative, a comparison between presence and absence of CyDisCo would determine the grade of specificity of the combination of SP-POI to be exported by Tat.

5.3.2.4. Folding of the POI and the cleavage of the SP from the POI

Once the fractionation technique is clean i.e. no cytoplasmic protein cross-contamination in the periplasmic fraction, intact mass of the POI should be assessed. In this way, DSB formation and correct SP cleavage from the POI can be analysed. Formation of native disulphide bonds can be an essential feature for the stabilization and for modulation of biological activities of the target protein (Wiedemann et al., 2020). Ideally, CyDisCo would have folded the protein and Tat proofread it. Additionally, SP's correct cleavage ensures the recovery of the mature protein.

5.3.3 Future prospects

Finding a completely Tat specific SP better than TorA SP remains a challenge. Mutations of AmiC and MdoD SP in order to make them more Tat specific were not completely successful, and even though they worked as SPs (Annex 3), lysis of the cells was too prominent (data not shown). Further research on possible mutations in SPs or screening of other POIs for these mutated SPs would be a way to go in order to broaden the knowledge that relates the POI with the successful/unsuccessful SP. Furthermore, it might also give insights into the characteristics needed to find a better and more universal Tat SP. On the other hand, attaching AmiC-YebF or MdoD-YebF to other difficult-to-express proteins for the export to the periplasm and media by Tat pathway in the presence of CyDisCo is a promising future research topic. This has a big industrial potential and initial fermentation runs of the AmiC-YebF construct with CyDisCo have been carried out as a proof of concept of the scalability of this construct (Figure 16). Additionally, these fermentation experiments corroborate Tat dependency of AmiC-YebF in the presence of CyDisCo i.e., TatExpress cells (TE) export more protein to the periplasm (P) than wild type (wt) strains and wild type spheroplasts (Sph) show uncleaved protein (AmiC-YebF). This is presumably due to the availability limitation of the Tat apparatus in the wild type strain and the impossibility of Sec export, as the protein has already been folded.

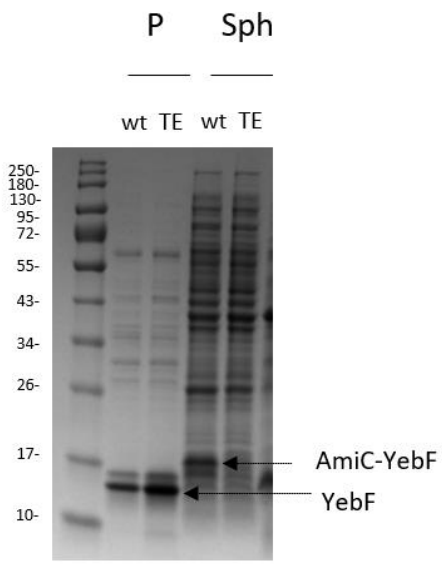


Figure 16. Periplasm and spheroplast fractions of BL21 wild type and TatExpress cells after 22 hours post induction in 0.5 L fermenters. Wild-type (wt) and TatExpress (TE) cells were fractionated following PureFrac fractionation protocol (Malherbe et al., 2019) to obtain periplasm (P) and spheroplast (Sph) fractions. YebF mature protein is marked by an arrow in the periplasm fraction. Pre-mature protein (AmiC-YebF) is marked by an arrow in the spheroplast fraction.

CHAPTER 6: Conclusion and overall future perspectives

Most *E. coli*- and bacilli-based production platforms use technology that is decades-old. Similarly for *Komagataella phaffii* (*Pichia pastoris)*, the basics have been developed in the 1990s, but new technology is urgently required to reach the status of 'mature chassis', especially when it comes to disulphide bond formation and glycosylation (Rettenbacher et al., 2022). In the case of *E. coli*, commercially available strains such as Origami and Shuffle strains have improved DSB formation of some heterologous proteins in their oxidizing cytoplasm (Bessette et al., 1999; Lobstein et al., 2012), but yields are rather low and in many cases their growth impairment makes these technologies not suitable for industrial use (Ren et al., 2016). A systematic comparison between the CyDisCo system and these commercial technologies can help in unravelling key limitations of oxidative folding and enhance the production capabilities of *E. coli* for disulphide-bonded biopharmaceuticals. On the other hand, successful attempts to transfer the *Campylobacter jejuni* bacteria (Wacker et al., 2002) or yeast glycosylation systems (Valderrama-Rincon et al., 2012) to *E. coli* have been made to allow the production of N-glycosylated heterologous proteins. Successful efforts to achieve O-glycosylation have also been made in *E. coli* (Du et al., 2019; Natarajan et al., 2020). These advances open new avenues for the production of glycoproteins in *E. coli*. However, these technologies are still in their infancy and have not been able to achieve a human-like N-glycosylation pattern which is a key requirement for industrially relevant therapeutic monoclonal antibodies (Rettenbacher et al., 2022).

Despite the impressive achievement of microbial production, current microbial systems have clear limitations and mammalian cell systems are increasingly used to produce complex molecules. However, these are difficult, time-consuming to set up and expensive (Aricescu et al., 2006). On the other hand, cell-free systems have been around for decades and are being improved and exploited as attractive alternatives to cell-based expression. However, they struggle due to low efficiency and high costs (Claassens et al., 2019). The efficient scale up of a cell-free protein synthesis for the production of virus-like-particles reported recently shows promise as a solution to a long-standing scalability issue (Armero-Gimenez et al., 2023)

The last decade has witnessed the emergence of an array of bispecific antibody variants, smaller antibody fragments and antibody mimetics (El Khatib & Salla, 2022). Many of these proteins are classed as difficult to express, and it is necessary to develop new platforms that can carry out disulphide bonding more efficiently (El Khatib & Salla, 2022), and which can secrete target proteins out of the cytoplasm into more protected environments. Overall, there is an urgent need to develop new, more efficient platforms in order to reduce costs and handle new format biotherapeutics that are emerging.

This project contributes to a better understanding of the production of difficult-to-express proteins and their export to the periplasm in *E. coli* with CyDisCo. Through the work presented here, an empirical baseline to produce multiple disulphide bonds containing proteins for export by Tat pathway at the shake-flask scale has been set and drawbacks have been identified and addressed. We report here that the CyDisCo system improves folding and soluble yields of disulphide-bonded proteins which potentially contributes to efficient Tat-dependent secretion. This improved experimental design, that combines the Tat pathway and CyDisCo technology for the production of high-value biopharmaceuticals, could be the starting point for many future research investigations involving the production of new antibody fragment formats and other disulphide-bonded proteins.

Additionally, in this study we report the limitations of mass spectrometry based intact mass analysis to detect heterogeneity in proteins. Therefore, we call for further analysis in order to identify as well as obtain more insights into the heterogeneity in recombinant proteins. On the other hand, the product heterogeneity profile was found to be similar in both *E. coli* BL21 and W3110 strains, which serves as an assurance of interchangeability of these strains at fermentation scale. Over other advantages, this could fasten and simplify the screening process to find the best host for industrial production based solely on the final product titer.

So far, Sec remains the pathway of choice for recombinant protein export to the periplasm; at least until a better Tat specific SP is found for the export of those target proteins. Promisingly, the combination of CyDisCo, TatExpress cells, and ΔTat cells can help with the discovery of these SPs for the target POIs, and our preliminary proof of concept findings suggest that yields

can be high and reproducible at the large scale. We showed that TorA SP could be one of the major issues when employing Tat as an export pathway for recombinant proteins and we found that in the presence of CyDisCo, alternative SPs increase the export of the target protein through the Tat pathway. Overall, this study has shown the potential of periplasmic production in *E. coli* within the biotechnology industry and establishes that it is possible to obtain highly efficient Tat-based secretion to the periplasm of disulphide-bonded protein YebF.

As a future perspective, testing the export of pre-folded difficult-to-express proteins to the periplasm and medium attached to YebF "passenger" by TatExpress cells with CyDisCo would be a very valuable research continuity. If this results in a correctly folded POI being proofread by the Tat system and secreted to the culture medium, it could potentially contribute to a simpler and more efficient recovery of the POI. Product recovery from the culture medium would be cheap, fast, and homogenous. At the same time, empirically testing more SP-POI combinations and trying to find a more universal Tat specific SP would be the ultimate goal. If this/these SPs is/are found, and if combined with CyDisCo and TatExpress cell lines, the mammalian cell lines and Sec as an export pathway would face a very strong competitor.

References

Adams, H., Scotti, P. A., De Cock, H., Luirink, J., & Tommassen, J. (2002). The presence of a helix breaker in the hydrophobic core of signal sequences of secretory proteins prevents recognition by the signal-recognition particle in Escherichia coli. *European Journal of Biochemistry*, *269*(22), 5564–5571. https://doi.org/10.1046/J.1432-1033.2002.03262.X

Adrio, J. L., & Demain, A. L. (2014). Microbial Enzymes: Tools for Biotechnological Processes. *Biomolecules*, *4*, 117–139. https://doi.org/10.3390/biom4010117

Agyei, D., Ahmed, I., Akram, Z., M. N. Iqbal, H., & K. Danquah, M. (2017). Protein and Peptide Biopharmaceuticals: An Overview. *Protein & Peptide Letters*, *24*(2), 94–101. https://doi.org/10.2174/0929866523666161222150444

Alami, M., Lüke, 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. *Molecular Cell*, *12*(4), 937–946. https://doi.org/10.1016/S1097-2765(03)00398-8

Alanen, H. I., Walker, K. L., Lourdes Velez Suberbie, M., Matos, C. F. R. O., Bönisch, S., Freedman, R. B., Keshavarz-Moore, E., Ruddock, L. W., & Robinson, C. (2015). 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. *Biochimica et Biophysica Acta*, *1853*(3), 756–763. https://doi.org/10.1016/J.BBAMCR.2014.12.027

Alcock, F., Baker, M. A. B., Greene, N. P., Palmer, T., Wallace, M. I., & Berks, B. C. (2013). Live cell imaging shows reversible assembly of the TatA component of the twin-arginine protein transport system. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), E3650–E3659. https://doi.org/10.1073/PNAS.1306738110

Aricescu, A. R., Lu, W., & Jones, E. Y. (2006). A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallographica Section D: Biological Crystallography*, *62*(10), 1243–1250. https://doi.org/10.1107/S0907444906029799

Armero-Gimenez, J., Wilbers, R., Schots, A., Williams, C., & Finnern, R. (2023). Rapid screening and scaled manufacture of immunogenic virus-like particles in a tobacco BY-2 cell-free protein synthesis system. *Frontiers in Immunology*, *14*. doi: 10.3389/fimmu.2023.1088852

Assadi-Porter, F. M., Aceti, D. J., & Markley, J. L. (2000). Sweetness Determinant Sites of Brazzein, a Small, Heat-Stable, Sweet-Tasting Protein. *Archives of Biochemistry and Biophysics*, *376*(2), 259–265. https://doi.org/10.1006/ABBI.2000.1726

Assadi-Porter, F. M., Patry, S., & Markley, J. L. (2008). Efficient and rapid protein expression and purification of small high disulfide containing sweet protein brazzein in E. coli. *Protein Expression and Purification*, *58*(2), 263. https://doi.org/10.1016/J.PEP.2007.11.009

Auclair, S. M., Bhanu, M. K., & Kendall, D. A. (2012). Signal peptidase I: Cleaving the way to mature proteins. *Protein Science*, *21*(1), 13–25. https://doi.org/10.1002/PRO.757

Baeshen, M. N., Al-Hejin, A. M., Bora, R. S., Ahmed, M. M. M., Ramadan, H. A. I., Saini, K. S., Baeshen, N. A., & Redwan, E. M. (2015). Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives. *J. Microbiol. Biotechnol.*, *25*(7), 953–962. https://doi.org/10.4014/JMB.1412.12079

Baeshen, N. A., Baeshen, M. N., Sheikh, A., Bora, R. S., Ahmed, M. M. M., Ramadan, H. A. I., Saini, K. S., & Redwan, E. M. (2014). Cell factories for insulin production. *Microbial Cell Factories*, *13*(1). https://doi.org/10.1186/S12934-014-0141-0

Bagherinejad, M. R., Sadeghi, H. M. M., Abedi, D., Chou, C. P., Moazen, F., & Rabbani, M. (2016). Twin arginine translocation system in secretory expression of recombinant human

growth hormone. *Research in Pharmaceutical Sciences*, *11*(6), 461. https://doi.org/10.4103/1735-5362.194871

Baglieri, J., Beck, D., Vasisht, N., Smith, C. J., & Robinson, C. (2012). Structure of TatA paralog, TatE, suggests a structurally homogeneous form of Tat protein translocase that transports folded proteins of differing diameter. *The Journal of Biological Chemistry*, *287*(10), 7335–7344. https://doi.org/10.1074/JBC.M111.326355

Bai, F., Li, Z., Umezawa, A., Terada, N., & Jin, S. (2018). Bacterial type III secretion system as a protein delivery tool for a broad range of biomedical applications. *Biotechnology Advances*, *36*(2), 482–493. https://doi.org/10.1016/J.BIOTECHADV.2018.01.016

Balasundaram, B., Harrison, S., & Bracewell, D. G. (2009). Advances in product release strategies and impact on bioprocess design. *Trends in Biotechnology*, *27*(8), 477–485. https://doi.org/10.1016/j.tibtech.2009.04.004

Barnett, J. P., Van Der Ploeg, R., Eijlander, R. T., Nenninger, A., Mendel, S., Rozeboom, R., Kuipers, O. P., Van Dijl, J. M., & Robinson, C. (2009). The twin-arginine translocation (Tat) systems from Bacillus subtilis display a conserved mode of complex organization and similar substrate recognition requirements. *The FEBS Journal*, *276*(1), 232–243. https://doi.org/10.1111/J.1742-4658.2008.06776.X

Behrendt, J., Standar, K., Lindenstrauss, U., & Brüser, T. (2004). Topological studies on the twin-arginine translocase component TatC. *FEMS Microbiology Letters*, *234*(2), 303–308. https://doi.org/10.1111/J.1574-6968.2004.TB09548.X

Berks, B. C. (1996). A common export pathway for proteins binding complex redox cofactors? *Molecular Microbiology*, *22*(3), 393–404. https://doi.org/10.1046/J.1365-2958.1996.00114.X

Berks, B. C. (2015). The Twin-Arginine Protein Translocation Pathway. *Https://Doi.Org/10.1146/Annurev-Biochem-060614-034251*, *84*, 843–864. https://doi.org/10.1146/ANNUREV-BIOCHEM-060614-034251

Berks, B. C., Palmer, T., & Sargent, F. (2003). The Tat protein translocation pathway and its role in microbial physiology. *Advances in Microbial Physiology*, *47*, 187–254. https://doi.org/10.1016/S0065-2911(03)47004-5

Berks, B. C., Palmer, T., & Sargent, F. (2005). Protein targeting by the bacterial twin-arginine translocation (Tat) pathway. *Current Opinion in Microbiology*, *8*(2), 174–181. https://doi.org/10.1016/J.MIB.2005.02.010

Berks, B. C., Sargent, F., & Palmer, T. (2000). The Tat protein export pathway. *Molecular Microbiology*, *35*(2), 260–274. https://doi.org/10.1046/J.1365-2958.2000.01719.X

Berlec, A., Tompa, G., Slapar, N., Fonović, U. P., Rogelj, I., & Štrukelj, B. (2008). Optimization of fermentation conditions for the expression of sweet-tasting protein brazzein in Lactococcus lactis. *Letters in Applied Microbiology*, *46*(2), 227–231. https://doi.org/10.1111/J.1472-765X.2007.02297.X

Berlec, A., Jevnikar, Z., Majhenič, A. Č., Rogelj, I., & Štrukelj, B. (2006). Expression of the sweet-tasting plant protein brazzein in Escherichia coli and Lactococcus lactis: a path toward sweet lactic acid bacteria. *Appl Microbiol Biotechnol*, *73*, 158–165. https://doi.org/10.1007/s00253-006-0438-y

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), 1171–1182. https://doi.org/10.1046/J.1365-2958.2003.03511.X

Bessette, P. H., Åslund, F., Beckwith, J., & Georgiou, G. (1999). Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 13703–13708.

https://doi.org/10.1073/PNAS.96.24.13703

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X. Z., & Zhou, T. (2021). Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. *Frontiers in Bioengineering and Biotechnology*, *9*. https://doi.org/10.3389/FBIOE.2021.630551

Blaudeck, N., Sprenger, G. A., Freudl, R., & Wiegert, T. (2001). Specificity of signal peptide recognition in Tat-dependent bacterial protein translocation. *Journal of Bacteriology*, *183*(2), 604–610. https://doi.org/10.1128/JB.183.2.604-610.2001

Blaudeck, Natascha, Kreutzenbeck, P., Freudl, R., & Sprenger, G. A. (2003). Genetic analysis of pathway specificity during posttranslational protein translocation across the Escherichia coli plasma membrane. *Journal of Bacteriology*, *185*(9), 2811–2819. https://doi.org/10.1128/JB.185.9.2811-2819.2003

Blount, Z. D. (2015). The unexhausted potential of E. coli. *ELife*, *4*. https://doi.org/10.7554/ELIFE.05826

Blümmel, A. S., Drepper, F., Knapp, B., Eimer, E., Warscheid, B., Müller, M., & Fröbel, J. (2017). Structural features of the TatC membrane protein that determine docking and insertion of a twin-arginine signal peptide. *The Journal of Biological Chemistry*, *292*(52), 21320. https://doi.org/10.1074/JBC.M117.812560

Blümmel, A. S., Haag, L. A., Eimer, E., Müller, M., & Fröbel, J. (2015). Initial assembly steps of a translocase for folded proteins. *Nature Communications*, *6*. https://doi.org/10.1038/NCOMMS8234

Bolhuis, A., Mathers, J. E., Thomas, J. D., Claire, M., Barrett, L., & Robinson, C. (2001). TatB and TatC Form a Functional and Structural Unit of the Twin-arginine Translocase from Escherichia coli*. *Journal of Biological Chemistry*, *276*(23), 20213–20219. http://www.jbc.org/article/S002192581940464X/fulltext

Browning, D. F., Godfrey, R. E., Richards, K. L., Robinson, C., & Busby, S. J. W. (2019). Exploitation of the *Escherichia coli lac* operon promoter for controlled recombinant protein production. *Biochemical Society transactions*, *47*(2), 755–763. https://doi.org/10.1042/BST20190059

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. *Biotechnology and Bioengineering*, *114*(12), 2828–2836. https://doi.org/10.1002/bit.26434

Burness, C. B., & Scott, L. J. (2015). Dulaglutide: A Review in Type 2 Diabetes. *BioDrugs*, *29*(6), 407–418. https://doi.org/10.1007/S40259-015-0143-4/FIGURES/1

Caldelari, I., Palmer, T., & Sargent, F. (2008). Escherichia coli tat mutant strains are able to transport maltose in the absence of an active malE gene. *Arch Microbiol*, *189*, 597–604. https://doi.org/10.1007/s00203-008-0356-8

Caldwell, J. E., Abildgaard, F., Dzakula, Z., Ming, D., Hellekant, G., & Markley, J. L. (1998). Solution structure of the thermostable sweet-tasting protein brazzein. *Nature Structural Biology 1998 5:6*, *5*(6), 427–431. https://doi.org/10.1038/nsb0698-427

Caspers, M., Brockmeier, U., Degering, C., Eggert, T., & Freudl, R. (2010). Improvement of Sec-dependent secretion of a heterologous model protein in Bacillus subtilis by saturation mutagenesis of the N-domain of the AmyE signal peptide. *APPLIED GENETICS AND MOLECULAR BIOTECHNOLOGY*. https://doi.org/10.1007/s00253-009-2405-x

Chen, H., Kim, J., & Kendall, D. A. (1996). Competition between functional signal peptides demonstrates variation in affinity for the secretion pathway. *Journal of Bacteriology*, *178*(23), 6658–6664. https://doi.org/10.1128/JB.178.23.6658-6664.1996

Chen, R. (2012). Bacterial expression systems for recombinant protein production: E. coli and beyond. *Biotechnology Advances*, *30*(5), 1102–1107. https://doi.org/10.1016/J.BIOTECHADV.2011.09.013

Claassens, N. J., Burgener, S., Vögeli, B., Erb, T. J., & Bar-Even, A. (2019). A critical comparison of cellular and cell-free bioproduction systems. *Current opinion in biotechnology*, *60*, 221-229. https://doi.org/10.1016/j.copbio.2019.05.003

Cline, K. (2015). Mechanistic Aspects of Folded Protein Transport by the Twin Arginine Translocase (Tat). *The Journal of Biological Chemistry*, *290*(27), 16530–16538. https://doi.org/10.1074/JBC.R114.626820

Correa, A., & Oppezzo, P. (2015). Overcoming the solubility problem in E. coli: Available approaches for recombinant protein production. *Methods in Molecular Biology*, *1258*, 27–44. https://doi.org/10.1007/978-1-4939-2205-5_2

Costa, S., Almeida, A., Castro, A., & Domingues, L. (2014). Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Frontiers in Microbiology*, *5*(FEB). https://doi.org/10.3389/FMICB.2014.00063

Costa, S. J., Coelho, E., Franco, L., Almeida, A., Castro, A., & Domingues, L. (2013). The Fh8 tag: A fusion partner for simple and cost-effective protein purification in Escherichia coli. *Protein Expression and Purification*, *92*(2), 163–170. https://doi.org/10.1016/J.PEP.2013.09.013

Cristóbal, S., De Gier, J. W., Nielsen, H., & Von Heijne, G. (1999). Competition between Sec- and TAT-dependent protein translocation in Escherichia coli. *The EMBO Journal*, *18*(11), 2982–2990. https://doi.org/10.1093/EMBOJ/18.11.2982

Croxen, M. A., Law, R. J., Scholz, R., Keeney, K. M., Wlodarska, M., & Finlay, B. B. (2013). Recent advances in understanding enteric pathogenic Escherichia coli. *Clinical Microbiology Reviews*, *26*(4), 822–880. https://doi.org/10.1128/CMR.00022-13

Daegelen, P., Studier, F. W., Lenski, R. E., Cure, S., & Kim, J. F. (2009). Tracing Ancestors and Relatives of Escherichia coli B, and the Derivation of B Strains REL606 and BL21(DE3). *Journal of Molecular Biology*, *394*(4), 634–643. https://doi.org/10.1016/J.JMB.2009.09.022

Date, T. (1983). Demonstration by a novel genetic technique that leader peptidase is an essential enzyme of Escherichia coli. *Journal of Bacteriology*, *154*(1), 76–83. https://doi.org/10.1128/JB.154.1.76-83.1983

de Marco, A. (2009). Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microbial Cell Factories*, *8*. https://doi.org/10.1186/1475-2859-8-26

De Oliveira, J. E., Soares, C. R. J., Peroni, C. N., Gimbo, E., Camargo, I. M. C., Morganti, L., Bellini, M. H., Affonso, R., Arkaten, R. R., Bartolini, P., & Ribela, M. T. C. P. (1999). High-yield purification of biosynthetic human growth hormone secreted in Escherichia coli periplasmic space. *Journal of Chromatography A*, *852*(2), 441–450. https://doi.org/10.1016/S0021-9673(99)00613-5

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), 6115–6120. https://doi.org/10.1073/PNAS.0937838100

Djoko, K. Y., Chong, L. X., Wedd, A. G., & Xiao, Z. (2010). Reaction mechanisms of the multicopper oxidase CueO from Escherichia coli support its functional role as a cuprous oxidase. *Journal of the American Chemical Society*, *132*(6), 2005–2015. https://doi.org/10.1021/JA9091903

Du, T., Buenbrazo, N., Kell, L., Withers, S. G., Defrees, S., & Wakarchuk, W. (2019). A Bacterial Expression Platform for Production of Therapeutic Proteins Containing Human-like O-Linked Glycans. *Cell Chemical Biology*, *26*, 203–212. https://doi.org/10.1016/j.chembiol.2018.10.017

El Khatib, S., & Salla, M. (2022). The Mosaic Puzzle of the Therapeutic Monoclonal Antibodies and Antibody Fragments A Modular Transition from Full Length Immunoglobulins to Antibody Mimetics. *Leukemia Research Reports*, 100335. https://doi.org/10.1016/j.lrr.2022.100335

Ellis, M., Patel, P., Edon, M., Ramage, W., Dickinson, R., & Humphreys, D. P. (2017). Development of a high yielding E. coli periplasmic expression system for the production of humanized Fab' fragments. *Biotechnology Progress*, *33*(1), 212–220. https://doi.org/10.1002/BTPR.2393

Ensuring Access to Safe, Affordable, and Effective Generic Drugs. (2020). www.fda.gov

Fisher, A. C., Kim, J. Y., Perez-Rodriguez, R., Tullman-Ercek, D., Fish, W. R., Henderson, L. A., & Delisa, M. P. (2008). Exploration of twin-arginine translocation for expression and purification of correctly folded proteins in Escherichia coli. *Microbial Biotechnology*, *1*(5), 403–415. https://doi.org/10.1111/J.1751-7915.2008.00041.X

Frain, K. M., Robinson, C., & van Dijl, J. M. (2019). Transport of Folded Proteins by the Tat System. *The Protein Journal 2019 38:4*, *38*(4), 377–388. https://doi.org/10.1007/S10930-019-09859-Y

Frampton, J. E., & Lyseng-Williamson, K. A. (2009). Romiplostim. *Drugs*, *69*(3), 307–317. https://doi.org/10.2165/00003495-200969030-00006

Francis, D. M., & Page, R. (2010). Strategies to Optimize Protein Expression in E. coli. *Current Protocols in Protein Science*, *61*(1), 5.24.1-5.24.29. https://doi.org/10.1002/0471140864.PS0524S61

Freudl, R. (2018). Signal peptides for recombinant protein secretion in bacterial expression systems. *Microbial Cell Factories*, *17*(1), 1–10. https://doi.org/10.1186/S12934-018-0901-3

Fröbel, J., Blümmel, A. S., Drepper, F., Warscheid, B., & Müller, M. (2019). Surface-exposed domains of TatB involved in the structural and functional assembly of the Tat translocase in Escherichia coli. *Journal of Biological Chemistry*, *294*(38), 13902–13914. https://doi.org/10.1074/JBC.RA119.009298

Fröbel, J., Rose, P., Lausberg, F., Blümmel, A. S., Freudl, R., & Müller, M. (2012). Transmembrane insertion of twin-arginine signal peptides is driven by TatC and regulated by TatB. *Nature Communications 2012 3:1*, *3*(1), 1–10. https://doi.org/10.1038/ncomms2308

Gąciarz, A., Khatri, N. K., Velez-Suberbie, M. L., Saaranen, M. J., Uchida, Y., Keshavarz-Moore, E., & Ruddock, L. W. (2017). Efficient soluble expression of disulfide bonded proteins in the cytoplasm of Escherichia coli in fed-batch fermentations on chemically defined minimal media. *Microbial Cell Factories*, *16*(1), 1–12. https://doi.org/10.1186/S12934-017-0721-X

Gąciarz, A., & Ruddock, L. W. (2017). Complementarity determining regions and frameworks contribute to the disulfide bond independent folding of intrinsically stable scFv. *PloS One*, *12*(12). https://doi.org/10.1371/JOURNAL.PONE.0189964

Gąciarz, A., Veijola, J., Uchida, Y., Saaranen, M. J., Wang, C., Hörkkö, S., & Ruddock, L. W. (2016). Systematic screening of soluble expression of antibody fragments in the cytoplasm of E. coli. *Microbial Cell Factories*, *15*(1), 22. https://doi.org/10.1186/S12934-016-0419-5

Georgiou, G., & Segatori, L. (2005). Preparative expression of secreted proteins in bacteria: status report and future prospects. *Current Opinion in Biotechnology*, *16*(5), 538–545.

https://doi.org/10.1016/J.COPBIO.2005.07.008

Goel, N., & Stephens, S. (2010). mAbs Certolizumab Pegol. *Certolizumab Pegol*, *2*(2), 137–147. https://doi.org/10.4161/mabs.2.2.11271

Gouridis, G., Karamanou, S., Gelis, I., Kalodimos, C. G., & Economou, A. (2009). Signal peptides are allosteric activators of the protein translocase. *Nature*, *462*(7271), 363–367. https://doi.org/10.1038/NATURE08559

Grassi, L., & Cabrele, C. (2019). *Susceptibility of protein therapeutics to spontaneous chemical modifications by oxidation, cyclization, and elimination reactions*. *51*, 1409–1431. https://doi.org/10.1007/s00726-019-02787-2

Green, E. R., & Mecsas, J. (2016). Bacterial Secretion Systems: An Overview. *Microbiology Spectrum*, *4*(1). https://doi.org/10.1128/MICROBIOLSPEC.VMBF-0012-2015

Guerrero Montero, I., Richards, K. L., Jawara, C., Browning, D. F., Peswani, A. R., Labrit, M., Allen, M., Aubry, C., Davé, E., Humphreys, D. P., Busby, S. J. W., & Robinson, C. (2019). Escherichia coli "TatExpress" strains export several g/L human growth hormone to the periplasm by the Tat pathway. *Biotechnology and Bioengineering*, *116*(12), 3282–3291. https://doi.org/10.1002/BIT.27147

Guerrero Montero, I., Dolata, K. M., Schlüter, R., Malherbe, G., Sievers, S., Zühlke, D., Sura, T., Dave, E., Riedel, K., & 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. *Microbial Cell Factories*, *18*(1). https://doi.org/10.1186/S12934-019-1071-7

Han, M. J., Kim, J. Y., & Kim, J. A. (2014). Comparison of the large-scale periplasmic proteomes of the Escherichia coli K-12 and B strains. *Journal of Bioscience and Bioengineering*, *117*(4), 437–442. https://doi.org/10.1016/J.JBIOSC.2013.09.008

Harrison, J. J., Ceri, H., Badry, E. A., Roper, N. J., Tomlin, K. L., & Turner, R. J. (2005). Effects of the twin-arginine translocase on the structure and antimicrobial susceptibility of Escherichia coli biofilms. *Canadian Journal of Microbiology*, *51*(8), 671–683. https://doi.org/10.1139/W05-048

Hatahet, F., Nguyen, V. D., Salo, K. E. H., & Ruddock, L. W. (2010). Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. *Microbial Cell Factories*, *9*(1), 67. https://doi.org/10.1186/1475-2859-9-67/TABLES/3

Hercus, T. R., Barry, E. F., Dottore, M., McClure, B. J., Webb, A. I., Lopez, A. F., Young, I. G., & Murphy, J. M. (2013). High Yield Production of a Soluble Human Interleukin-3 Variant from E. coli with Wild-Type Bioactivity and Improved Radiolabeling Properties. *PLoS ONE*, *8*(8). https://doi.org/10.1371/JOURNAL.PONE.0074376

Hou, B., Heidrich, E. S., Mehner-Breitfeld, D., & Brüser, T. (2018). The TatA component of the twin-arginine translocation system locally weakens the cytoplasmic membrane of Escherichia coli upon protein substrate binding. *Journal of Biological Chemistry*, *293*(20), 7592–7605. https://doi.org/10.1074/JBC.RA118.002205

Huang, K. xue, Badger, M., Haney, K., & Evans, S. L. (2007). Large scale production of Bacillus thuringiensis PS149B1 insecticidal proteins Cry34Ab1 and Cry35Ab1 from Pseudomonas fluorescens. *Protein Expression and Purification*, *53*(2), 325–330. https://doi.org/10.1016/J.PEP.2007.01.010

Idalia, V. M. N., & Bernardo, F. (2017). *Escherichia coli* as a Model Organism and Its Application in Biotechnology. *Recent Advances on Physiology, Pathogenesis and Biotechnological Applications*. https://doi.org/10.5772/67306

Ilbert, M., Méjean, V., & Iobbi-Nivol, C. (2004). Functional and structural analysis of members of the TorD family, a large chaperone family dedicated to molybdoproteins. *Microbiology*,

150(4), 935–943. https://doi.org/10.1099/MIC.0.26909-0

Inouye, S., Soberon, X., Franceschini, T., Nakamura, K., & Itakura, K. (1982). Role of positive charge on the amino-terminal region of the signal peptide in protein secretion across the membrane. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(11), 3438–3441. https://doi.org/10.1073/PNAS.79.11.3438

Ize, B., Stanley, N. R., Buchanan, G., & Palmer, T. (2003). Role of the Escherichia coli Tat pathway in outer membrane integrity. *Molecular Microbiology*, *48*(5), 1183–1193. https://doi.org/10.1046/J.1365-2958.2003.03504.X

Jack, R. L., Sargent, F., Berks, B. C., Sawers, G., & Palmer, T. (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), 1801–1804. https://doi.org/10.1128/JB.183.5.1801-1804.2001

Jack, Rachael L., Buchanan, G., Dubini, A., Hatzixanthis, K., Palmer, T., & Sargent, F. (2004). Coordinating assembly and export of complex bacterial proteins. *The EMBO Journal*, *23*(20), 3962–3972. https://doi.org/10.1038/SJ.EMBOJ.7600409

Janzen, N. H., Striedner, G., Jarmer, J., Voigtmann, M., Abad, S., & Reinisch, D. (2019). Implementation of a Fully Automated Microbial Cultivation Platform for Strain and Process Screening. *Biotechnology Journal*, *14*(10), 1800625. https://doi.org/10.1002/BIOT.201800625

Jimenez-Solem, E., Rasmussen, M. H., Christensen, M., & Knop, F. K. (2010). Dulaglutide, a long-acting GLP-1 analog fused with an Fc antibody fragment for the potential treatment of type 2 diabetes. *Current Opinion in Molecular Therapeutics*, *12*(6), 790–797.

Jo, H. J., Noh, J. S., & Kong, K. H. (2013). Efficient secretory expression of the sweet-tasting protein brazzein in the yeast Kluyveromyces lactis. *Protein Expression and Purification*, *90*(2), 84–89. https://doi.org/10.1016/J.PEP.2013.05.001

Johnson, I. S. (1983). Human Insulin from Recombinant DNA Technology. *Science*, *219*(4585), 632–637. https://doi.org/10.1126/SCIENCE.6337396

Jones, A. S., Austerberry, J. I., Dajani, R., Warwicker, J., Curtis, R., Derrick, J. P., & Robinson, C. (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), 3116–3124. https://doi.org/10.1016/J.BBAMCR.2016.09.006

Jong, W. S. P., Vikström, D., Houben, D., Berg van Saparoea, H. B., Gier, J. W., & Luirink, J. (2017). Application of an E. coli signal sequence as a versatile inclusion body tag. *Microbial Cell Factories*, *16*(1), 1–13. https://doi.org/10.1186/S12934-017-0662-4

Jozala, A. F., Geraldes, D. C., Tundisi, L. L., Feitosa, V. de A., Breyer, C. A., Cardoso, S. L., Mazzola, P. G., de Oliveira-Nascimento, L., Rangel-Yagui, C. de O., Magalhães, P. de O., de Oliveira, M. A., & Pessoa, A. (2016). Biopharmaceuticals from microorganisms: from production to purification. *Brazilian Journal of Microbiology*, *47*, 51–63. https://doi.org/10.1016/J.BJM.2016.10.007

Kamionka, M. (2011). Engineering of therapeutic proteins production in Escherichia coli. *Current Pharmaceutical Biotechnology*, *12*(2), 268–274. https://doi.org/10.2174/138920111794295693

Karamyshev, A. L., Karamysheva, Z. N., Kajava, A. V., Ksenzenko, V. N., & Nesmeyanova, M. A. (1998). Processing of Escherichia coli alkaline phosphatase: role of the primary structure of the signal peptide cleavage region. *Journal of Molecular Biology*, *277*(4), 859–870. https://doi.org/10.1006/JMBI.1997.1617

Karyolaimos, A., Ampah-Korsah, H., Hillenaar, T., Mestre Borras, A., Dolata, K. M., Sievers, S.,

Riedel, K., Daniels, R., & de Gier, J. W. (2019). Enhancing Recombinant Protein Yields in the E. coli Periplasm by Combining Signal Peptide and Production Rate Screening. *Frontiers in Microbiology*, *10*, 1511. https://doi.org/10.3389/FMICB.2019.01511

Keating, G. M. (2012). Romiplostim: A review of its use in immune thrombocytopenia. *Drugs*, *72*(3), 415–435. https://doi.org/10.2165/11208260-000000000-00000

Khushoo, A., Pal, Y., Singh, B. N., & Mukherjee, K. J. (2004). Extracellular expression and single step purification of recombinant Escherichia coli l-asparaginase II. *Protein Expression and Purification*, *38*(1), 29–36. https://doi.org/10.1016/J.PEP.2004.07.009

Kleiner-Grote, G. R. M., Risse, J. M., & Friehs, K. (2018). Secretion of recombinant proteins from E. coli. *Engineering in Life Sciences*, *18*(8), 532–550. https://doi.org/10.1002/ELSC.201700200

Klint, J. K., Senff, S., Saez, N. J., Seshadri, R., Lau, H. Y., Bende, N. S., Undheim, E. A. B., Rash, L. D., Mobli, M., & King, G. F. (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). https://doi.org/10.1371/JOURNAL.PONE.0063865

Lange, C., Müller, S. D., Walther, T. H., Bürck, J., & Ulrich, A. S. (2007). Structure analysis of the protein translocating channel TatA in membranes using a multi-construct approach. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1768*(10), 2627–2634. https://doi.org/10.1016/J.BBAMEM.2007.06.021

Lausberg, F., Fleckenstein, S., Kreutzenbeck, P., Fröbel, J., Rose, P., Müller, M., & Freudl, R. (2012). Genetic Evidence for a Tight Cooperation of TatB and TatC during Productive Recognition of Twin-Arginine (Tat) Signal Peptides in Escherichia coli. *PLOS ONE*, *7*(6), e39867. https://doi.org/10.1371/JOURNAL.PONE.0039867

Leake, M. C., Greene, N. P., Godun, R. M., Granjon, T., Buchanan, G., 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. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(40), 15376–15381. https://doi.org/10.1073/PNAS.0806338105

Lee, J. J., Kong, J. N., Do, H. D., Jo, D. H., & Kong, K. H. (2010). Design and Efficient Soluble Expression of a Sweet Protein, Brazzein and Minor-Form Mutant. *Bulletin of the Korean Chemical Society*, *31*(12), 3830–3833. https://doi.org/10.5012/BKCS.2010.31.12.3830

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. *Journal of Bacteriology*, *184*(21), 5871–5879. https://doi.org/10.1128/JB.184.21.5871-5879.2002

Lens, J., & Evertzen, A. (1952). The difference between insulin from cattle and from pigs. *Biochimica et Biophysica Acta*, *8*(C), 332–338. https://doi.org/10.1016/0006-3002(52)90048-6

Linton, E., Walsh, M. K., Sims, R. C., & Miller, C. D. (2012). Translocation of green fluorescent protein by comparative analysis with multiple signal peptides. *Biotechnology Journal*, *7*(5), 667–676. https://doi.org/10.1002/BIOT.201100158

Lobstein, J., Emrich, C. A., Jeans, C., Faulkner, M., Riggs, P., & Berkmen, M. (2012). SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microbial Cell Factories*, *11*(1), 1–16. https://doi.org/10.1186/1475-2859-11-56

Low, K. O., Mahadi, N. M., & Illias, R. M. (2013). Optimisation of signal peptide for recombinant protein secretion in bacterial hosts. *Applied Microbiology and Biotechnology*, *97*(9), 3811–3826. https://doi.org/10.1007/S00253-013-4831-Z

Malherbe, G., Humphreys, D. P., & Davé, E. (2019). A robust fractionation method for protein subcellular localization studies in Escherichia coli. *BioTechniques*, *66*(4), 171–178. https://doi.org/10.2144/BTN-2018-0135

Marisch, K., Bayer, K., Cserjan-Puschmann, M., Luchner, M., & Striedner, G. (2013). Evaluation of three industrial Escherichia coli strains in fed-batch cultivations during high-level SOD protein production. *Microbial Cell Factories*, *12*(1), 1–11. https://doi.org/10.1186/1475-2859-12-58

Matias, V. R. F., & Beveridge, T. J. (2005). Cryo-electron microscopy reveals native polymeric cell wall structure in Bacillus subtilis 168 and the existence of a periplasmic space. *Molecular Microbiology*, *56*(1), 240–251. https://doi.org/10.1111/J.1365-2958.2005.04535.X

Matos, C. F. R. O., Branston, S. D., Albiniak, A., Dhanoya, A., Freedman, R. B., 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. *Biotechnology and Bioengineering*, *109*(10), 2533–2542. https://doi.org/10.1002/BIT.24535

Matos, C. F. R. O., Di Cola, A., & Robinson, C. (2009). TatD is a central component of a Tat translocon-initiated quality control system for exported FeS proteins in Escherichia coli. *EMBO Reports*, *10*(5), 474–479. https://doi.org/10.1038/EMBOR.2009.34

Matos, C. F. R. O., Robinson, C., Alanen, H. I., Prus, P., Uchida, Y., Ruddock, L. W., Freedman, R. B., & Keshavarz-Moore, E. (2014). Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in Escherichia coli CyDisCo strains. *Biotechnology Progress*, *30*(2), 281–290. https://doi.org/10.1002/BTPR.1858

Matos, C. F. R. O., Robinson, C., & Di Cola, A. (2008). The Tat system proofreads FeS protein substrates and directly initiates the disposal of rejected molecules. *The EMBO Journal*, *27*(15), 2055–2063. https://doi.org/10.1038/EMBOJ.2008.132

Méjean, V., Lobbi-Nivol, C., Lepelletier, M., Giordano, G., Chippaux, M., & Pascal, M. -C. (1994). TMAO anaerobic respiration in Escherichia coli: involvement of the tor operon. *Molecular Microbiology*, *11*(6), 1169–1179. https://doi.org/10.1111/J.1365-2958.1994.TB00393.X

Mergulhão, F. J. M., Summers, D. K., & Monteiro, G. A. (2005). Recombinant protein secretion in Escherichia coli. *Biotechnology Advances*, *23*(3), 177–202. https://doi.org/10.1016/J.BIOTECHADV.2004.11.003

Mirzadeh, K., Shilling, P. J., Elfageih, R., Cumming, A. J., Cui, H. L., Rennig, M., Nørholm, M. H. H., & Daley, D. O. (2020). Increased production of periplasmic proteins in Escherichia coli by directed evolution of the translation initiation region. *Microbial Cell Factories*, *19*(1), 1–12. https://doi.org/10.1186/S12934-020-01339-8

Müller, D., Bayer, K., & Mattanovich, D. (2006). Potentials and limitations of prokaryotic and eukaryotic expression systems for recombinant protein production – a comparative view. *Microbial Cell Factories 2006 5:1*, *5*(1), 1–2. https://doi.org/10.1186/1475-2859-5-S1-P61

Müntjes, K., Philipp, M., Hüsemann, L., Heucken, N., Weidtkamp-Peters, S., Schipper, K., Zurbriggen, M. D., & Feldbrügge, M. (2020). Establishing Polycistronic Expression in the Model Microorganism Ustilago maydis. *Frontiers in Microbiology*, *11*, 1384. https://doi.org/10.3389/FMICB.2020.01384/BIBTEX

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*, *1778*(9), 1735–1756. https://doi.org/10.1016/J.BBAMEM.2007.07.015

Natarajan, A., Jaroentomeechai, T., Cabrera-Sánchez, M., Mohammed, J. C., Cox, E. C., Young, O., Shajahan, A., Vilkhovoy, M., Vadhin, S., Varner, J. D., Azadi, P., & DeLisa, M. P. (2020). Engineering orthogonal human O-linked glycoprotein biosynthesis in

bacteria. *Nature chemical biology*, *16*(10), 1062-1070. https://doi.org/10.1038/s41589-020-0595-9

Nilsson, J., Jonasson, P., Samuelsson, E., Ståhl, S., & Uhlén, M. (1996). Integrated production of human insulin and its C-peptide. *Journal of Biotechnology*, *48*(3), 241–250. https://doi.org/10.1016/0168-1656(96)01514-3

Nothaft, H., & Szymanski, C. M. (2010). Protein glycosylation in bacteria: sweeter than ever. *Nature Reviews. Microbiology*, *8*(11), 765–778. https://doi.org/10.1038/NRMICRO2383

Orfanoudaki, G., & Economou, A. (2014). Proteome-wide subcellular topologies of E. coli polypeptides database (STEPdb). *Molecular and Cellular Proteomics*, *13*(12), 3674–3687. https://doi.org/10.1074/MCP.O114.041137

Palmer, I., & Wingfield, P. T. (2004). Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from Escherichia coli. *Current Protocols in Protein Science*, *38*(1), 6.3.1-6.3.18. https://doi.org/10.1002/0471140864.PS0603S38

Palmer, T., & Berks, B. C. (2012). The twin-arginine translocation (Tat) protein export pathway. *Nature Reviews Microbiology 2012 10:7*, *10*(7), 483–496. https://doi.org/10.1038/nrmicro2814

Pan, S. H., & Malcolm, B. A. (2000). Reduced background expression and improved plasmid stability with pET vectors in BL21 (DE3). *BioTechniques*, *29*(6), 1234–1238. https://doi.org/10.2144/00296ST03

Patel, R., Smith, S. M., & Robinson, C. (2014). Protein transport by the bacterial Tat pathway. *Biochimica et Biophysica Acta*, *1843*(8), 1620–1628. https://doi.org/10.1016/J.BBAMCR.2014.02.013

Pham, J. V., Yilma, M. A., Feliz, A., Majid, M. T., Maffetone, N., Walker, J. R., Kim, E., Cho, H. J., Reynolds, J. M., Song, M. C., Park, S. R., & Yoon, Y. J. (2019). A review of the microbial production of bioactive natural products and biologics. *Frontiers in Microbiology*, *10*(JUN), 1404. https://doi.org/10.3389/FMICB.2019.01404

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. *Journal of Biotechnology*, *58*(1), 1–11. https://doi.org/10.1016/S0168-1656(97)00116-8

Poirier, N., Roudnitzky, N., Brockhoff, A., Belloir, C., Maison, M., Thomas-Danguin, T., Meyerhof, W., & Briand, L. (2012). Efficient production and characterization of the sweet-tasting brazzein secreted by the yeast pichia pastoris. *Journal of Agricultural and Food Chemistry*, *60*(39), 9807–9814. https://doi.org/10.1021/JF301600M

Prahlad, J., Struble, L. R., Lutz, W. E., Wallin, S. A., Khurana, S., Schnaubelt, A., Broadhurst, M. J., Bayles, K. W., & Borgstahl, G. E. O. (2021). CyDisCo production of functional recombinant SARS-CoV-2 spike receptor binding domain. *Protein Science*, *30*(9), 1983–1990. https://doi.org/10.1002/PRO.4152

Prinz, W. A., Åslund, F., Holmgren, A., & Beckwith, J. (1997). The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm. *The Journal of Biological Chemistry*, *272*(25), 15661–15667. https://doi.org/10.1074/JBC.272.25.15661

Pugsley, A. P. (1993). The complete general secretory pathway in gram-negative bacteria. *Microbiological Reviews*, *57*(1), 50–108. https://doi.org/10.1128/MR.57.1.50-108.1993

Pylypenko, O., Welz, T., Tittel, J., Kollmar, M., Chardon, F., Malherbe, G., Weiss, S., Michel, C. I. L., Samol-Wolf, A., Grasskamp, A. T., Hume, A., Goud, B., Baron, B., England, P., Titus, M. A., Schwille, P., Weidemann, T., Houdusse, A., & Kerkhoff, E. (2016). Coordinated recruitment of spir actin nucleators and myosin V motors to rab11 vesicle membranes. *ELife*, *5*(September). https://doi.org/10.7554/ELIFE.17523

Qamsari, E. S., Sharifzadeh, Z., Bagheri, S., Riazi-Rad, F., Younesi, V., Abolhassani, M., Ghaderi, S. S., Baradaran, B., Somi, M. H., & Yousefi, M. (2017). Isolation and characterization of anti c-met single chain fragment variable (scFv) antibodies. *Journal of Immunotoxicology*, *14*(1), 23–30. https://doi.org/10.1080/1547691X.2016.1251512

Ramazi, S., & Zahiri, J. (2021). Posttranslational modifications in proteins: resources, tools and prediction methods. *Database : The Journal of Biological Databases and Curation*, *2021*. https://doi.org/10.1093/DATABASE/BAAB012

Randall, L. L., & Hardy, S. 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*, *46*(6), 921–928. https://doi.org/10.1016/0092-8674(86)90074-7

Ren, G., Ke, N., & Berkmen, M. (2016). Use of the SHuffle strains in production of proteins. *Current Protocols in Protein Science*, *85*(1), 5-26. https://doi.org/10.1002/cpps.11

Rettenbacher, L. A., Arauzo-Aguilera, K., Buscajoni, L., Castillo-Corujo, A., Ferrero-Bordera, B., Kostopoulou, A., Moran-Torres, R., Núñez-Nepomuceno, D., Öktem, A., Palma, A., Pisent, B., Puricelli, M., Schilling, T., Tungekar, A. A., Walgraeve, J., Humphreys, D., von der Haar, T., Gasser, B., Mattanovich, D., Ruddock, L., & van Dijl, J. M. (2021). Microbial protein cell factories fight back? *Trends in Biotechnology 40*(5), 576–590. https://doi.org/10.1016/j.tibtech.2021.10.003

Reuten, R., Nikodemus, D., Oliveira, M. B., Patel, T. R., Brachvogel, B., Breloy, I., Stetefeld, J., & Koch, M. (2016). Maltose-Binding Protein (MBP), a Secretion-Enhancing Tag for Mammalian Protein Expression Systems. *PLoS ONE*, *11*(3). https://doi.org/10.1371/JOURNAL.PONE.0152386

Richter, S., & Brüser, T. (2005). Targeting of Unfolded PhoA to the TAT Translocon of Escherichia coli. *Journal of Biological Chemistry*, *280*(52), 42723–42730. https://doi.org/10.1074/JBC.M509570200

Rinas, U., & Hoffmann, F. (2004). Selective Leakage of Host-Cell Proteins during High-Cell-Density Cultivation of Recombinant and Non-recombinant Escherichia coli. *Biotechnology Progress*, *20*(3), 679–687. https://doi.org/10.1021/BP034348K

Rocco, M. A., Waraho-Zhmayev, D., & DeLisa, M. P. (2012). Twin-arginine translocase mutations that suppress folding quality control and permit export of misfolded substrate proteins. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(33), 13392–13397. https://doi.org/10.1073/PNAS.1210140109

Rodrigue, A., Chanal, A., Beck, K., Müller, M., & Wu, L. F. (1999). Co-translocation of a periplasmic enzyme complex by a hitchhiker mechanism through the bacterial tat pathway. *The Journal of Biological Chemistry*, *274*(19), 13223–13228. https://doi.org/10.1074/JBC.274.19.13223

Rodriguez, F., Rouse, S. L., Tait, C. E., Harmer, J., De Riso, A., Timmel, C. R., Sansom, M. S. P., Berks, B. C., & Schnell, J. R. (2013). Structural model for the protein-translocating element of the twin-arginine transport system. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(12), E1092–E1101. https://doi.org/10.1073/PNAS.1219486110

Roifman, C. M., Mills, G. B., Chu, M., & Gelfand, E. W. (1985). Functional comparison of recombinant interleukin 2 (IL-2) with IL-2-containing preparations derived from cultured cells. *Cellular Immunology*, *95*(1), 146–156. https://doi.org/10.1016/0008-8749(85)90303-X

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*, *5*(APR). https://doi.org/10.3389/FMICB.2014.00172

Rosano, G. L., Morales, E. S., & Ceccarelli, E. A. (2019). New tools for recombinant protein

production in Escherichia coli: A 5-year update. *Protein Science : A Publication of the Protein Society*, *28*(8), 1412. https://doi.org/10.1002/PRO.3668

Rose, P., Fröbel, J., Graumann, P. L., & Müller, M. (2013). Substrate-Dependent Assembly of the Tat Translocase as Observed in Live Escherichia coli Cells. *PLOS ONE*, *8*(8), e69488. https://doi.org/10.1371/JOURNAL.PONE.0069488

Sachelaru, I., Petriman, N. A., Kudva, R., & Koch, H. G. (2014). Dynamic Interaction of the Sec Translocon with the Chaperone PpiD. *The Journal of Biological Chemistry*, *289*(31), 21706. https://doi.org/10.1074/JBC.M114.577916

Sanchez-Garcia, L., Martín, L., Mangues, R., Ferrer-Miralles, N., Vázquez, E., & Villaverde, A. (2016). Recombinant pharmaceuticals from microbial cells: A 2015 update. *Microbial Cell Factories*, *15*(1), 1–7. https://doi.org/10.1186/S12934-016-0437-3

Sandomenico, A., Sivaccumar, J. P., & Ruvo, M. (2020). Evolution of *Escherichia coli* Expression System in Producing Antibody Recombinant Fragments. *International Journal of Molecular Science*, *21*(17), 6324. https://doi.org/10.3390/ijms21176324

Sanford, M. (2014). Dulaglutide: First global approval. *Drugs*, *74*(17), 2097–2103. https://doi.org/10.1007/S40265-014-0320-7

Santini, C. L., Ize, B., Chanal, A., Müller, M., Giordano, G., & Wu, L. F. (1998). A novel Sec-independent periplasmic protein translocation pathway in Escherichia coli. *The EMBO Journal*, *17*(1), 101–112. https://doi.org/10.1093/EMBOJ/17.1.101

Sargent, F., Berks, B. C., & Palmer, T. (2002). Assembly of membrane-bound respiratory complexes by the Tat protein-transport system. *Archives of Microbiology*, *178*(2), 77–84. https://doi.org/10.1007/S00203-002-0434-2

Sargent, F., Berks, B. C., & Palmer, T. (2005). The Twin-Arginine Transport System. *Protein Movement Across Membranes*, 71–84. https://doi.org/10.1007/0-387-30871-7_6

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. *The EMBO Journal*, *17*(13), 3640–3650. https://doi.org/10.1093/EMBOJ/17.13.3640

Shiloach, J., & Fass, R. (2005). Growing E. coli to high cell density—A historical perspective on method development. *Biotechnology Advances*, *23*(5), 345–357. https://doi.org/10.1016/J.BIOTECHADV.2005.04.004

Shriver-Lake, L. C., Goldman, E. R., Zabetakis, D., & Anderson, G. P. (2017). Improved production of single domain antibodies with two disulfide bonds by co-expression of chaperone proteins in the Escherichia coli periplasm. *Journal of Immunological Methods*, *443*, 64–67. https://doi.org/10.1016/J.JIM.2017.01.007

Silhavy, T. J., Kahne, D., & Walker, S. (2010). The Bacterial Cell Envelope. *Cold Spring Harbor Perspectives in Biology*, *2*(5). https://doi.org/10.1101/CSHPERSPECT.A000414

Silhavy, T. J., & Mitchell, A. M. (2019). Genetic Analysis of Protein Translocation. *Protein Journal*, *38*(3), 217–228. https://doi.org/10.1007/S10930-019-09813-Y

Sohail, A. A., Gaikwad, M., Khadka, P., Saaranen, M. J., & Ruddock, L. W. (2020). Production of Extracellular Matrix Proteins in the Cytoplasm of E. coli: Making Giants in Tiny Factories. *International Journal of Molecular Sciences 2020, Vol. 21, Page 688*, *21*(3), 688. https://doi.org/10.3390/IJMS21030688

Sørensen, H. P., & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories*, *4*, 1. https://doi.org/10.1186/1475-2859-4-1

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. *The Journal of Biological Chemistry*, *275*(16), 11591–11596. https://doi.org/10.1074/JBC.275.16.11591

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. *Molecular Microbiology*, *43*(4), 1005–1021. https://doi.org/10.1046/J.1365-2958.2002.02797.X

Stolle, P., Hou, B., & Brüser, T. (2016). The Tat substrate CueO is transported in an incomplete folding state. *Journal of Biological Chemistry*, *291*(26), 13520–13528. https://doi.org/10.1074/jbc.M116.729103

Studier, F. W. (2005). *Protein production by auto-induction in high-density shaking cultures*. https://doi.org/10.1016/j.pep.2005.01.016

Studier, F. W., Daegelen, P., Lenski, R. E., Maslov, S., & Kim, J. F. (2009). Understanding the Differences between Genome Sequences of Escherichia coli B Strains REL606 and BL21(DE3) and Comparison of the E. coli B and K-12 Genomes. *Journal of Molecular Biology*, *394*(4), 653–680. https://doi.org/10.1016/J.JMB.2009.09.021

Suominen, I., Meyer, P., Tilgmann, C., Glumoff, T., Glumoff, V., Kapyla, J., & Mantsala, P. (1995). Effects of signal peptide mutations on processing of Bacillus stearothermophilus α-amylase in Escherichia coli. *Microbiology*, *141*(3), 649–654. https://doi.org/10.1099/13500872-141-3-649

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., Tuffery, P., Derreumaux, P., Leslie Dutton, P., Robinson, C., Hitchcock, A., & Neil Hunter, C. (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), 6672–6681. https://doi.org/10.1074/JBC.RA117.000880

Tarry, M. J., Schäfer, E., Chen, S., Buchanan, G., Greene, N. P., Lea, S. M., Palmer, T., Saibil, H. R., & Berks, B. C. (2009). Structural analysis of substrate binding by the TatBC component of the twin-arginine protein transport system. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(32), 13284–13289. https://doi.org/10.1073/PNAS.0901566106

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology*, *72*(2), 211–222. https://doi.org/10.1007/s00253-006-0465-8

Thomas, J. D., Daniel, R. A., Errington, J., & Robinson, C. (2001). Export of active green fluorescent protein to the periplasm by the twin-arginine translocase (Tat) pathway in Escherichia coli. *Molecular Microbiology*, *39*(1), 47–53. https://doi.org/10.1046/j.1365-2958.2001.02253.x

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), 1185–1196. https://doi.org/10.1046/J.1365-2958.1996.651436.X

Tsirigotaki, A., De Geyter, J., Šoštarić, N., Economou, A., & Karamanou, S. (2016). Protein export through the bacterial Sec pathway. *Nature Reviews Microbiology 2016 15:1*, *15*(1), 21–36. https://doi.org/10.1038/nrmicro.2016.161

Tsukazaki, T., Mori, H., Echizen, Y., Ishitani, R., Fukai, S., Tanaka, T., Perederina, A., Vassylyev, D. G., Kohno, T., Maturana, A. D., Ito, K., & Nureki, O. (2011). Structure and function of a membrane component SecDF that enhances protein export. *Nature 2011 474:7350*, *474*(7350), 235–238. https://doi.org/10.1038/nature09980

Tullman-Ercek, D., DeLisa, M. P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., & Georgiou, G. (2007). Export pathway selectivity of Escherichia coli twin arginine translocation signal peptides. *The Journal of Biological Chemistry*, *282*(11), 8309–8316. https://doi.org/10.1074/JBC.M610507200

Tungekar, A. A., Castillo-Corujo, A., & Ruddock, L. W. (2021). So you want to express your protein in Escherichia coli? *Essays in Biochemistry*, *65*(2), 247–260. https://doi.org/10.1042/EBC20200170

Turner, R. J., Papish, A. L., & Sargent, F. (2004). Sequence analysis of bacterial redox enzyme maturation proteins (REMPs). *Https://Doi.Org/10.1139/W03-117*, *50*(4), 225–238. https://doi.org/10.1139/W03-117

Ukkonen, K., Veijola, J., Vasala, A., & Neubauer, P. (2013). Effect of culture medium, host strain and oxygen transfer on recombinant Fab antibody fragment yield and leakage to medium in shaken E. coli cultures. *Microbial Cell Factories*, *12*(1), 1–14. https://doi.org/10.1186/1475-2859-12-73/FIGURES/7

Ullers, R. S., Ang, D., Schwager, F., Georgopoulos, C., & Genevaux, P. (2007). Trigger factor can antagonize both SecB and DnaK/DnaJ chaperone functions in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3101–3106. https://doi.org/10.1073/PNAS.0608232104/SUPPL_FILE/08232TABLE1.PDF

Unden, G., Becker, S., Bongaerts, J., Schirawski, J., & Six, S. (1994). Oxygen regulated gene expression in facultatively anaerobic bacteria. *Antonie van Leeuwenhoek 1994 66:1*, *66*(1), 3–22. https://doi.org/10.1007/BF00871629

Valderrama-Rincon, J. D., Fisher, A. C., Merritt, J. H., Fan, Y. Y., Reading, C. A., Chhiba, K., Heiss, C., Azadi, P., Aebi, M., & DeLisa, M. P. (2012). An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nature Chemical Biology 2012 8:5*, *8*(5), 434–436. https://doi.org/10.1038/nchembio.921

van Dijl, J. M., Braun, P. G., Robinson, C., Quax, W. J., Antelmann, H., Hecker, M., Rg Mü Ller D, J., Tjalsma, H., Bron, S., & Jongbloed, J. D. H. (2002). Functional genomic analysis of the Bacillus subtilis Tat pathway for protein secretion. *Journal of Biotechnology*, *98*, 243–254. https://doi.org/10.1016/s0168-1656(02)00135-9

Vasala, A., Panula, J., Bollók, M., Illmann, L., Hälsig, C., & Neubauer, P. (2006). A new wireless system for decentralised measurement of physiological parameters from shake flasks. *Microbial Cell Factories*, *5*(1), 1–6. https://doi.org/10.1186/1475-2859-5-8

Vecchio, I., Tornali, C., Bragazzi, N. L., & Martini, M. (2018). The Discovery of Insulin: An Important Milestone in the History of Medicine. *Frontiers in Endocrinology*, *9*, 613. https://doi.org/10.3389/FENDO.2018.00613

Wacker, M., Linton, D., Hitchen, P. G., Nita-Lazar, M., Haslam, S. M., North, S. J., Panico, M., Morris, H. R., Dell, A., Wren, B. W., & Aebi, M. (2002). N-Linked Glycosylation in Campylobacter jejuni and Its Functional Transfer into E. coli. *Science*, *298*(5599), 1790–1793. https://doi.org/10.1126/SCIENCE.298.5599.1790

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 . *Pharmaceutical Bioprocessing*, *3*(6), 387–396. https://doi.org/10.4155/PBP.15.21

Weiner, J. H., Bilous, P. T., Shaw, G. M., Lubitz, S. P., Frost, L., Thomas, G. H., Cole, J. A., & Turner, R. J. (1998). A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins. *Cell*, *93*(1), 93–101. https://doi.org/10.1016/S0092-8674(00)81149-6

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. *Journal of Biological Chemistry*, *275*(22), 16717–16722. https://doi.org/10.1074/JBC.M000800200

Wiedemann, C., Kumar, A., Lang, A., & Ohlenschläger, O. (2020). Cysteines and Disulfide Bonds as Structure-Forming Units: Insights From Different Domains of Life and the Potential for Characterization by NMR. *Frontiers in Chemistry*, *8*, 280. https://doi.org/10.3389/FCHEM.2020.00280

Wong, C. H. (2005). Protein glycosylation: New challenges and opportunities. *Journal of Organic Chemistry*, *70*(11), 4219–4225. https://doi.org/10.1021/JO050278F

Wong, W. K. R., & Sutherland, M. L. (1993). Excretion of heterologous protein from E. coli. United States patent application US19890395797 19890818.

Wood, W. N., Smith, K. D., Ream, J. A., & Kevin Lewis, L. (2017). Enhancing yields of low and single copy number plasmid DNAs from Escherichia coli cells. *Journal of Microbiological Methods*, *133*, 46–51. https://doi.org/10.1016/J.MIMET.2016.12.016

Yahr, T. L., & Wickner, W. T. (2001). Functional reconstitution of bacterial Tat translocation in vitro. *The EMBO Journal*, *20*(10), 2472. https://doi.org/10.1093/EMBOJ/20.10.2472

Yamaguchi, H., & Miyazaki, M. (2014). Refolding Techniques for Recovering Biologically Active Recombinant Proteins from Inclusion Bodies. *Biomolecules*, *4*, 235–251. https://doi.org/10.3390/biom4010235

Yang, Z., Zhang, L., Zhang, Y., Zhang, T., & Feng, Y. (2011). Highly Efficient Production of Soluble Proteins from Insoluble Inclusion Bodies by a Two-Step-Denaturing and Refolding Method. *PLoS ONE*, *6*(7), 22981. https://doi.org/10.1371/journal.pone.0022981

Yoon, S. H., Han, M. J., Jeong, H., Lee, C. H., Xia, X. X., Lee, D. H., Shim, J. H., Lee, S. Y., Oh, T. K., & Kim, J. F. (2012). Comparative multi-omics systems analysis of Escherichia coli strains B and K-12. *Genome Biology*, *13*(5), 1–13. https://doi.org/10.1186/GB-2012-13-5-R37

Yu, F., Yamada, H., Daishima, K., & Mizushima, S. (1984). Nucleotide sequence of the lspA gene, the structural gene for lipoprotein signal peptidase of Escherichia coli. *FEBS Letters*, *173*(1), 264–268. https://doi.org/10.1016/0014-5793(84)81060-1

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., Dao, P., Cenk Sahinalp, S., Ester, M., Foster, L. J., & Brinkman, F. S. L. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics (Oxford, England)*, *26*(13), 1608–1615. https://doi.org/10.1093/BIOINFORMATICS/BTQ249

Zafar, A., Aftab, M. N., ud Din, Z., Aftab, S., Iqbal, I., & ul Haq, I. (2016). Cloning, Purification and Characterization of a Highly Thermostable Amylase Gene of Thermotoga petrophila into Escherichia coli. *Applied Biochemistry and Biotechnology*, *178*(4), 831–848. https://doi.org/10.1007/S12010-015-1912-8

Zeece, M. (2020). Flavors. *Introduction to the Chemistry of Food*, 213–250. https://doi.org/10.1016/B978-0-12-809434-1.00006-2

Zhang, Y., Hu, Y., Li, H., & Jin, C. (2014a). Structural Basis for TatA Oligomerization: An NMR Study of Escherichia coli TatA Dimeric Structure. *PLOS ONE*, *9*(8), e103157. https://doi.org/10.1371/JOURNAL.PONE.0103157

Zhang, Y., Wang, L., Hu, Y., & Jin, C. (2014b). Solution structure of the TatB component of the twin-arginine translocation system. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, *1838*(7), 1881–1888. https://doi.org/10.1016/J.BBAMEM.2014.03.015

Zhu, M. M., Mollet, M., Hubert, R. S., Kyung, Y. S., & Zhang, G. G. (2017). Industrial Production of Therapeutic Proteins: Cell Lines, Cell Culture, and Purification. *Handbook of Industrial Chemistry and Biotechnology*, 1639–1669. https://doi.org/10.1007/978-3-319-52287-6_29

Annex 1: SECRETERS ITN published review: Microbial protein cell factories fight back?

Trends in Biotechnology

CellPress OPEN ACCESS

Review

Microbial protein cell factories fight back?

SECRETERS - European Union's Horizon 2020 Programme[1,2,*]

The biopharmaceutical market is growing faster than ever, with two production systems competing for market dominance: mammalian cells and microorganisms. In recent years, based on the rise of antibody-based therapies, new biotherapeutic approvals have favored mammalian hosts. However, not only has extensive research elevated our understanding of microbes to new levels, but emerging therapeutic molecules also facilitate their use; thus, is it time for microbes to fight back? In this review, we answer this timely question by cross-comparing four microbial production hosts and examining the innovations made to both their secretion and post-translational modification (PTM) capabilities. Furthermore, we discuss the impact of tools, such as omics and systems biology, as well as alternative production systems and emerging biotherapeutics.

Small cells in an expanding market

Recombinant proteins have dramatically changed our lives, and their market size and impact are projected to keep expanding[i]. Despite the widespread use of recombinant proteins in many industrial sectors, one of the main driving forces for continuous market expansion are biopharmaceuticals, the fastest growing group in the pharmaceutical industry [1]. This has triggered the development of a large spectrum of industrial expression platforms for their production, including both microbial and mammalian cell hosts [2].

The trend in recent years has seen mammalian cell lines increasingly outcompete their microbial counterparts (Figure 1A). From 2014 to mid-2018, more than 87% of the genuinely new biopharmaceutical active ingredients that were released to the market were proteins [3]. Of these, 84% were expressed in mammalian expression systems, with **Chinese hamster ovary** (CHO) (see Glossary) cell-based systems being the most widely used (Box 1). This surge in the biotherapeutics sector can be explained mainly by the increasing dominance of monoclonal antibodies (mAbs), which require humanized **PTMs** [3].

Despite the numerous advantages of mammalian-based protein production, there are also drawbacks (Box 1). Compared with mammalian hosts, microbial expression systems are characterized by being easy to work with, robust, and cost-effective, all highly desirable features in the context of biopharmaceutical production. In fact, microbial platforms are capable of delivering in a scalable and affordable manner a range of functional recombinant therapeutic proteins [4], such as vaccines, hormones, interferons, and growth factors (Figure 1B), as well as nonpharma products, such as industrial enzymes. Nevertheless, these platforms also come with some disadvantages, particularly with respect to their PTM capabilities and secretion.

In the diverse area of scientific developments and engineering strategies, this review follows an integrative approach to summarize and compare the major innovations in the field of microbial recombinant protein production, with a focus on four microbial platforms: *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, and *Pichia pastoris* (syn. *Komagataella* spp.). In this review, we discuss recent scientific progress made in innovation-rich research areas, such as **omics**, systems

Highlights

Microbial production systems for biopharmaceutics have been outcompeted by mammalian systems in recent years, mostly due to the increased demand for antibody-based products. In most other categories, microbes have maintained their dominant position in bioproduction.

The post-translational modification capabilities of *Escherichia coli* have been significantly improved, specifically regarding disulfide bond formation. Recent developments of *Bacillus subtilis* platforms hold promise for biopharmaceutical protein production.

The toolbox for protein glycosylation and secretion in *Saccharomyces cerevisiae* and *Pichia pastoris* has been significantly improved.

Omics and systems biology tools are significantly enhancing our understanding of cellular processes.

Antibody mimetics are promising biotherapeutics that are likely to increase the demand for microbial biopharmaceutical production platforms.

[1]Consortium authors. Consortium author information: Lukas A. Rettenbacher,[3] Klaudia Arauzo-Aguilera,[3] Luisa Buscajoni,[3] Angel Castillo-Corujo,[3] Borja Ferrero-Bordera,[3] Aliki Kostopoulou,[3] Rafael Moran-Torres,[3] David Núñez-Nepomuceno,[3] Ayşegül Öktem,[3] Arianna Palma,[3] Beatrice Pisent,[3] Martina Puricelli,[3] Tobias Schilling,[3] Aatir A. Tungekar,[3] Jonathan Walgraeve,[3] David Humphreys, Tobias von der Haar, Brigitte Gasser, Diethard Mattanovich, Lloyd Ruddock, and Jan Maarten van Dijl.

Affiliations: Lukas A. Rettenbacher, Division of Natural Sciences, School of Biosciences, University of Kent, Canterbury, UK; Klaudia Arauzo-Aguilera, Centre for Molecular

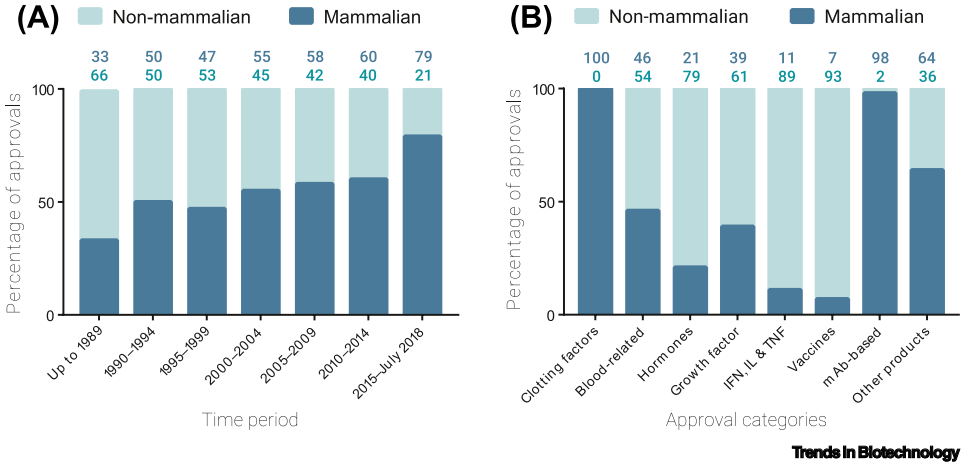


Figure 1. Market share comparison between biotherapeutic production hosts. (A) Approval trend of mammalian versus non-mammalian biopharmaceuticals from 1989 to 2018. (B) Percentage of biopharmaceuticals approved in the USA and EU up until mid-2018, categorized by product type and compared between mammalian and non-mammalian hosts. Both graphs are built on data collected by Gary Walsh [3]. Abbreviations: IFN, inteferon, IL, interleukin; mAb, monoclonal antibody; TNF, tumor necrosis factor.

Box 1. The other side of the story: mammalian systems

CHO cells: advantages

CHO cells are the current number-one choice of host for the production of antibodies and other complex biopharmaceutical proteins [3]. Compared with similar mammalian cell lines, they grow in chemically defined media (as opposed to serum-based media, which can give rise to batch-to-batch variation issues), reach relatively high cell densities in suspension cultures, and have a lower risk of contracting human viruses [80,81]. Over the past two decades, major improvements in media composition, production strategies, and cell line development resulted in the achievement of higher cell densities and product titers, leading to an overall reduction in production costs [82]. Moreover, CHO cells can produce proteins with a range of PTMs, including complex disulfide patterns. Specifically in the realm of human-like glycosylation, they currently have significant advantages over microbes [83].

CHO cells: disadvantages

CHO cells are derived from immortalized cell lines and, as such, they face genome stability issues. This can result in chromosome rearrangements, gene silencing, and gene loss, which, in turn, may lead to increased susceptibility to cellular stress and reduced product titers [84]. These karyotypic variations make long-term fermentations difficult and can impact product quality via PTM-associated heterogeneities [85]. Specifically, *N*-glycan heterogeneity is a complex area of science invoking the use of specialized analytical methods and tools [86].

Over the past 2 years, our ability to tackle global health crises via vaccines has been impressively demonstrated. However, resolving other global health issues, such as HIV infections or Alzheimer's disease, could require completely new protein-based drugs and long-term treatments, calling for even greater efforts and production scales. Addressing these kinds of challenge requires production scales and resource efficiency that will be difficult to achieve in an economically feasible way using CHO technology [87,88].

Another point rarely discussed is the significant lack of knowledge exchange with regards to the various CHO strains used in the pharmaceutical industry. CHO cells have been used for over 50 years and, in part due to their unstable genome, many 'quasispecies' of CHO have emerged [89]. However, since these constitute a certain production advantage for some manufacturers, they are kept hidden from the community, restricting potential innovations for the entire biotech industry.

Non-CHO, non-microbial alternatives

In the vast biotechnology market, there are many more successful platforms besides microbes and CHO cells. However, since they are beyond the scope of this review, we point the interested reader toward key reviews on protein production in human cell lines [90], insect cells [91], and plants [92].

Processing, School of Biosciences, University of Kent, Canterbury, UK; Luisa Buscajoni, Department of Biotechnology, BOKU University of Natural Resources and Life Sciences, Vienna, Austria and Process Science Department, Boehringer Ingelheim RCV GmbH & CO KG, Vienna, Austria; Angel Castillo-Corujo, Protein and Structural Biology Research Unit, Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu 90220, Finland; Borja Ferrero-Bordera, Department of Microbial Proteomics, Institute of Microbiology, Universität Greifswald, Greifswald, Germany; Aliki Kostopoulou, Department of Biotechnology, BOKU University of Natural Resources and Life Sciences, Vienna, Austria and Austrian Centre of Industrial Biotechnology (ACIB), Vienna, Austria; Rafael Moran-Torres, Theoretical Biophysics, Humboldt-Universität zu Berlin, Berlin, Germany; David Núñez-Nepomuceno, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany; Ayşegül Öktem, Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, The Netherlands; Arianna Palma, Department of Biotechnology, BOKU University of Natural Resources and Life Sciences, Vienna, Austria and Austrian Centre of Industrial Biotechnology (ACIB), Vienna, Austria; Beatrice Pisent, Centre for Molecular Processing, School of Biosciences, University of Kent, Canterbury, UK and Celltech R&D Limited, Discovery Science, Slough, UK; Martina Puricelli, Department of Biotechnology, BOKU University of Natural Resources and Life Sciences, Vienna, Austria and Microbial Research & Development, Lonza Pharma & Biotech, Visp, Switzerland; Tobias Schilling, Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, The Netherlands; Aatir A. Tungekar, Protein and Structural Biology Research Unit, Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu 90220, Finland; Jonathan Walgraeve, Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, The Netherlands and Molecular Biology department, AB Enzymes, Darmstadt, Germany; David Humphreys, Celltech R&D Limited, Discovery Science, Slough, UK; Tobias von der Haar, Division of Natural Sciences, School of Biosciences, University of Kent, Canterbury, UK; Brigitte Gasser, Department of Biotechnology, BOKU University of Natural Resources and Life

biology, protein secretion, and PTMs, and place them into the wider context of the biopharmaceutical market. By examining not only recent improvements, but also competing production methods and organisms, we predict the future roles of microbes in the pharmaceutical market.

Current and future proteins of interest

Antibody formats and mimetics

Full-length mAbs represent a large part of the global biopharmaceutical market. Due to their high complexity, they are mostly produced in mammalian cell lines, which require long processing times and elevated production costs. More recently, we have seen a shift toward the production of antibody fragments, among which **single-chain variable fragments** (scFvs) and **fragmented antigen-binding** (Fabs) are the most exploited [5]. These small antibody-based formats have several advantages and, due to the lack of requirement for glycosylation, can be expressed in microbial platforms. In 2017, Gupta and Shukla reviewed an exhaustive survey of expression technologies for the production of antibody-like molecules in microbes, particularly in *E. coli* [6]. In the same year, another review covered antibody fragments and, more in general, biopharmaceuticals production in *Bacillus* strains [7]. Yeast species, notably *S. cerevisiae* and *P. pastoris*, are also used as hosts for the expression of antibody formats, such as scFvs, single-domain antibodies (vHHs), and Fabs [8,9]. In addition to the four hosts discussed here, some promising alternative microorganisms have also made significant attempts to enter the antibody formats markets (Box 2).

Antibody-like scaffolds, also known as **antibody mimetics**, are considered to be future alternatives to mAbs. These include adnectins, affibodies, anticalins, avimers, DARPins, fynomers, and Kunitz domains [10,11]. Similar to antibodies, these compounds have high target-binding specificity and affinity, but are characterized by additional benefits, such as a much smaller size, increased thermostability, and low immunogenic potential. Moreover, they can be easily and efficiently produced in microbial host cells: Binz and colleagues reported yields of 15 g/l of DARPins expressed solubly in *E. coli* [12].

Non-antibodies

Some categories of biopharmaceuticals are still heavily dominated by microbial hosts (Figure 1B). These include interferons (e.g., IFNα-2b), cytokines [e.g., granulocyte colony stimulating factor (G-CSF) and tumor necrosis factors (TNFs)], hormones [e.g., insulin glargine and human growth hormone (hGH)], and interleukins (IL-2 and IL-11), all of which can be produced in *E. coli* and yeasts [4,13,14]. High recombinant human serum albumin (rHSA) yields were recently reported (17.5 g/l) in *P. pastoris* by means of medium optimization [15].

Industrial proteins

Due to its efficient secretory production, *P. pastoris* is widely used also for the recombinant production of protein-based polymers [16] and enzymes. Some examples are reviewed by Vieira Gomes and colleagues [17] and Gifre and colleagues [18], the latter of which illustrates an extensive array of enzymes with interest in the feed industry. These include phytases, which are produced recombinantly mainly in *P. pastoris* and *E. coli*. Among the most important industrial enzyme producers are *Bacillus* spp., which are capable of secreting 20–25 g/l of proteins into the culture medium [19]. These enzymes can be applied in numerous industrial applications, such as food, feed, detergent, textile, and waste treatment processes [20].

Titer comparison of selected proteins of interest

Comparing titer numbers between organisms can be difficult because of significant differences both in terms of production process (e.g., fermentation mode, and media composition

Sciences, Vienna, Austria and Austrian Centre of Industrial Biotechnology (ACIB), Vienna, Austria; Diethard Mattanovich,[1] Department of Biotechnology, BOKU University of Natural Resources and Life Sciences, Vienna, Austria and Austrian Centre of Industrial Biotechnology (ACIB), Vienna, Austria; Lloyd Ruddock, Protein and Structural Biology Research Unit, Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu 90220, Finland; Jan Maarten van Dijl, Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, The Netherlands
[2]Website: http://secreters-msca-itn.eu/
[3]These authors contributed equally.

*Correspondence:
l.rettenbacher@kent.ac.uk
(L. Rettenbacher).

Box 2. Alternative microbial hosts

Microbial protein production is a rich field comprising many more production hosts than the four discussed in this review. *Lactococcus lactis, Pseudomonas* spp., and *Aspergillus niger* are just a few of the many microbes that, over the past decades, have been successfully used in recombinant protein production. However, here we highlight some highly promising alternatives to these better-established organisms.

Myceliophthora thermophila

Myceliophthora thermophila is a filamentous fungus (dubbed C1 by its discoverers) that is one of the biggest potential disrupters of the recombinant protein production industry. As a side-effect of random mutagenesis strain development, C1 adapted a change to its cellular morphology, resulting in reduced fermentation viscosity [118]. This facilitates scalability and sets *M. thermophila* apart from most other filamentous fungi. C1 is already successfully used in cellulase production, but has also shown high potential for pharmaceutical applications, achieving g/l antibody yields [119]. Similar to most filamentous fungi, C1 is unable to produce human-like glycosylation; however, breakthroughs in this area could elevate C1 to the forefront of biopharmaceutical protein production.

Leishmania tarentolae

Leishmania tarentolae, a trypanosomatid protozoan, is a unicellular nonpathogenic parasite of the white-spotted wall gecko. Its high potential as a biotherapeutic production host stems from its innate ability to construct *N*-glycans that are similar to mammalian *N*-glycan structures [120]. Furthermore, *L. tarentolae* glycosylation is homogeneous, making its glycosylated products particularly useful for both studying glycosylation effects and expressing biotherapeutic proteins. GlycoEra (a business unit of LimmaTech) has developed a platform capable of producing glycoengineered mAbs based on *L. tarentolae*, one of the most promising systems for addressing complex glycosylation challenges [121].

Corynebacterium glutamicum

The Gram-positive bacterium *Corynebacterium glutamicum* is a well-established host for the production of amino acids on an industrial scale [122]. Recent advances have been made in its use for recombinant protein production, especially regarding expression constructs and secretion capabilities [123]. Of particular note is the CORYNEX system, which uses *C. glutamicum* and has been shown to produce and secrete high amounts of both human epidermal growth factor and Fabs [123].

Phaeodactylum tricornutum and *Chlamydomonas reinhardtii*

Microalgae are promising hosts for recombinant protein production, particularly due to their comparatively high growth rates and phototropic lifestyle, making them solar powered and, therefore, potentially cost effective. The diatom *Phaeodactylum tricornutum* as well as the single-cell green algae *Chlamydomonas reinhardtii* can both produce bioactive antibodies [124,125] and both show promising secretion levels.

and optimization) and product analytics (including product quality, correct folding, biological activity, and quantification of titers). Despite this difficulty in benchmarking, it is clear that, once microbial systems can produce a protein, they generally outperform CHO cells in terms of volumetric titers (Table 1). For individual proteins of interest (POIs), there can be pronounced differences between the four compared microbes, but there is no clear overall winner.

Omics and systems biology techniques impacting microbes

Omics

In recent years, omics tools, such as genomics, transcriptomics, proteomics, metabolomics, and fluxomics, have been increasingly used to expand our understanding of cell physiology and stress resistance, together with improving existing industrial biotech processes. Currently, genomics is widely implemented in strain development because it provides fundamental information about the DNA sequence, gene functions, DNA structure, and epigenetics. Proteomics, transcriptomics, and metabolomics identify the molecular phenotype of the cell and are used for monitoring media and process development.

As omics tools have become more accessible, it is increasingly feasible to combine them into multi-omics-based research approaches. For example, research by Kohlstedt and colleagues

Glossary

Antibody mimetics: proteins or other organic molecules that fulfil similar roles as antibodies (i.e., they are able to bind antigens). Compared with antibodies, they are relatively small molecules (5–20 kDa compared with ~150 kDa) and lack complex PTMs, making them easier to produce.
Chinese hamster ovary (CHO) cells: the current predominant choice for pharmaceutical antibody production. Compared with microorganisms, they are more closely related to humans, which makes their protein production machinery deliver products that are relatively similar to those of humans.
Disulfide bond (DSB): a covalent connection between two sulfur atoms commonly introduced into proteins via the enzymatically catalyzed oxidation of the side chains of two cysteine residues. Introducing this modification into heterologous proteins can be challenging for many microbes specifically if there are multiple cysteines to be connected into disulfides.
Fragmented antigen-binding (Fab): describes the region of an antibody that binds to an antigen. Producing a Fab is often easier rather than producing the antibody as a whole.
Genome-scale metabolic model (GSMM): aims to encapsulate an organism on the genomic scale in terms of a mathematical model.
***N*- or *O*- glycosylation:** glycosylation is a PTM that connects carbohydrates (i.e., sugars) to the amino acid chain of proteins. Depending on where in the protein they are attached, this PTM can be classified into two main groups, namely *N*-linked glycosylation [connected via an asparagine (N) residue] or *O*-linked glycosylated (connected via an oxygen in the side chain of serine or threonine residues).
Omics: a branch of science that aims to encapsulate the entirety of a field. Examples are genomics (the genes of an organism), transcriptomics (the transcription processes of an organism), proteomics (the proteins in an organism), metabolomics (the metabolites in an organism) and fluxomics (the rate of metabolic reactions in an organism).
Post-translational modifications (PTMs): introduce specific chemical modifications to the primary amino acid chain of proteins. These include disulfide bonds and glycosylation and determine

Table 1. Comparative list of reported titers of therapeutic proteins[a]

Protein	Production system	Cell culture process	Site of accumulation	Titer[b]	Refs
hEGF: 3 DSBs	*Bacillus* spp.	Shake flask	Culture medium	360 mg/l[c]	[93]
	Escherichia coli	Fed-batch bioreactor	Culture medium	250 mg/l	[94]
	Saccharomyces cerevisiae	Fed-batch bioreactor	Culture medium	259.2 mg/l	[95]
	Pichia pastoris	Baffled shake flask	Culture medium	2.5 mg/l	[96]
	CHO	–	–	–	–
hGH: 1 DSB	*Bacillus* spp.	Fed-batch bioreactor	Culture medium	497 mg/l	[97]
	E. coli	Fed-batch bioreactor	Periplasm	2390 mg/l[d]	[42]
	S. cerevisiae	Shake flask	Culture medium	0.9 mg/l	[98]
	P. pastoris	Fed-batch bioreactor	Culture medium	640 mg/l	[99]
	CHO	Semi-continuous batch culture in spinner flasks	Culture medium	75 mg/l	[100]
hIFN-α: 2 DSBs; 1 *O*-glycosylation site	*Bacillus spp.*	Shake flask	Culture medium	15 mg /l	[101]
	E. coli	Fed-batch bioreactor	Cytoplasm	300 mg/l[d]	[102]
	S. cerevisiae	Fed-batch bioreactor	Culture medium	276 mg/l	[103]
	P. pastoris	Batch bioreactor	Culture medium	436 mg/l	[38]
	CHO	–	–	–	–
hIL-6: 2 DSBs;1 *N*-glycosylation site	*Bacillus spp.*	Shake flask	Culture medium	200 mg/l[d]	[104]
	E. coli	Fed-batch bioreactor	Cytoplasm	1100 mg/l[d]	[53]
	S. cerevisiae	Batch bioreactor	Culture medium	30 mg/l	[105]
	P. pastoris	Fed-batch bioreactor	Culture medium	170 mg/l[d]	[106]
	CHO	35-mm dishes	Culture medium	1.4 µg/ 10^6 cells/day	[107]
hG-CSF: 2 DSBs; 1 *O*-glycosylation site	*Bacillus* spp.	Batch bioreactor	Culture medium	120 mg/l	[108]
	E. coli	Fed-batch bioreactor	Cytoplasm	4200 mg/l*	[109]
	S. cerevisiae	Fed-batch bioreactor	Culture medium	98 mg/l	[110]
	P. pastoris	Fed-batch bioreactor	Culture medium	35 mg/l	[111]
	CHO	100-mm dishes	Culture medium	90 µg/ 10^6 cells/day	[112]
Insulin: 3 DSBs	*Bacillus spp.*	Batch bioreactor	Culture medium	1000 mg/l	[113]
	E. coli	Fed-batch bioreactor	Cytoplasm	4340 mg/l*	[114]
	S. cerevisiae	Shake flask	Culture medium	79 mg/l	[115]
	P. pastoris	Fed-batch bioreactor	Culture medium	3075 mg/l	[116]
	CHO	T-75 flasks	Culture medium	1.98 ng/ 10^6 cells/day	[117]

[a]For improved comparison, high-density CHO cultures have average cell densities of 2–4×10^6 cells/ml in a suspension-batch mode and 10–15×10^6 cells/ml in a suspension-perfusion mode using stirred-tank bioreactors. DSBs and *N*- and *O*-glycosylation sites for each protein are also listed.
[b]Unless indicated (*), all proteins were obtained in a soluble form.
[c]Produced as inclusion bodies. For a state-of-the-art overview on how IBs are processed, please refer to [135].
[d]Purified titers.

the regulation, function, and stability of proteins.
Resource balance analysis and flux balance analysis (RBA and FBA): mathematical methods that describe metabolic networks. Compared with similar methods, they require fewer input data and are less computationally expensive.
Signal peptide: short peptide sequences usually attached to the N terminus of newly translated proteins. The SP determines which secretory pathway the protein will access. During translocation, the SP is cleaved from the rest of the protein.
Single-chain variable fragment (scFv): fusion proteins between the antigen-binding domains of both the light and heavy chain of antibodies.
Systems biology: uses omics data and computational and mathematical concepts to characterize complex biological systems in their entirety.

combines four omics tools to investigate the core carbon metabolism of *B. subtilis* under different carbon-limiting growth conditions [21]. In the search for CHO alternatives, the MIT-based AltHost Consortium used a combination of genomics and transcriptomics tools to identify *P. pastoris*

strains presenting the highest protein titers and transformation efficiencies [22]. Long and colleagues recently showcased the advantages of combining genomics and fluxomics by demonstrating how metabolic bottlenecks in *E. coli* can be identified and overcome [23]. Both transcriptomics and proteomics were applied to elucidate mechanisms that lead to increased protein production in *E. coli* [24] or to improved mutants of S. cerevisiae [25]. Together, these studies highlight the immense potential of multi-omics-based approaches, and we anticipate that their implementation in the production of biopharmaceutics will massively increase in future years. For further reading on the impact and importance of omics on industrial microbe engineering, we recommend recent reviews by Becker and Wittmann [26] and Subramanian and colleagues [27].

Genome-scale metabolic models

Over the past decade, **genome-scale metabolic models** (GSMMs) have provided a huge variety of new insights into the functioning of microbial cells and, specifically during the past few years, the predictive quality of these models has significantly improved [28].

Both **resource balance analysis** (RBA) and **flux balance analysis** (FBA) have been applied to microbes to predict cell behavior under different growth and stress conditions. RBA based on underlying GSMMs has been shown to deliver accurate quantitative metabolic predictions for both *E. coli* and *B. subtilis,* and it is easily applicable to a variety of other microbial hosts [29]. In *P. pastoris*, the influence of heterologous protein expression was investigated using a GSMM [30]. The resulting model was able to accurately predict effects of overexpression and gene deletions on product yields.

Correct protein folding, as discussed in the PTM section of this review, can be a major limitation to the protein production process. Using a GSMM for *S. cerevisiae*, Qi and colleagues identified bottlenecks in the production of α-amylase [25]. Based on the predictions of the model, they were able to engineer the yeast and double the protein yield. In another approach, Chen and colleagues constructed a genome-scale protein-folding network for *E. coli*, which describes the adaptations of the cells to different heat-stress conditions [31]. Both approaches show how **systems biology** approaches, such as GSMM, can not only increase our understanding of complex cellular processes, but also improve production yields.

What is holding microbes back? PTMs

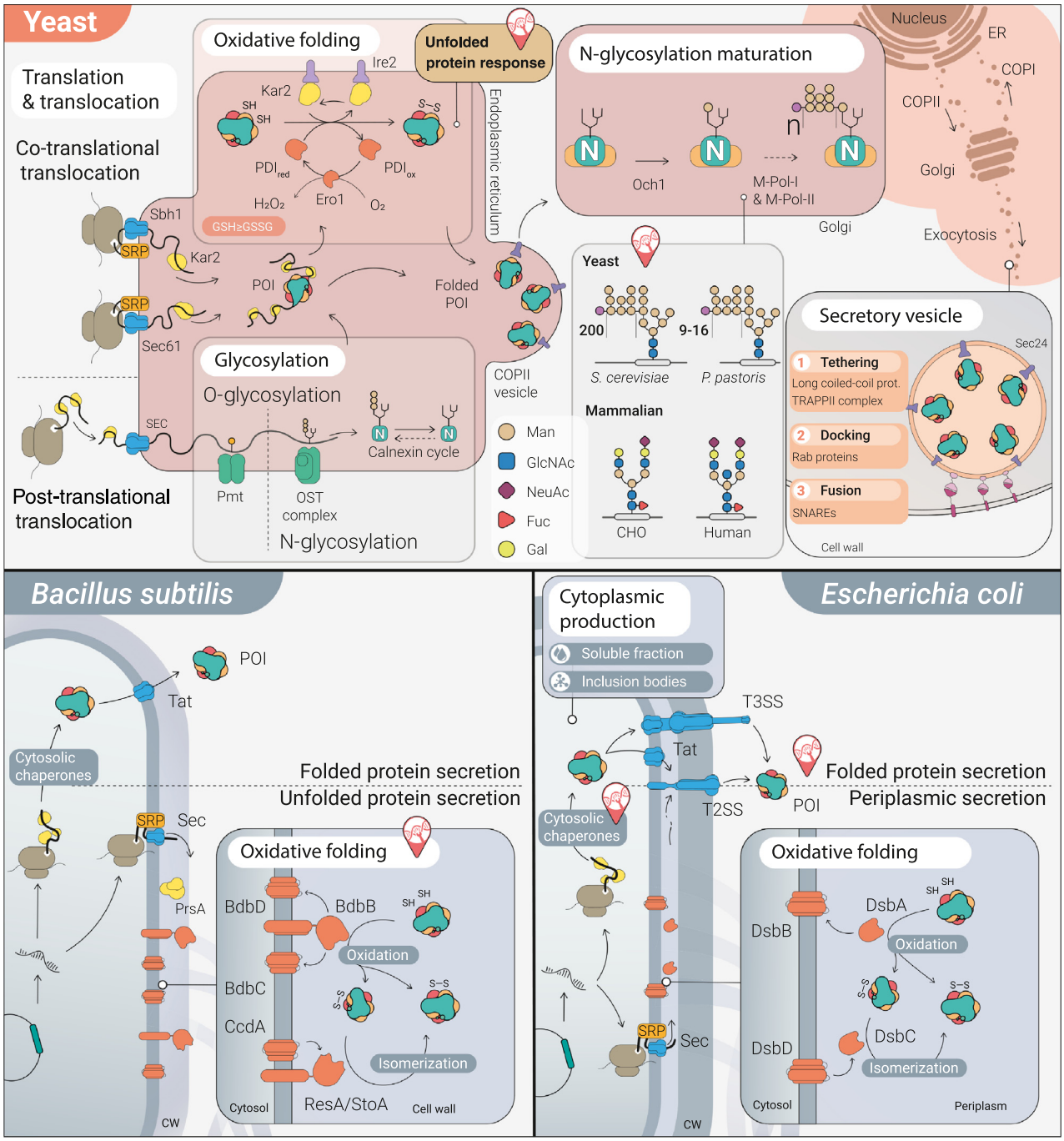
PTMs occur during or after protein synthesis. They change the physicochemical properties and potentially the activity of the protein. They range from small chemical modifications to the amino acid chain, to the addition of complex branching structures on the protein backbone [32,33]. Such modifications can vary immensely in terms of form and function. Many PTMs are involved in metabolic crosstalk, such as phosphorylation and glycosylation, or in improved folding and stability, in which glycosylation and disulfide bonds have a major role, or in signaling for aberrations such as nitrosylation and deamination. The most commonly found PTMs in recombinant proteins are glycosylation and **disulfide bond (**DSB) formation. The endogenous mechanisms carrying out these PTMs are illustrated in Figure 2 (Key figure). While other PTMs, such as methylation and carboxylation exist, glycosylation and DSB formation are frequently required for correct protein folding and biological activity and, therefore, are discussed in more detail below.

Glycosylation (*N*- or *O*-glycosylation)

Glycosylation is a challenging PTM. Its huge variability comes from the number of sugars, attachment sites, sugar chain length, and branching, all of which affect the final product quality and also functionality [34]. For many POIs, proper glycosylation is essential for acquiring correct folding

Key figure

Key cellular machineries involved in protein production in yeasts, *Escherichia coli* and *Bacillus subtilis*



Trends in Biotechnology

(See figure legend at the bottom of the next page.)

conformation. Some sites even display heterogeneity determined by the host and culture conditions, impairing reproducibility of industrial production.

Prokaryotes and lower and higher eukaryotes all have different glycosylation signatures, and natural bacterial pathways are generally *N*-acetylgalactosamine-rich and less branched than their eukaryotic counterparts. *E. coli* does not display natural protein glycosylation; however, attempts have been made by introducing genes of *Campylobacter jejuni* and yeasts to engineer *E. coli* so that it is able to ***N*-glycosylate** [35]. ***O*-glycosylation** was also successfully achieved with *E. coli* through a plasmid-based expression approach encoding the required human genes [36]. In their current state, these engineered cells produce proteins at titers that are not yet economically competitive. However, once optimized, these engineered systems may also offer analytical and clinical advantages due to inherent homogenous glycosylation patterns.

In yeast, both *N*- and *O*-glycosylation machineries are present, but the process of glycan maturation in the Golgi is typically more mannose rich than in higher eukaryotes (Figure 2). Prevention of hypermannosylation and pioneering attempts to 'humanize' the glycoprofiles in *S. cerevisiae* and *P. pastoris* were achieved by a combination of cell engineering and chemical enzymatic treatments [37]. This led to biopharmaceuticals with a more homogeneous human-type Man5GlcNAc2 glycan (e.g., IFNα2b produced in glycoengineered SuperMan5 *P. pastoris* [38]). For *P. pastoris*, even terminally sialylated mammalian-type *N*-glycans were achieved [39]. Glycoengineered yeasts can also efficiently produce therapeutic enzymes for the treatment of lysosomal storage diseases, which require mannose-6-phosphate glycans [40].

Despite successful glycoengineering, glycosylation efficiency still represents a target-specific problem in both bacteria and yeast. Promising modular cell-free systems are now under study to acquire glycosylation by means of *E. coli* cell lysates (Box 3). Such systems can be used to acquire different glycosylation architectures in a controlled manner [41].

Disulfide bonds

Many pharmaceutically relevant proteins are disulfide bonded. While *E. coli* and *B. subtilis* both have endogenous systems for disulfide bond formation in the periplasm and cell wall, respectively (Figure 2), they can struggle with disulfide-rich proteins with complex folding patterns. Nevertheless, it is possible to obtain g/l yields in some cases [42]. By contrast, yeasts have a DSB-forming system located in the endoplasmic reticulum (ER), an environment that is capable of pairing up more complex DSBs for their endogenous proteins. However, a strong unfolded protein response (UPR) when expressing heterologous disulfide-bonded proteins can cause issues [43]. Several attempts at mitigating the UPR response in both *S. cerevisiae* and *P. pastoris* have been made.

Figure 2. Side-by-side comparison of important endogenous cellular functions involved in protein maturation and transport, secretion pathways, oxidative folding, and glycosylation. Cell machineries are represented in different colors according to their function. Secretion machinery components and transporters are shown in blue, oxidative folding elements in orange, chaperones in yellow, and glycosylation elements in green. In addition, points of engineering interest are highlighted for each organism. Top (yeast): unfolded proteins are translocated into the endoplasmic reticulum (ER). Once there, some proteins undergo initial *N*-glycosylation and/or *O*-glycosylation, while elongation of the glycan chain is mainly achieved in the Golgi apparatus. The typical final glycosylation patterns for the two yeasts as well as Chinese hamster ovary (CHO) cells are shown. In the oxidative ER environment, but after the potential initial glycosylation, chaperones, such as Kar2, can act on unfolded proteins and facilitate correct folding and PDI/ERO catalyze disulfide bond formation. Bottom (*E. coli* and *B. subtilis*): both organisms can form disulfide bonds in their oxidative periplasmic environment. Both use the Sec and Tat pathways for secreting unfolded and folded proteins, respectively. Extracellular secretion happens directly in *Bacillus* spp. In *E. coli*, extracellular secretion across the inner and outer membranes occurs via the type II (T2SS) or type III secretion system (T3SS) pathways. Abbreviations: CHO, Chinese hamster ovary; CW, cell wall; DSB, disulfide bond; POI, protein of interest.

Box 3. Cell-free systems as competing production systems

New technologies for manufacturing different proteins are not limited to conventional expression platforms that rely on living cells. CFPS systems have been around for decades and are being improved and exploited as attractive alternatives to cell-based expression. Numerous CFPS crude cell extracts are being developed and/or used, ranging from prokaryotic [126,127] to eukaryotic [128,129] systems.

CFPS: advantages

CFPS offers various advantages: the separation of cell growth and protein synthesis, and the ease of manipulating, monitoring, and optimizing the reaction environment in real-time are the most noteworthy. *In vitro* expression systems are particularly suitable for the production of cytotoxic and membrane proteins [130,131]. Expressing cytotoxic proteins *in vivo* impacts cell viability, leading to increased lysis, stress, and low product titers. By contrast, membrane proteins are inherently hydrophobic and, therefore, difficult to express, harvest, and purify *in vivo* without losing high levels of product to aggregation. Another advantage of CFPS is the (comparative) ease with which unnatural amino acids can be incorporated into proteins [132]. As discussed above, microbes can struggle with glycosylation. CFPS are well suited for manufacturing glycosylated proteins in both pro- and eukaryotic systems [41].

CFPS: disadvantages

While the development of cell-free systems has made significant progress over the past decade, big challenges remain. Even though CFPS are able to achieve protein glycosylation and incorporation of unnatural amino acids, they still struggle with both the efficiency and homogeneity of proteins with these modifications [133]. As mentioned above, there are advantages to CFPS-based membrane protein synthesis, but there are also significant drawbacks. Folding membrane proteins often requires pre-existing membrane structures and those are mostly missing in cell-free systems. Alternative folding environments can be provided, but issues with non-native folding and orientation remain [134]. However, the biggest issues are still cost and scalability [133]. Regeneration or addition of ATP as the primary energy source for protein synthesis remain the biggest economic bottlenecks of cell-free systems.

A common approach for all four microbial systems is the overexpression of chaperones and isomerases, such as Kar2 and PDI, in yeast, the extracellular chaperone PrsA in *B. subtilis*, and periplasmic chaperons DsbC, FkpA, DsbA, and SurA in *E. coli* [44–46]. All of these have resulted in improved protein titers. Furthermore, these overexpressed folding catalysts not only improve protein solubility, but may also facilitate improved quality control because misfolded proteins are either refolded or tagged for proteolytic destruction.

Over the past decade, several different approaches aiming to improve DSB formation in *E. coli* have been researched. Disruption of the native reducing pathways resulted in the Origami strain, capable of forming DSBs in the cytoplasm [47]. Later, the Shuffle strain also incorporated the native periplasmic isomerase DsbC (Figure 2) in the cytoplasm, further improving cytoplasmic DSB formation [48]. Both strains have historically been used in research labs. However, while they both facilitate improved DSB formation in some heterologous proteins, they lack growth fitness and sufficient protein yields for industrial relevance [49]. Similar attempts have been made to enhance DSB formation in proteins secreted by *B. subtilis* but, despite proof of principle, major breakthroughs in terms of industrial bioproduction have not yet been achieved [50].

Owing to its reducing nature, the cytoplasm of *E. coli* does not have any endogenous mechanisms for DSB formation (Figure 2). However, a different approach toward improving DSB formation is based on the cytoplasmic expression of recombinant oxidases and isomerases. This was first established in the CyDisCo strain, which expresses eukaryotic Erv1p and PDI from a plasmid [51]. Over the past decade, this approach has yielded promising results both in terms of DSB complexity and yields. An example of the former is the production of a 44-DSB containing extracellular matrix protein by Sohail and colleagues [52]. A noteworthy yield achievement is the reported production of around 1 g/l of both hGH1 and IL6 [53].

A key area of Improvement: secretion

Signal sequences and process control

The secretory **signal peptide** (SP) is a ubiquitous N-terminal protein-sorting signal that targets proteins for translocation across the ER membrane in eukaryotes and the cytoplasmic membrane in prokaryotes. SPs may need to be optimized for each POI to ensure the best secretion titers. Unoptimized SP–POI pairs can result in protein mistargeting and SP miscleavage, which in turn leads to misfolding, degradation, and an overall reduction in secretion yields [54]. Combinatorial studies and computational methods to infer the most suitable signal sequences are becoming increasingly popular [55]. Computational modeling is widely used to predict the optimal SP–POI combination [56]. Given the size of the possible solution space, high-throughput screenings are becoming a hot topic. A β-lactamase (Bla)-based assay was recently developed for *E. coli*, in which Bla was fused to the C terminus of a scFv enabling the correlation of its activity to periplasmic translocation [57].

The microbial secretion machinery can become overloaded, causing, for example, the product to be partially retained in the cytoplasm. Fine-tuning of the expression machinery can avoid product overload and match expression levels and secretion capacity [58]. Therefore, tools to facilitate expression control are attractive. One example is RiboTite, which allows the modulation of basal protein expression in the absence of, or post, induction in bacteria, thereby protecting the Sec and Tat systems from overloading [59].

Improving secretion in yeasts

Secretion is a distinctive natural feature of yeasts and requires several transport steps across membranes (Figure 2). SPs enable translocation of nascent peptides into the ER, while leader peptides facilitate later trafficking of folded proteins to the extracellular medium via vesicular transport. High secretion yields are usually also associated with strong promoters and elevated gene copy numbers, but imbalanced overproduction may lead to UPR induction and stress. In addition to protein folding, both translocation and vesicular transport were reported to be rate limiting for heterologous protein secretion [60,61]. Successful engineering strategies include alternative signal peptides, promoter engineering, increasing the ER capacity by overexpressing organelle-reshaping factors and chaperones, stress minimization, and enhanced trafficking [62].

Bacterial secretion: Sec and Tat pathways

Sec and Tat are the two distinct, key pathways in bacteria responsible for translocating secretory proteins across the cytoplasmic membrane (Figure 2). The industrial standard is Sec, which transports proteins in an unfolded state, often followed by periplasmic oxidative folding. By contrast, the Tat pathway exports proteins in a folded state (e.g., redox cofactor-binding proteins), which have to reach the periplasm in their mature form to perform their biological function. Tat has a unique proofreading capacity, which favors the export of homogeneously folded products. This proofreading system senses conformational flexibility and can detect the presence of unfolded peptides and transiently exposed inner protein regions [63]. Even though secretion titers are comparatively lower with Tat, it is this quality check feature that makes the pathway interesting for recombinant protein production, particularly biopharmaceuticals. A notable example is given by the W3110 'TatExpress' *E. coli* strain, which has a modified chromosomal *tatABC* operon tuned by IPTG induction. This was successfully used to produce hGH in a homogeneous active state during extended fed-batch fermentations, with titers reaching at least 2.39 g/l culture, highlighting its potential for industrial use [42].

Soil bacteria, such as *B. subtilis* (or *Bacillus licheniformis*), evolved to use a wide spectrum of different substrates by secreting a variety of extracellular macromolecular depolymerases into

their environment via the Sec pathway [19]. The high-capacity secretion system of *B. subtilis* can deliver enzyme yields of up to 25 g per liter culture in industrially optimized processes [64]. While the secretion capabilities of *B. subtilis* are competitive, it is a lack of PTM capabilities that currently limits its exploitation for the pharmaceutical market. In similar fashion to *E. coli*, attempts have been made to use the proofreading Tat pathway for protein secretion in *B. subtilis*. Although attempts to transport GFP in this way remain unsuccessful, examples of protein rerouting into this pathway have improved our understanding of this system [65].

Alternative extracellular transport

Direct transport of POIs into the fermentation broth has the potential to improve downstream processing and can help facilitate continuous production strategies. *E. coli* faces clear limitations with regards to its extracellular secretion capabilities. However, recent progress has been made by using the flagellar type III secretion system (T3SS) secretion pathway. By removing selected flagellar components, the pathway can be repurposed for recombinant protein secretion [66]. Several passive transport strategies have also been developed, particularly during the past decade. One of the most relevant examples is inducing outer-membrane leakiness. Of notable relevance here is the *E. coli* X-press strain [67], which can express a T7 RNAp inhibitor that not only stops cell growth, but also makes the outer membrane leaky without causing significant cell lysis. Excreted protein A and VHH titers in a lab-scale bioreactor setup were reported to reach 349 and 19.6 mg/g dry cell weight (DCW), respectively [68]. Secretion into culture media has also reached commercial reality with controlled 'leaky' mutant forms of *E. coli*, notably lppA marketed as ESETEC strains [69]. Regarding active secretion, autosecretion mechanisms, which are not yet fully understood, are increasingly being studied. In this approach, target proteins are fused to fast-folding tags, which help facilitate extracellular transport. The most noteworthy development in this field is the *E. coli* sfGFP Protein Secretion System (EGPSS) [70]. By fusing target proteins to superfolder GFP, Liu and colleagues were able to successfully autosecrete an immunologically active scFv [71].

Other areas of improvement

A variety of other improvements have been reported for all four microbial systems over the past decade. These are diverse and range from better understanding initiation of translation [72], novel fusion partners, such as Ffu [73], to new promoter systems (e.g., glucose-dependent regulated promoters for *P. pastoris* and its methylotrophic relatives [74]), which avoid the many drawbacks of methanol-induced expression systems [75].

Even after translocation to the periplasm, many POIs, particularly Fabs, still face proteolytic degradation pressure from host cell proteases. To mitigate this issue, Ellis and coworkers developed a protease-deficient *E. coli* strain and combined it with co-expression of the isomerase DsbC, which resulted in a Fab yield of 2.4 g/l [44].

Combining strain and process engineering in *P. pastoris* has yielded promising results, as showcased by two recent examples from Che and colleagues and Bankefa and colleagues [76,77]. In the former, a combination of codon optimization, gene dosage, and fermentation optimization resulted in the production of 12 690 U/ml of fibase, while the latter showed how a mixed-feeding strategy combined with secretory pathway engineering led to a threefold increase in glucose oxidase production to 787.4 U/ml.

Inclusion body (IB) formation has long been a way to overcome problems, such as proteolytic degradation and cell toxicity [78]. However, proteins produced by means of IBs need to be subsequently isolated, solubilized, refolded, and purified to obtain functional products. Current research in this area is focused on developing Process Analytical Technologies (PATs) as tools to

increase the understanding, control, and efficiency of this process. A recent review by Humer and Spadiut summarizes the most important developments in this currently underrated research area [79].

Concluding remarks and future perspectives

Research dedicated to the improvement of microbial production systems has yielded impressive achievements. However, in many cases, microbial systems are still outperformed by mammalian systems, in particular when it comes to pharmaceutical protein production. For a more balanced view on the pros and cons of microbial platforms, this review compared the key microbial production hosts side-by-side. The most significant recent improvement to microbes comes from a holistic understanding of cellular function based on omics, GSMMs, and systems biology. In *E. coli*, the capabilities for folding DSBs have significantly improved, which widens the spectrum of expressible POIs and exploitation of the T3SS secretion system opens new avenues. *Bacillus* keeps achieving impressive titers for industrial proteins and efforts to translate these titers to pharmaceutical proteins have been accelerated over the past decade. *S. cerevisiae* remains the key target of yeast-based research and modern tools have elevated our understanding of this host. *P. pastoris* continues its rise to the forefront of protein production hosts and probably has, together with cell-free protein synthesis (CFPS), the best shot at challenging mammalian systems in producing complex glycosylated proteins.

The near future will see the continued development of alternative protein forms, such as antibody mimetics, which will facilitate a shift away from the current dominance of mammalian production systems. Should these novel protein formats prove successful in clinical applications, they will cause a paradigm shift back toward microbial production hosts.

Nevertheless, many questions and avenues for future research remain (see Outstanding questions). In essence, microbial protein production bottlenecks in secretion, disulfide bond formation, and human-like glycosylation will need to be further reduced before they can compete with CHO cells, particularly when it comes to molecules such as full-length antibodies. Furthermore, until now, CHO-based protein production had the advantage when it came to cost-effective downstream processing mainly due to cleaner extracellular secretion capabilities. Microbials will need to improve significantly before they can economically compete with these (already well-established) processes. However, we are confident it can be done. A holistic understanding of the cellular processes based on omics and systems biology can give microbes the necessary edge. Pharmaceutical process development is still heavily burdened by insufficient predictability of both host and POI behavior. If our understanding of microbial platforms can reach a point where it is possible to accurately predict product yields as well as best-use organisms, signal peptides, expression levels, and process conditions, microbes will take a huge step toward outcompeting mammalian systems.

In conclusion, we propose that microbes are inherently more suited for recombinant protein production and that future research and innovations will guide their application to the pharmaceutical market. We anticipate that CHO dominance has peaked and that the coming years will see a graduate shift back toward microbial-based pharmaceutical protein production. This will come on the back of an increased demand for antibody mimetics and a holistic understanding of both microbes and their POI production processes.

Acknowledgments

This work was funded by the People Programme (Marie Skłodowska-Curie Actions) of the European Union's Horizon 2020 Programme under REA grant agreement no. 813979 (SECRETERS).

Outstanding questions

Can *E. coli* engineering achieve competitive titers for complex disulfide-bonded proteins?

Can *E. coli*-centered efforts to enhance disulfide bond formation be translated to *B. subtilis*?

Can *B. subtilis* translate its impressive secretory protein titers from enzyme production to the biopharma industry?

Will yeast cell engineering enable consistent high secreted titers for a variety of proteins?

How will the integration of folding mechanics impact the predictive qualities of GSMMs and other models?

How will the continuous adoption of omics and systems biology tools impact our understanding of microbial protein production? How far away is a complete and holistic understanding of all cellular processes?

Will new bacterial or yeast strains make it to the forefront of microbial production systems, or is the existing knowledge and experience regarding the established hosts too great to be overcome?

Will the industry evolve toward a unified production platform that is simple but versatile and completely understood, or will it rather continue to diversify into myriad purpose-dedicated platforms?

What will be the impact of cell-free systems and will they outcompete cell-based expression systems?

How can we best implement novel tools and innovations into the biomanufacturing processes? Can a quicker and more streamlined adaptation processes benefit the microbe-based product development process?

How long will the new approvals be dominated by antibodies for, and will alternative proteins establish themselves in the market any time soon?

How will alternatives to protein-based treatments, such as viral delivery systems for DNA or RNA, small interfering

Declaration of interests

None declared by authors.

Resources

[i]www.marketsandmarkets.com/Market-Reports/protein-expression-market-180323924.html

References

1. Rader, R.A. and Langer, E.S. (2018) Fifteen years of progress: biopharmaceutical industry survey results. *Pharm. Technol. Eur.* 30, 10–12
2. Tripathi, N.K. and Shrivastava, A. (2019) Recent developments in bioprocessing of recombinant proteins: expression hosts and process development. *Front. Bioeng. Biotechnol.* 7, 420
3. Walsh, G. (2018) Biopharmaceutical benchmarks 2018. *Nat. Biotechnol.* 36, 1136–1145
4. Sanchez-Garcia, L. *et al.* (2016) Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microb. Cell Factories* 15, 1–7
5. Sandomenico, A. *et al.* (2020) Evolution of *Escherichia coli* expression system in producing antibody recombinant fragments. *Int. J. Mol. Sci.* 21, 1–39
6. Gupta, S.K. and Shukla, P. (2017) Microbial platform technology for recombinant antibody fragment production: a review. *Crit. Rev. Microbiol.* 43, 31–42
7. Lakowitz, A. *et al.* (2018) Mini review: recombinant production of tailored bio-pharmaceuticals in different *Bacillus* strains and future perspectives. *Eur. J. Pharm. Biopharm.* 126, 27–39
8. Liu, Y. and Huang, H. (2018) Expression of single-domain antibody in different systems. *Appl. Microbiol. Biotechnol.* 102, 539–551
9. Spadiut, O. *et al.* (2014) Microbials for the production of monoclonal antibodies and antibody fragments. *Trends Biotechnol.* 32, 54–60
10. Owens, B. (2017) Faster, deeper, smaller-the rise of antibody-like scaffolds. *Nat. Biotechnol.* 35, 602–603
11. Simeon, R. and Chen, Z. (2018) In vitro-engineered non-antibody protein therapeutics. *Protein Cell* 9, 3–14
12. Binz, H.K. *et al.* (2017) Design and characterization of MP0250, a tri-specific anti-HGF/anti-VEGF DARPin® drug candidate. *MAbs* 9, 1262–1269
13. Baeshen, M.N. *et al.* (2015) Production of biopharmaceuticals in *E. coli*: current scenario and future perspectives. *J. Microbiol. Biotechnol.* 25, 953–962
14. Jozala, A.F. *et al.* (2016) Biopharmaceuticals from microorganisms: from production to purification. *Braz. J. Microbiol.* 47, 51–63
15. Zhu, W. *et al.* (2021) Medium optimization for high yield production of human serum albumin in *Pichia pastoris* and its efficient purification. *Protein Expr. Purif.* 181, 1–11
16. Werten, M.W.T. *et al.* (2019) Production of protein-based polymers in *Pichia pastoris*. *Biotechnol. Adv.* 37, 642–666
17. Vieira Gomes, A. *et al.* (2018) Comparison of yeasts as hosts for recombinant protein production. *Microorganisms* 6, 38
18. Gifre, L. *et al.* (2017) Trends in recombinant protein use in animal production. *Microb. Cell Factories* 16, 40
19. van Dijl, J.M. and Hecker, M. (2013) *Bacillus subtilis*: from soil bacterium to super-secreting cell factory. *Microb. Cell Factories* 12, 1–6
20. Singh, R. *et al.* (2016) Microbial enzymes: industrial progress in 21st century. *3 Biotech* 6, 174
21. Kohlstedt, M. *et al.* (2014) Adaptation of *Bacillus subtilis* carbon core metabolism to simultaneous nutrient limitation and osmotic challenge: a multi-omics perspective. *Environ. Microbiol.* 16, 1898–1917
22. Brady, J.R. *et al.* (2020) Comparative genome-scale analysis of *Pichia pastoris* variants informs selection of an optimal base strain. *Biotechnol. Bioeng.* 117, 543–555
23. Long, C.P. *et al.* (2018) Dissecting the genetic and metabolic mechanisms of adaptation to the knockout of a major metabolic enzyme in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 115, 222–227
24. Huangfu, J. *et al.* (2020) Omics analysis reveals the mechanism of enhanced recombinant protein production under simulated microgravity. *Front. Bioeng. Biotechnol.* 8, 30
25. Qi, Q. *et al.* (2020) Different routes of protein folding contribute to improved protein production in saccharomyces cerevisiae. *MBio* 11, 1–12
26. Becker, J. and Wittmann, C. (2018) From systems biology to metabolically engineered cells — an omics perspective on the development of industrial microbes. *Curr. Opin. Microbiol.* 45, 180–188
27. Subramanian, I. *et al.* (2020) Multi-omics data integration, interpretation, and its application. *Bioinform. Biol. Insights* 14, 24
28. Chen, Y. *et al.* (2020) Systems and synthetic biology tools for advanced bioproduction hosts. *Curr. Opin. Biotechnol.* 64, 101–109
29. Bulović, A. *et al.* (2019) Automated generation of bacterial resource allocation models. *Metab. Eng.* 55, 12–22
30. Nocon, J. *et al.* (2014) Model based engineering of *Pichia pastoris* central metabolism enhances recombinant protein production. *Metab. Eng.* 24, 129–138
31. Chen, K. *et al.* (2017) Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation. *Proc. Natl. Acad. Sci. U. S. A.* 114, 11548–11553
32. Müller, M.M. (2018) Post-translational modifications of protein backbones: unique functions, mechanisms, and challenges. *Biochemistry* 57, 177–185
33. Macek, B. *et al.* (2019) Protein post-translational modifications in bacteria. *Nat. Rev. Microbiol.* 17, 651–664
34. Schjoldager, K.T. *et al.* (2020) Global view of human protein glycosylation pathways and functions. *Nat. Rev. Mol. Cell Biol.* 21, 729–749
35. Du, T. *et al.* (2019) A bacterial expression platform for production of therapeutic proteins containing human-like O-linked glycans. *Cell Chem. Biol.* 26, 203–212
36. Mueller, P. *et al.* (2018) High level in vivo mucin-type glycosylation in *Escherichia coli*. *Microb. Cell Factories* 17, 168
37. De Wachter, C. *et al.* (2018) Engineering of yeast glycoprotein expression. *Adv. Biochem. Eng. Biotechnol.* 175, 93–135
38. Katla, S. *et al.* (2019) Novel glycosylated human interferon alpha 2b expressed in glycoengineered *Pichia pastoris* and its biological activity: N-linked glycoengineering approach. *Enzym. Microb. Technol.* 128, 49–58
39. Gong, B. *et al.* (2013) Glycosylation characterization of recombinant human erythropoietin produced in glycoengineered *Pichia pastoris* by mass spectrometry. *J. Mass Spectrom.* 48, 1308–1317
40. Kang, J.Y. *et al.* (2018) Lysosomal targeting enhancement by conjugation of glycopeptides containing mannose-6-phosphate glycans derived from glyco-engineered yeast. *Sci. Rep.* 8, 8730
41. Hershewe, J. *et al.* (2020) Cell-free systems for accelerating glycoprotein expression and biomanufacturing. *J. Ind. Microbiol. Biotechnol.* 47, 977–991
42. Guerrero Montero, I. *et al.* (2019) *Escherichia coli* 'TatExpress' strains export several g/l human growth hormone to the periplasm by the Tat pathway. *Biotechnol. Bioeng.* 116, 3282–3291
43. Tredwell, G.D. *et al.* (2017) Rapid screening of cellular stress responses in recombinant *Pichia pastoris* strains using metabolite profiling. *J. Ind. Microbiol. Biotechnol.* 44, 413–417
44. Ellis, M. *et al.* (2017) Development of a high yielding *E. coli* periplasmic expression system for the production of humanized Fab' fragments. *Biotechnol. Prog.* 33, 212–220
45. Quesada-Ganuza, A. *et al.* (2019) Identification and optimization of PrsA in *Bacillus subtilis* for improved yield of amylase. *Microb. Cell Factories* 18, 1–16
46. Navone, L. *et al.* (2021) Disulfide bond engineering of AppA phytase for increased thermostability requires co-expression

(si)RNA, gene editing, and stem cells affect the industry?

How can more cost-, energy- and resource-effective 'greener' expression platforms be developed for sustainable production of biopharmaceutics and industrial enzymes?

of protein disulfide isomerase in *Pichia pastoris*. *Biotechnol. Biofuels* 14, 1–14
47. Levy, R. *et al.* (2001) Production of correctly folded fab antibody fragment in the cytoplasm of *Escherichia coli* trxB gor mutants via the coexpression of molecular chaperones. *Protein Expr. Purif.* 23, 338–347
48. Lobstein, J. *et al.* (2016) Erratum to: SHuffle, a novel *Escherichia coli* protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Factories* 15, 124
49. Ren, G. *et al.* (2016) Use of the SHuffle strains in production of proteins. *Curr. Protoc. Protein Sci.* 2016, 5.26.1–5.26.21
50. Kouwen, T.R.H.M. and Van Dijl, J.M. (2009) Applications of thiol-disulfide oxidoreductases for optimized in vivo production of functionally active proteins in *Bacillus*. *Appl. Microbiol. Biotechnol.* 85, 45–52
51. Hatahet, F. *et al.* (2010) Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of *E. coli*. *Microb. Cell Fact.* 9, 67
52. Sohail, A.A. *et al.* (2020) Production of extracellular matrix proteins in the cytoplasm of *E. coli*: making giants in tiny factories. *Int. J. Mol. Sci.* 21, 688
53. Gaciarz, A. *et al.* (2017) Efficient soluble expression of disulfide bonded proteins in the cytoplasm of *Escherichia coli* in fed-batch fermentations on chemically defined minimal media. *Microb. Cell Factories* 16, 1–12
54. Delic, M. *et al.* (2014) Engineering of protein folding and secretion—strategies to overcome bottlenecks for efficient production of recombinant proteins. *Antioxid. Redox Signal.* 21, 414–437
55. Karyolaimos, A. *et al.* (2019) Enhancing recombinant protein yields in the *E. coli* periplasm by combining signal peptide and production rate screening. *Front. Microbiol.* 10, 1511
56. Dastjerdeh, M.S. *et al.* (2019) In silico analysis of different signal peptides for the secretory production of recombinant human keratinocyte growth factor in *Escherichia coli*. *Comput. Biol. Chem.* 80, 225–233
57. Selas Castiñeiras, T. *et al.* (2018) Development of a generic β-lactamase screening system for improved signal peptides for periplasmic targeting of recombinant proteins in *Escherichia coli*. *Sci. Rep.* 8, 1–18
58. Kasli, I.M. *et al.* (2019) Use of a design of experiments approach to optimise production of a recombinant antibody fragment in the periplasm of *Escherichia coli*: selection of signal peptide and optimal growth conditions. *AMB Express* 9, 5
59. Horga, L.G. *et al.* (2018) Tuning recombinant protein expression to match secretion capacity. *Microb. Cell Factories* 17, 199
60. Zahrl, R.J. *et al.* (2018) The impact of ERAD on recombinant protein secretion in pichia pastoris (Syn komagataella spp.). *Microbiology (United Kingdom)* 164, 453–463
61. Barrero, J.J. *et al.* (2018) An improved secretion signal enhances the secretion of model proteins from *Pichia pastoris*. *Microb. Cell Factories* 17, 1–13
62. Zahrl, R.J. *et al.* (2019) Detection and elimination of cellular bottlenecks in protein-producing yeasts. *Methods Mol. Biol.* 1923, 75–95
63. 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. *Biochim. Biophys. Acta, Mol. Cell Res.* 1863, 3116–3124
64. Neef, J. *et al.* (2020) Relative contributions of non-essential Sec pathway components and cell envelope-associated proteases to high-level enzyme secretion by *Bacillus subtilis*. *Microb. Cell Factories* 19, 1–13
65. van der Ploeg, R. *et al.* (2012) High-salinity growth conditions promote tat-independent secretion of tat substrates in *Bacillus subtilis*. *Appl. Environ. Microbiol.* 78, 7733–7744
66. Green, C.A. *et al.* (2019) Engineering the flagellar type III secretion system: improving capacity for secretion of recombinant protein. *Microb. Cell Factories* 18, 10
67. Stargardt, P. *et al.* (2020) Bacteriophage inspired growth-decoupled recombinant protein production in *Escherichia coli*. *ACS Synth. Biol.* 9, 1336–1348
68. Kastenhofer, J. *et al.* (2021) Inhibition of *E. coli* host RNA polymerase allows efficient extracellular recombinant protein production by enhancing outer membrane leakiness. *Biotechnol. J.* 16
69. Thön, M.P. and Koebsch, I.P. (2021) A unique *E. coli*-based secretion system. *Genet Eng. Biotech. News* 41, 68–70
70. Zhang, Z. *et al.* (2017) Non-peptide guided auto-secretion of recombinant proteins by super-folder green fluorescent protein in *Escherichia coli*. *Sci. Rep.* 7, 6990
71. Liu, M. *et al.* (2019) Soluble expression of single-chain variable fragment (scFv) in *Escherichia coli* using superfolder green fluorescent protein as fusion partner. *Appl. Microbiol. Biotechnol.* 103, 6071–6079
72. Merrick, W.C. and Pavitt, G.D. (2018) Protein synthesis initiation in eukaryotic cells. *Cold Spring Harb. Perspect. Biol.* 10, a033092
73. Cheng, C. *et al.* (2017) A novel Ffu fusion system for secretory expression of heterologous proteins in *Escherichia coli*. *Microb. Cell Factories* 16, 231
74. Prielhofer, R. *et al.* (2018) Superior protein titers in half the fermentation time: promoter and process engineering for the glucose-regulated GTH1 promoter of *Pichia pastoris*. *Biotechnol. Bioeng.* 115, 2479–2488
75. Jungo, C. *et al.* (2007) Mixed feeds of glycerol and methanol can improve the performance of *Pichia pastoris* cultures: a quantitative study based on concentration gradients in transient continuous cultures. *J. Biotechnol.* 128, 824–837
76. Che, Z. *et al.* (2020) An effective combination of codon optimization, gene dosage, and process optimization for high-level production of fibrinolytic enzyme in *Komagataella phaffii* (*Pichia pastoris*). *BMC Biotechnol.* 20, 1–13
77. Bankefa, O.E. *et al.* (2018) Enhancing the secretion pathway maximizes the effects of mixed feeding strategy for glucose oxidase production in the methylotrophic yeast *Pichia pastoris*. *Bioresour. Bioprocess.* 5, 4–11
78. García-Fruitós, E. *et al.* (2012) Bacterial inclusion bodies: making gold from waste. *Trends Biotechnol.* 30, 65–70
79. Humer, D. and Spadiut, O. (2018) Wanted: more monitoring and control during inclusion body processing. *World J. Microbiol. Biotechnol.* 34, 1–9
80. Lai, T. *et al.* (2013) Advances in mammalian cell line development technologies for recombinant protein production. *Pharmaceuticals* 6, 579–603
81. Berting, A. *et al.* (2010) Virus susceptibility of Chinese hamster ovary (CHO) cells and detection of viral contaminations by adventitious agent testing. *Biotechnol. Bioeng.* 106, 598–607
82. Ritacco, F.V. *et al.* (2018) Cell culture media for recombinant protein expression in Chinese hamster ovary (CHO) cells: history, key components, and optimization strategies. *Biotechnol. Prog.* 34, 1407–1426
83. Amann, T. *et al.* (2019) Glyco-engineered CHO cell lines producing alpha-1-antitrypsin and C1 esterase inhibitor with fully humanized N-glycosylation profiles. *Metab. Eng.* 52, 143–152
84. Dahodwala, H. and Lee, K.H. (2019) The fickle CHO: a review of the causes, implications, and potential alleviation of the CHO cell line instability problem. *Curr. Opin. Biotechnol.* 60, 128–137
85. Ko, P. *et al.* (2018) Probing the importance of clonality: single cell subcloning of clonally derived CHO cell lines yields widely diverse clones differing in growth, productivity, and product quality. *Biotechnol. Prog.* 34, 624–634
86. Higel, F. *et al.* (2016) N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *Eur. J. Pharm. Biopharm.* 100, 94–100
87. Jiang, H. *et al.* (2019) Challenging the workhorse: comparative analysis of eukaryotic micro-organisms for expressing monoclonal antibodies. *Biotechnol. Bioeng.* 116, 1449–1462
88. Buyel, J.F. *et al.* (2017) Very-large-scale production of antibodies in plants: the biologization of manufacturing. *Biotechnol. Adv.* 35, 458–465
89. Wurm, F.M. (2013) CHO quasispecies-implications for manufacturing processes. *Processes* 1, 296–311

90. Dumont, J. *et al.* (2016) Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit. Rev. Biotechnol.* 36, 1110–1122
91. Martínez-Solís, M. *et al.* (2019) Engineering of the baculovirus expression system for optimized protein production. *Appl. Microbiol. Biotechnol.* 103, 113–123
92. Hidalgo, D. *et al.* (2018) Biotechnological production of pharmaceuticals and biopharmaceuticals in plant cell and organ cultures. *Curr. Med. Chem.* 25, 3577–3596
93. Su, H.H. *et al.* (2020) Production of recombinant human epidermal growth factor in *Bacillus subtilis*. *J. Taiwan Inst. Chem. Eng.* 106, 86–91
94. Yadwad, V.B. *et al.* (1994) Production of human epidermal growth factor by an ampicillin resistant recombinant *Escherichia coli* strain. *Biotechnol. Lett.* 16, 885–890
95. Valdés, J. *et al.* (2009) Physiological study in *Saccharomyces cerevisiae* for overproduction of a homogeneous human epidermal growth factor molecule. *Biotechnol. Appl.* 26, 166–167
96. Eissazadeh, S. *et al.* (2017) Production of recombinant human epidermal growth factor in *Pichia pastoris*. *Braz. J. Microbiol.* 48, 286–293
97. Şahin, B. *et al.* (2015) Feeding strategy design for recombinant human growth hormone production by *Bacillus subtilis*. *Bioprocess Biosyst. Eng.* 38, 1855–1865
98. Hahm, M.S. and Chung, B.H. (2001) Secretory expression of human growth hormone in *Saccharomyces cerevisiae* using three different leader sequences. *Biotechnol. Bioprocess Eng.* 6, 306–309
99. Çalik, P. *et al.* (2013) Effect of co-substrate sorbitol different feeding strategies on human growth hormone production by recombinant *Pichia pastoris*. *J. Chem. Technol. Biotechnol.* 88, 1631–1640
100. Kunaparaju, R. *et al.* (2005) Epi-CHO, an episomal expression system for recombinant protein production in CHO cells. *Biotechnol. Bioeng.* 91, 670–677
101. Breitling, R. *et al.* (1989) Secretory expression in *Escherichia coli* and *Bacillus subtilis* of human interferon α genes directed by staphylokinase signals. *MGG Mol. Gen. Genet.* 217, 384–391
102. Babu, K.R. *et al.* (2000) Production of interferon-α in high cell density cultures of recombinant *Escherichia coli* and its single step purification from refolded inclusion body proteins. *Appl. Microbiol. Biotechnol.* 53, 655–660
103. Chu, J. *et al.* (2003) Fermentation process optimization of recombinant *Saccharomyces cerevisiae* for the production of human interferon-α2a. *Appl. Biochem. Biotechnol. A Enzym. Eng. Biotechnol.* 111, 129–137
104. Shiga, Y. *et al.* (2000) Efficient production of N-terminally truncated biologically active human interleukin-6 by *Bacillus brevis*. *Biosci. Biotechnol. Biochem.* 64, 665–669
105. Guisez, Y. *et al.* (1991) Production and purification of recombinant human interleukin-6 secreted by the yeast *Saccharomyces cerevisiae*. *Eur. J. Biochem.* 198, 217–222
106. Li, H. *et al.* (2011) Large-scale production, purification and bioactivity assay of recombinant human interleukin-6 in the methylotrophic yeast *Pichia pastoris*. *FEMS Yeast Res.* 11, 160–167
107. Zhang, Y. *et al.* (1997) Efficient and inducible production of human interleukin 6 in Chinese hamster ovary cells using a novel expression system. *Cytotechnology* 25, 53–60
108. Lee, S.W. *et al.* (2019) Functional secretion of granulocyte colony stimulating factor in *Bacillus subtilis* and its thermogenic activity in brown adipocytes. *Biotechnol. Bioprocess Eng.* 24, 298–307
109. Dasari, V. *et al.*, Natco Pharma Limited, A novel process for production of recombinant human G-CSF. WO2008096368A2.
110. Bae, C.S. *et al.* (1999) Improved process for production of recombinant yeast-derived monomeric human G-CSF. *Appl. Microbiol. Biotechnol.* 52, 338–344
111. Bahrami, A. *et al.* (2007) Production of recombinant human granulocyte-colony stimulating factor by *Pichia pastoris*. *Iran. J. Biotechnol.* 5, 162–169
112. Rotondaro, L. *et al.* (1997) High-level expression of a cDNA for human granulocyte colony-stimulating factor in Chinese hamster ovary cells: effect of 3′-noncoding sequences. *Appl. Biochem. Biotechnol. B Mol. Biotechnol.* 7, 231–240
113. Olmos-Soto, J. and Contreras-Flores, R. (2003) Genetic system constructed to overproduce and secrete proinsulin in *Bacillus subtilis*. *Appl. Microbiol. Biotechnol.* 62, 369–373
114. Shin, C.S. *et al.* (1997) Enhanced production of human miniproinsulin in fed-batch cultures at high cell density of *Escherichia coli* BL21(DE3)[pET-3aT2M2]. *Biotechnol. Prog.* 13, 249–257
115. Kjeldsen, T. (2000) Yeast secretory expression of insulin precursors. *Appl. Microbiol. Biotechnol.* 54, 277–286
116. Gurramkonda, C. *et al.* (2010) Application of simple fed-batch technique to high-level secretory production of insulin precursor using *Pichia pastoris* with subsequent purification and conversion to human insulin. *Microb. Cell Factories* 9, 31
117. Pak, S.C.O. *et al.* (2006) Expression of recombinant human insulin in Chinese hamster ovary cells is complicated by intracellular insulin-degrading enzymes. In *New Developments and New Applications in Animal Cell Technology* (Merten, O.W. *et al.*, eds), pp. 59–67, Springer
118. Visser, H. *et al.* (2011) Development of a mature fungal technology and production platform for industrial enzymes based on a *Myceliophthora thermophila* isolate, previously known as *Chrysosporium lucknowense* C1. *Ind. Biotechnol.* 7, 214–223
119. Emalfarb, M.A. Dyadic International USA Inc. Expression and high-throughput screening of complex expressed DNA libraries in filamentous fungi. US8680252B2.
120. Lai, J.Y. *et al.* (2019) Potential application of *Leishmania tarentolae* as an alternative platform for antibody expression. *Crit. Rev. Biotechnol.* 39, 380–394
121. Mally, M. *et al.* Limmatech Biologics Ag. Glycoengineered monoclonal antibody. WO2019234021A1.
122. Hermann, T. (2003) Industrial production of amino acids by coryneform bacteria. *J. Biotechnol.* 104, 155–172
123. Liu, X. *et al.* (2016) Expression of recombinant protein using *Corynebacterium glutamicum*: progress, challenges and applications. *Crit. Rev. Biotechnol.* 36, 652–664
124. Vanier, G. *et al.* (2018) Alga-made anti-Hepatitis B antibody binds to human Fcγ receptors. *Biotechnol. J.* 13, 1–10
125. Ramos-Martinez, E.M. *et al.* (2017) High-yield secretion of recombinant proteins from the microalga *Chlamydomonas reinhardtii*. *Plant Biotechnol. J.* 15, 1214–1224
126. Yin, G. *et al.* (2012) Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. *MAbs* 4, 217–225
127. Kelwick, R. *et al.* (2016) Development of a *Bacillus subtilis* cell-free transcription-translation system for prototyping regulatory elements. *Metab. Eng.* 38, 370–381
128. Su, Y. *et al.* (2020) *Bacillus subtilis*: a universal cell factory for industry, agriculture, biomaterials and medicine. *Microb. Cell Factories* 19, 1–12
129. Heide, C. *et al.* (2021) Design, development and optimization of a functional mammalian cell-free protein synthesis platform. *Front. Bioeng. Biotechnol.* 8, 1–10
130. Salehi, A.S.M. *et al.* (2016) Cell-free protein synthesis of a cytotoxic cancer therapeutic: onconase production and a just-add-water cell-free system. *Biotechnol. J.* 11, 274–281
131. Henrich, E. *et al.* (2015) Membrane protein production in *Escherichia coli* cell-free lysates. *FEBS Lett.* 589, 1713–1722
132. Yin, G. *et al.* (2017) RF1 attenuation enables efficient non-natural amino acid incorporation for production of homogeneous antibody drug conjugates. *Sci. Rep.* 7, 1–13
133. Gao, W. *et al.* (2019) Advances and challenges in cell-free incorporation of unnatural amino acids into proteins. *Front. Pharmacol.* 10, 611
134. Zemella, A. *et al.* (2015) Cell-free protein synthesis: pros and cons of prokaryotic and eukaryotic systems. *ChemBioChem* 16, 2420–2431
135. Haslbeck, M. and Buchner, J. (2018) In vitro refolding of proteins. In *Oxidative Folding of Proteins: Basic Principles, Cellular Regulation and Engineering* (Feige, M.J., ed.), pp. 129–151, Royal Society of Chemistry

Annex 2: The structures and sequences of Brazzein, Dulaglutide and Romiplostim

Polypeptide sequence for Brazzein:

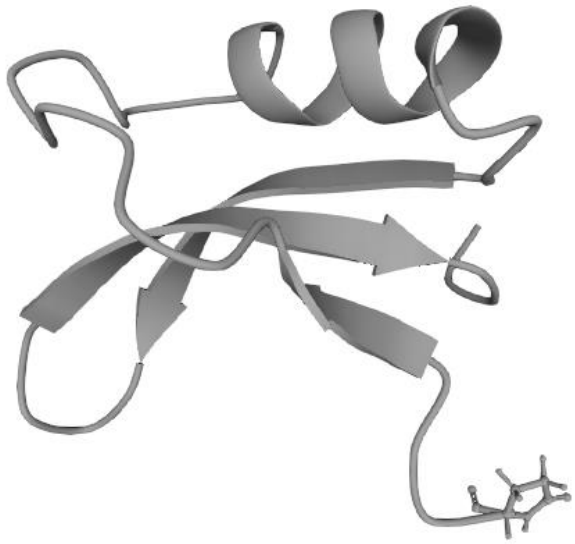
TorA SP-

QDKCKKVYENYPVSKCQLANQCNYDCKLDKHARSGECFYDEKRNLQCICDYCEY-HHHHHH

Pink: TorA SP sequence

Highlighted in yellow: Cysteines

Black: His6 tag



Molecular weight:

TorA SP-Brazzein-His (precursor): 11.5 kDa

Brazzein-His (mature): 7.3 kDa

His6 tag: 0.8 kDa

TorA SP: 4.2 kDa

Polypeptide sequence for Romiplostim (monomer):

TorA SP-
DKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHN
AKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLP
PSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQ
QGNVFSCSVMHEALHNHYTQKSLSLSPGKGGGGGIEGPTLRQWLAARAGGGGGGGGIEGPTLRQW
LAARA-HHHHHH

Pink: TorA SP sequence

Blue: Hinge

Orange: CH2

Green: CH2-CH3 interface residues (Fc-ball and socket joint)

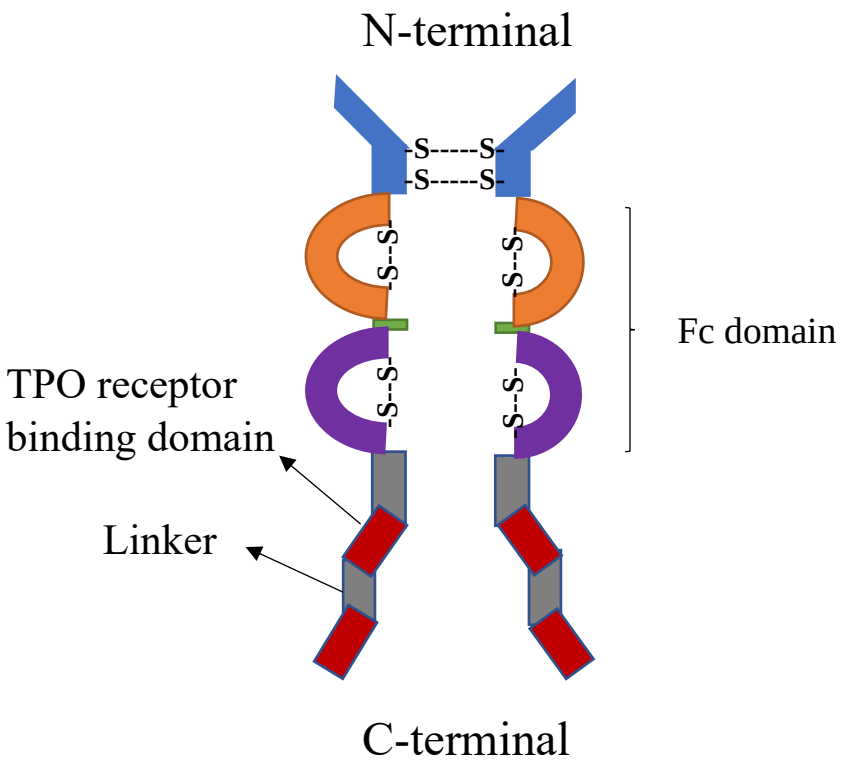
Purple: CH3

Grey: Glycine linker

Red: TPO-receptor domain (Peptide of Romiplostim)

Black: His6 tag

Highlighted in yellow: Cysteines



Molecular weight:

TorA SP-Dulaglutide-His (precursor): 34.4 kDa

Dulaglutide-His (mature): 30.2 kDa

His6 tag: 0.8 kDa

TorA SP: 4.2 kDa

Polypeptide sequence for Dulaglutide (monomer):

TorA SP-
HGEGTFTSDVSSYLEEQAAKEFIAWLVKGGGGGGGSGGGGSGGGGSAESKYGPPCPPCPAPEAAG
GPSVFLFPPKPKDTLMISRTPEVTCVVVDVSQEDPEVQFNWYVDGVEVHNAKTKPREEQFNSTYRV
VSVLTVLHQDWLNGKEYKCKVSNKGLPSSIEKTISKAKGQPREPQVYTLPPSQEEMTKNQVSLTCL
VKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSRLTVDKSRWQEGNVFSCSVMHEALHN
HYTQKSLSLSLG-HHHHHH

Pink: TorA SP sequence

Red: GLP-1 analogue (Peptide of Dulaglutide)

Grey: Glycine-Serine linker

Blue: Hinge

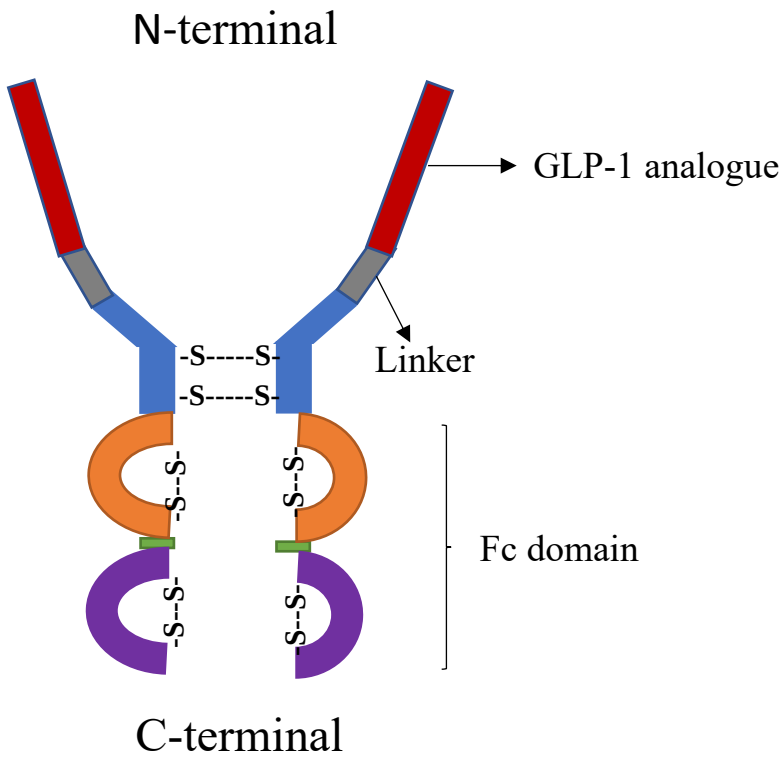
Orange: CH2

Green: CH2-CH3 interface residues (Fc-ball and socket joint)

Purple: CH3

Black: His6 tag

Highlighted in yellow: Cysteines



Molecular weight:

TorA SP-Romiplostim-His (precursor): 34.8 kDa

Romiplostim-His (mature): 30.6 kDa

His6 tag: 0.8 kDa

TorA SP: 4.2 kDa

Annex 3: List of other proteins of interest expressed at shake flask scale to the periplasm by Tat pathway

Protein	Signal peptide	Protein description	Export to the periplasm by Tat pathway
BIWA4-scFv	TorA	scFv fragment shared by Boehringer-Ingelheim (Vienna)	None
BIWA4-Fab	TorA	Fab fragment shared by Boehringer-Ingelheim (Vienna)	None
G. luciferase	PhoD*n (mutation: L51N)	*Gaussia princeps luciferase* (SECRETERS consortium model protein)	Inconclusive
YebF	AmiC (mutation: QV27RR)	Native *E. coli* protein shared by Lloyd Ruddock	Yes. Excessive cell lysis
YebF	MdoD (mutation: QA28RR)	Native *E. coli* protein shared by Lloyd Ruddock	Yes. Excessive cell lysis
YebF	PhoD	Native *E. coli* protein obtained by Lloyd Ruddock	Inconclusive
YebF	YcbK	Native *E. coli* protein obtained by Lloyd Ruddock	None
scFv-IL-1b	AmiC	scFv fragment shared by Lloyd Ruddock	None
scFv-IL-1b	MdoD	scFv fragment shared by Lloyd Ruddock	None
scFv-IL-1b	PhoD	scFv fragment shared by Lloyd Ruddock	None
scFv-IL-1b	TorA	scFv fragment shared by Lloyd Ruddock	Low
Herceptin scFv	AmiC	scFv fragment shared by Lloyd Ruddock	None
Herceptin scFv	MdoD	scFv fragment shared by Lloyd Ruddock	None
Herceptin scFv	PhoD	scFv fragment shared by Lloyd Ruddock	None
Herceptin scFv	TorA	scFv fragment shared by Lloyd Ruddock	Low
Herceptin scFv	YcbK	scFv fragment shared by Lloyd Ruddock	None
IgA_1 scFv	AmiC	scFv fragment shared by Lloyd Ruddock	None
IgA_1 scFv	PhoD	scFv fragment shared by Lloyd Ruddock	None
IgA_1 scFv	TorA	scFv fragment shared by Lloyd Ruddock	Low
IgA_1 scFv	YcbK	scFv fragment shared by Lloyd Ruddock	None

The POIs were cloned with a 6xHis-tag in a pET23 plasmid and express together with the plasmid containing the CyDisCo components (pMJS205) or empty plasmid (pAG82) in BL21 cells. Firstly, 6 h preculture at 37°C 250 rpm in LB media supplemented with 2 g/L of glucose and corresponding antibiotics was carried out. Proteins highlighted in yellow were grown as a main culture in 100 mL flasks with 10 mL culture in each flask in AIM media. The flasks were covered with oxygen permeable membranes, grown at 30°C, 250 rpm for 24 hours and harvested for fractionation. The rest of the POI in the table, were grown as a main culture in 24 deep well plates covered with air permeable membrane at 30°C in AIM media, 250 rpm, and harvested after 24 hours. The cells were collected by centrifugation and were lysed by freeze-thawing. Finally, the soluble fraction was purified and analysed by SDS-PAGE (more detailed information in Materials and Methods in section 4.2.). "Mutation": SP were mutated with the aim of making them exclusively Tat-dependent. "Inconclusive": experimental constrains or lack of follow-up experiments do not permit a final answer.

Annex 4: tac promoter mutations

-35 **-10**

ptac ATTCTGAAATGAGCTG**TTGACA**ATTAATCATCGGCTCG**TATAAT**GTGTGGAA
TTGTGAGCGGATAACAATTTCACACAG

ptacM1 ATTCTGAAATGAGCTG**TTGACA**ATTAATCATCGGCTCG**TAT**GTGGTGTGGA
ATTGTGAGCGGATAACAATTTCACACAG

ptacM2 ATTCTGAAATGAGCTG**TT**T**ACA**ATTAATCATCGGCTCG**TATAAT**GTGTGGA
ATTGTGAGCGGATAACAATTTCACACAG

tac promoter wild type and mutations DNA sequences. tac promoter -35 consensus element is underlined in bold and highlighted in yellow; -10 consensus element is underlined in bold and highlighted in green. Mutated base-pairs are in red. ptac M1 has three base pair substitution in -10 promoter element, while ptac M2 has only one base pair substitution in -35 consensus element.

Annex 5: Publication 1. Yields and product comparison between *Escherichia coli* BL21 and W3110 in industrially relevant conditions: anti-c-Met scFv as a case study

Microbial Cell Factories

RESEARCH

Open Access

Yields and product comparison between *Escherichia coli* BL21 and W3110 in industrially relevant conditions: anti-c-Met scFv as a case study

Klaudia Arauzo-Aguilera[1†], Luisa Buscajoni[2†], Karin Koch[2], Gary Thompson[3], Colin Robinson[1] and Matthias Berkemeyer[2*]

Abstract

Introduction In the biopharmaceutical industry, *Escherichia coli* is one of the preferred expression hosts for large-scale production of therapeutic proteins. Although increasing the product yield is important, product quality is a major factor in this industry because greatest productivity does not always correspond with the highest quality of the produced protein. While some post-translational modifications, such as disulphide bonds, are required to achieve the biologically active conformation, others may have a negative impact on the product's activity, effectiveness, and/or safety. Therefore, they are classified as product associated impurities, and they represent a crucial quality parameter for regulatory authorities.

Results In this study, fermentation conditions of two widely employed industrial *E. coli* strains, BL21 and W3110 are compared for recombinant protein production of a single-chain variable fragment (scFv) in an industrial setting. We found that the BL21 strain produces more soluble scFv than the W3110 strain, even though W3110 produces more recombinant protein in total. A quality assessment on the scFv recovered from the supernatant was then performed. Unexpectedly, even when our scFv is correctly disulphide bonded and cleaved from its signal peptide in both strains, the protein shows charge heterogeneity with up to seven distinguishable variants on cation exchange chromatography. Biophysical characterization confirmed the presence of altered conformations of the two main charged variants.

Conclusions The findings indicated that BL21 is more productive for this specific scFv than W3110. When assessing product quality, a distinctive profile of the protein was found which was independent of the *E. coli* strain. This suggests that alterations are present in the recovered product although the exact nature of them could not be determined. This similarity between the two strains' generated products also serves as a sign of their interchangeability. This study encourages the development of innovative, fast, and inexpensive techniques for the detection of heterogeneity while also provoking a debate about whether intact mass spectrometry-based analysis of the protein of interest is sufficient to detect heterogeneity in a product.

[†]Luisa Buscajoni and Klaudia Arauzo-Aguilera have contributed equally to this work.

*Correspondence:
Matthias Berkemeyer
matthias.berkemeyer@boehringer-ingelheim.com
Full list of author information is available at the end of the article

Keywords Sec pathway, Fermentation, *Escherichia coli* BL21 and W3110, Disulphide bond, Product heterogeneity, Protein purification

Introduction

Escherichia coli is one of the expression hosts of choice in the biopharmaceutical industry for large-scale production of therapeutic proteins because of its rapid growth, high product yield, cost effective production and easy scale-up processes [41]. If the protein of interest (POI) contains disulphide bonds (DSBs), as in the case of antibody fragments, periplasmic expression via the Sec pathway is often preferred [45]. Thereby the POI is transported in an unfolded state to the periplasm by fusing a signal peptide (SP) to the N-terminus of the POI. Once in the periplasm, correct DSB formation is achieved [23]. Furthermore, product translocation into the medium can be enforced which also simplifies downstream processing [58].

A good understanding of fermentation parameters and their impact on *E. coli* cell growth and final product yield is critical in defining biopharmaceutical production processes. Until today, many process adaptations to maximise product yield are based on optimizations of temperature, dissolved oxygen (DO) levels, pH, media composition, feeding strategies, etc. [26, 53]. However, maximum productivity does not always coincide with the highest quality of the recombinantly expressed protein [14]. Correct folding and in vivo stability of the recombinant protein are two crucial factors that must be controlled while optimising the cultivation conditions. Correct folding includes both the acquisition of the correct 3D structure as well as addition of post-translational modifications (PTMs), such as DSB formation. In *E. coli,* PTMs that occur during and after protein synthesis can represent a limitation when compared to other microorganisms. In the past decades several approaches have been explored in *E. coli* to overcome this drawback, reviewed by Rettenbacher et al. [41].

Some PTMs and physiochemical transformations of recombinant proteins can also originate from non-enzymatic reactions at all steps of the production process from cell culture to purification and storage [4]. In this case, PTMs are caused by chemical reactions occurring between the amino acid side chain and reagents present in either culture media or buffers in specific conditions of pH, temperature and oxygenation level. Some of these modifications can negatively affect the activity, efficacy and safety of the desired product by altering the product stability and its biological active conformation. Therefore, the percentage of product harbouring these modifications, within the heterogenous product pool generated, is identified as product related impurities and represents a crucial quality parameter for regulatory authorities [44]. Common PTMs are methionine oxidation, asparagine and glutamine deamidation, and aspartate isomerization. The importance of such unwanted modifications has been evaluated and ranked for recombinant monoclonal antibodies (mAbs) [29]. While N-terminal pyroglutamate, for example, is not considered as a critical quality attribute, modifications occurring in the complementary determining regions are of high importance as they could affect the antigen recognition capacity [29].

To investigate product heterogeneity, ion exchange chromatography [28, 34, 35] coupled with enzymatic digestion followed by mass spectrometry (MS) analysis (peptide mapping) [22, 49] are often the methods of choice. The former allows the separation of the protein heterogeneity based on the charge properties while the latter allows the identification of mass changes and the exact position of the modification within the protein expressed. However, since peptide mapping requires extensive work and can also generate artefactual modifications, the research for the improvement of this method is ongoing [10, 40].

In this study, fermentation conditions of two widely employed industrial *E. coli* strains, namely BL21 (B strain) and W3110 (K-12 strain) are compared for recombinant protein production of a single-chain variable fragment (scFv) in an industrial setting. Rather than analysing the well-known and studied performance and behavioural differences [31, 36, 47, 48], we focused on yields, and analysed the differences in product structure and heterogeneity between strains grown in 5 L fed-batch bioreactors using a number of different downstream and analytical techniques. The results reveal surprising difference in protein quantity and quality between the two strains, and equally surprising heterogeneity in the final preparations of this relatively simple biopharmaceutical.

Results and discussion

BL21 strain produces more soluble scFvM than W3110 strain

One of the major aims of this study was to directly compare the production of a biopharmaceutical product under industrial conditions in the two extensively used *E. coli* strains: BL21 and W3110. The viable and cost-effective production of a POI using *E. coli* varies enormously depending on many different factors. POI related factors

and upstream process parameters such as pH, temperature, media composition, strain type and others influence the recombinant expression [25, 52].

Jones et al. [20] and Edwardraja et al. [8] have produced this scFv in *E. coli* in the periplasm after export by Tat pathway and expression in the cytoplasm, respectively, both at shake flask level. Traditional upstream bioprocess development involves the use of shaken bioreactor systems (usually shake flask). Cultivation in shake flasks is normally performed in a batch manner, provides very limited variable monitoring, and produces low cell densities and product yields. Furthermore, they rely on uncontrolled surface aeration leading to limited oxygen transfer rates and low batch-to-batch reproducibility [1]. Therefore, cultivation conditions that are used during shaken culture bioprocess development may be changed or completely discarded once they are optimized at pilot scale [39]. To overcome the limitations described above, there has been a concerted effort to develop fully automated high-throughput cultivation systems to significantly accelerate the identification of the optimal expression systems and process conditions [2, 16].

In our research, a screening of different conditions for the optimization of soluble yields of the scFvM was carried out for both strains listed above combining different temperatures, pH and inducer concentrations (described in "Materials and Methods"). This screening showed only minimal changes in OD_{550} and titer. However, a further optimization of media composition and induction time could not be carried out since the implementation of these changes would require a complete revaluation of this automated protocol for 10 mL fermentation [16].

In Fig. 1A, the specific soluble product formation was calculated for BL21 and W3110 in 10 mL and 5 L bioreactors following the standard protocol at 7 h post-induction (T7) (refer to "Materials and Methods"). To eliminate a possible impact of the different optical density (OD) levels at both scales, the specific soluble product titer was determined by dividing soluble product titer by OD_{550} for both strains and scales at time point T7 (end of fermentation in 10 mL scale) (Fig. 1A). The BL21 strain shows a specific soluble product titer of 23.5 mg/OD in 10 mL fermenters and 12.3 mg/OD in benchmark 5 L fermenters. Janzen et al. [16] also described a higher specific soluble product titer formation in small-scale cultivations than in 5 L fermenters when employing a B strain. It has been suggested by Kang et al. [21] that BL21 may suffer from DO limitations in large-scale cultures, however this hypothesis could not be verified by our research due to the absence of comparison data between the two fermentation scales. The W3110 strain, on the other hand, shows very similar specific soluble product titers in both scales: 5.7 mg/OD in 10 mL fermenters and 4.9 mg/OD in 5 L fermenters. These results validate the robustness and reproducibility that this strain provides in industry [21, 56]. However, with respect to the expression strains used, BL21 showed significantly higher titers in all direct comparisons (Fig. 1A). Specific soluble product titer

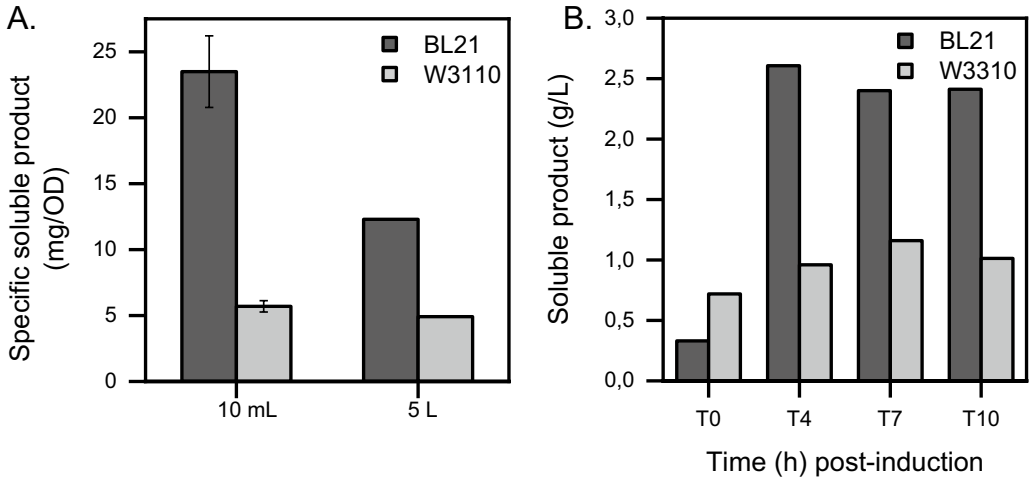


Fig. 1 Soluble scFvM production in BL21 and W3110 in 10 mL and 5 L scale fermentations. **A** Specific soluble product titer in BL21 and W3110 in 10 mL and 5 L bioreactors expressed in mg/OD. To calculate these values, the 7 h post-induction soluble scFvM titer is divided by the OD_{550} of the culture at that time in both scale bioreactors. 10 mL fermentations specific soluble product formation values are the average of four replicates, error bars have been added to the figure. **B** Soluble titer variation of the scFvM in BL21 and W3110 strains at different time points [0 h (T0), 4 h (T4), 7 h (T7) and 10 h (T10) post-induction] in 5 L fermentations expressed in g/L. In both graphs, BL21 strain is represented in black and W3110 in grey. Soluble scFvM product was determined by immunoassay from suspension samples. 5 L bioreactors were run two times for each strain with slight variations in temperature, inductor concentration and pH resulting in comparable profiles for OD_{550} and titer

comparison between scales was also carried out with the optimized conditions after the screening experiments in 10 mL fermenters (refer to "Materials and Methods"). Since a very similar pattern of soluble protein was obtained when comparing scales and strains, the data set is not shown because it had a comparable trend.

Looking more closely at the benchmark process, Fig. 1B shows the soluble production of scFvM in BL21 and W3110 in 5 L fermenters and standard conditions at different time points: 0 h (T0), 4 h (T4), 7 h (T7) and 10 h (T10) post-induction. Soluble scFvM was quantified by an immunoassay from suspension samples. Overall, BL21 shows a higher soluble protein content (two-fold) during the entire induction period in the 5 L fermentation system compared to W3110 (Fig. 1B). In BL21, the peak production of soluble protein is achieved 4 h after induction (2.61 g/L) and it remains stable until T10 (end of fermentation, 2.41 g/L). BL21 shows a tight regulation of the tac promoter under non-induced conditions (0 h post-induction): it leaks 0.33 g/L. On the other hand, W3110's peak production of soluble protein is achieved 7 h after induction (1.16 g/L) and it also remains stable until T10 (end of fermentation, 1.01 g/L). Unlike BL21, W3110's tac promoter is leakier and produces more than half of the total soluble scFvM before induction (0.72 g/L). This leaky expression in W3110 could be linked to plasmid instability, which many times explains a poor yield of target protein [43]. However, in this case, differences in yield between the chosen strains are not connected to plasmid loss or instability, as plasmid copy number (PCN) remains stable and comparable between them throughout the whole fermentation process (observed: ≈ 12–18 copies/cell,expected: 15–20 copies/cell). When plasmid instability is discarded, BL21 and W3110 critical genome differences for recombinant protein production should be considered to understand these yield differences. Even though BL21 and W3110 are both widely used in recombinant protein production, B strains are deficient in the Lon protease, which degrades many recombinant proteins. The B strain also lacks the outer membrane protease OmpT, whose function is to degrade extracellular proteins [43]. These genetic differences between strains may explain the higher yields obtained with BL21. In addition, it should be noted that BL21 reaches a lower OD_{550} value than the W3110 strain (BL21: 212 and W3110: 272) at the end of fermentation. These OD differences between the compared strains might correlate with the metabolic burden caused by the continuous export to the periplasm by Sec pathway [13] and/or the lethal outer membrane punctures occurred as a result of limited periplasmic capacity [46, 51] in BL21. A second run of bioreactors with parameters optimised for BL21 (refer to "Materials and Methods") was carried out and similar patterns for yields and OD_{550} values were obtained compared to standard conditions (data set is not shown because it had a comparable trend).

Even when the focus is on soluble production, an additional inherent part of disulphide bonded protein production in *E. coli* cannot be dismissed: inclusion body (IB) formation. The Coomassie blue-stained gel in Fig. 2 shows the lysates of BL21 and W3110 from 5 L fermenters when expressing OmpA-scFvM at T0, T4, T7 and T10 time-points in standard conditions. Cell suspension was analysed and fractionated in total titer (TT, comprising soluble and insoluble proteins), total soluble (TS, comprising intracellular and extracellular soluble proteins) and supernatant samples (SN, only proteins located in the extracellular medium). The insoluble POI production is remarkably different between the compared strains (Fig. 2). When looking at the Coomassie blue-stained gel, it is important to notice that total production of the POI (TT) is visually higher in W3110 than in BL21 at all time points. This result suggests that the majority of the protein is produced as IBs and only a small part is translocated to the periplasm and extracellular medium and is therefore soluble (TS). As explained before in Fig. 1B, and as it can be noticed in the Coomassie blue-stained gel in Fig. 2 (see and visually compare T0 scFvM production in both strains), W3110 pre-induction leakiness is higher than BL21's. We hypothesise that due to this early high-level expression in W3110, hydrophobic stretches in the polypeptide are present at high concentrations very early in the cell and are available for interaction with similar regions. This may lead to protein instability and aggregation (IB formation) [6, 43]. Over time, and in both strains, but more remarkably in BL21, the POI starts to be detectable in the supernatant due to the leakiness of the outer membrane [46, 51], active export [55] and/or lysis of the cells [24].

The purified scFvM shows multiple charged variants on a CEX

The second objective of this research was to determine whether the protein expressed by the two strains, after translocation in the periplasm, was similarly folded and contained the same charge heterogeneity. The produced scFvM from the T10 (end of fermentation) was purified from periplasmic extract and from culture supernatant by nickel immobilised metal affinity chromatography (IMAC). Table 1 shows the amount of soluble protein obtained after purification of the same volume of either supernatant or extracted periplasm. It should be noted that the intracellular soluble titer (periplasmic fraction) was higher than the extracellular one in both strains. However, this was not reflected in the product titer after purification. This difference depends on the low amount

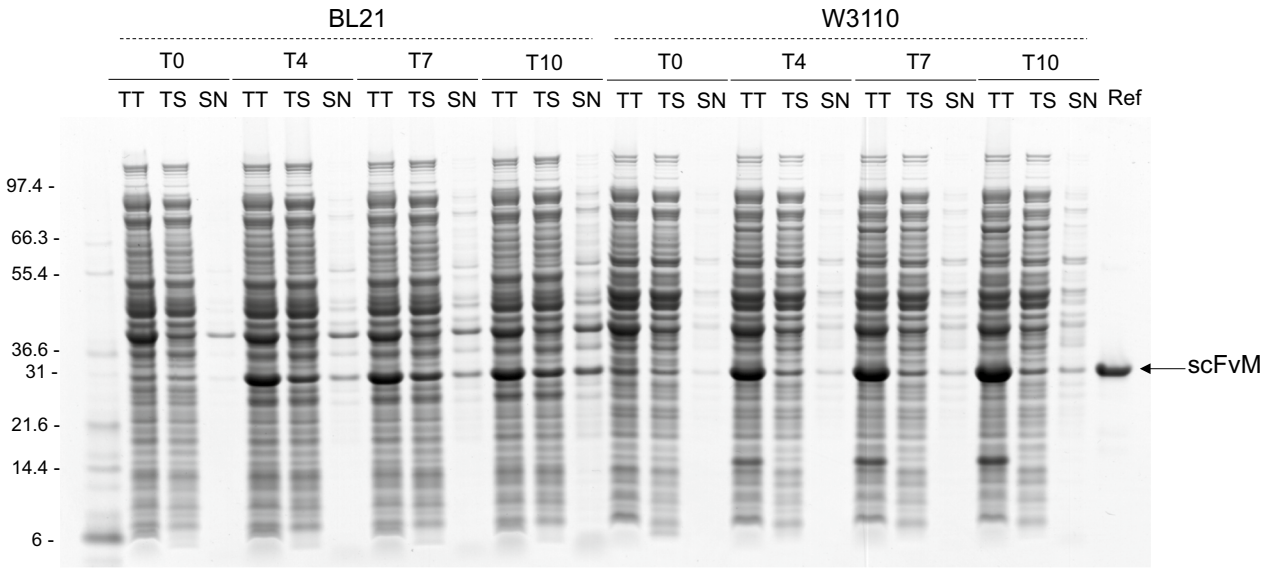


Fig. 2 Lysates of BL21 and W3110 from 5 L fermenters when expressing OmpA-scFvM. Suspension and supernatant samples of BL21 and W3110 expressing OmpA-scFvM were recovered at different time points: T0, T4, T7 and T10 for SDS-PAGE analysis in reducing condition. Representative Coomassie blue-stained gel of the total titer (TT, comprising soluble and insoluble proteins), total soluble (TS, comprising intracellular and extracellular soluble proteins) and supernatant samples (SN, only proteins located in the supernatant) and scFvM reference protein.Ladder (Mark12™ Unstained Standard) on the left in kDa. The same volume of sample was treated and loaded for comparison. OD of the samples: BL21: T0 (167); T4 (229); T7 (231); T10 (212) and W3110: T0 (165); T4 (217); T7 (236); T10(272)

Table 1 Comparative analysis of the scFvM produced by the two strains in the different compartments

Strain	Location	Protein concentration (g/L)	Protein purified (mg)	Purification yield (%)	Purification factor[a]	Purity (%)
BL21	Periplasm	2.1	0.5	51.8	7.1	92.2
	Supernatant	0.5	45.9	81.4	3.7	91.5
W3110	Periplasm	0.9	0.3	77.5	9.4	91.3
	Supernatant	0.2	26.3	99.6	7.2	92.3

[a] The purification factor was calculated dividing the purification yield (%) over the relative abundance (%) of the target protein in the examined compartment

of periplasm that could be extracted due to setting constraints in the maximal cell pellet that can be processed ("see Materials and Methods"). Analysis via Coomassie blue-stained gel (BL21 Fig. 3A, W3110 SI Fig. 1A) shows that the scFvM was obtained with high purity, independently of the expression host and purified compartment.

To further investigate the presence of possible heterogeneity of the expressed scFvM, the IMAC purified material from culture supernatants was separated by cation exchange chromatography (CEX) via gradient elution, since the scFvM has a basic isoelectric point of 7.8 (Expasy Protparam). CEX is typically considered a gold standard technique to separate and purify charge variants [54, 57]. However, the results from this technique can be strongly influenced by differences in operational parameters such as column type, particle size and flow rate [9, 18]. Since previous studies demonstrated the importance of the diameter resin particles and flow rate on the separation performance [18, 57], a small resin particle (10 µm diameter resin) coupled with a slow flow rate (0.5 mL/min) was selected in our case. The results indicated a high separation performance. The chromatograms show the presence of two main peaks: one more acidic (N°2) and another one more basic (N°5), both coupled with some minor subforms (Fig. 3B). Between the two strains, the elution pattern is maintained, however, the relativity of the peaks slightly changes BL21 being richer in acidic variants while W3110 in basic ones (Fig. 3C). To verify, firstly, if the multiple peaks showed different masses and impurities, a non-reducing gel was assessed. However, no differences could be detected (BL21 Fig. 3D, W3110 SI Fig. 1B).

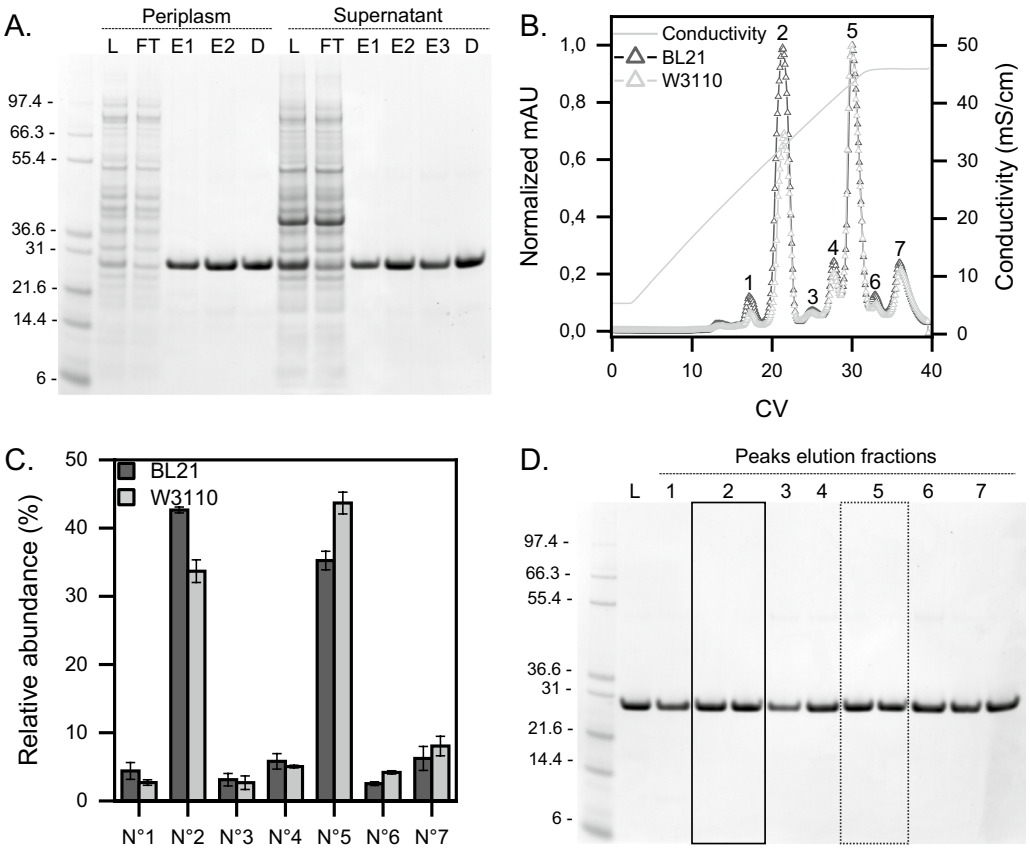


Fig. 3 scFvM purification via two steps chromatography. **A** Representative Coomassie blue-stained gel of the scFvM produced with BL21 and purified via IMAC. The samples analysed are the total protein sample loaded (L), column flow-though (FT), eluates (E) from either the periplasm or the culture supernatant and the purified protein after dialysis (D) in non-reducing condition. Ladder (Mark12™ Unstained Standard) on the left in kDa. **B** Representative MonoS normalised chromatograms of scFvM purified from the culture supernatant. Peaks are labelled with numbers ranging from 1 to 7. **C** Relative abundance of the CEX peaks in BL21 and W3110. BL21 strain is represented in black and W3110 in grey. **D** Representative Coomassie blue-stained gel of a CEX run in non-reducing condition on the scFvM produced with the BL21 strain. peak N°2 elution fractions are in the continued box while peak N°5 ones are in the dotted box

scFvM is correctly disulphide bonded and cleaved from its SP in BL21 and W3110

In this study an offline approach was applied, consisting in the isolation of the separated forms from CEX followed by individual analysis for better understanding of possible modifications. The workflow involved coupling size-exclusion chromatography (SEC) directly to mass spectrometry (MS). This was done to verify that the scFvM was correctly folded and that different peaks on CEX were not caused by a pool of species with free thiols or uncleaved SP. IMAC and CEX purified samples (peaks N°2 and N°5) from both periplasm and supernatant samples from BL21 and W3110 were analysed by LC–MS. This analysis confirmed that all scFvM samples had the expected molecular weight, consistent with the cleavage of the SP when the POI is exported from the cytoplasm to the periplasm and its four cysteines in two DSBs (Table 2). The main component in all these samples is the unmodified scFvM molecule. In addition to it and as second species, both BL21 and W3110 samples show comparable reduction of −17 Da, probably due to N-terminal pyroglutamate modification (pyroQ), while gluconoylation is only seen in BL21 samples (+178 Da). PyroQ modification is generated after a non-enzymatic cyclization of N-terminal glutamine whose rate of formation can be affected by various environmental factors during purification and storage [3, 4]. In previous studies, CEX has been reported as a method of choice for the separation of pyroQ modifications since the loss of a primary amine causes an acidity shift of the antibody [4, 5]. However, in this study the use of a strong cation exchange did not show the same results. In fact, LC–MS analysis run on each of the CEX peaks showed the presence of a −17 Da modifications, ranging from 12–25%, in each sample (Table 2). The strain selectivity of the non-enzymatic gluconoylation

Table 2 Analysis of free thiol content, SP cleavage and secondary modifications based on MS

Strain	Technique	Location	N° of Cysteine	$M_{ox\,Theor}$ [a]	M_{Exp} [b]	Δ mass	+178 Da (5–11%)	−17 Da (12–25%)
BL21	IMAC	Periplasm	4	27478.27	27478.27	0	+ [c]	+
		Supernatant		27478.27	27478.27	0	+	+
	CEX	Peak N°2		27478.27	27478.27	0	+	+
		Peak N°5		27478.27	27478.27	0	+	+
W3110	IMAC	Periplasm		27478.27	27478.27	0	−	+
		Supernatant		27478.27	27478.27	0	−	+
	CEX	Peak N°2		27478.27	27478.27	0	−	+
		Peak N°5		27478.27	27478.27	0	−	+

[a] Calculated theoretical oxidised molecular weight ($M_{OxTheor}$)

[b] Experimental molecular weight (M_{Exp})

[c] The sign "+" indicates the presence of the modification (+178 Da or −17 Da) while "−" indicates the absence of it

modification agrees with previous literature. B strains accumulate 6-phosphogluconolactone due to the lack of 6-phosphogluconolactonase [33], which favours gluconoylation, so it is not unexpected that this strain produces some gluconoylated proteins. Although this modification can adversely affect protein quality, the gluconoylation is not very stable and can transform back into unmodified protein and gluconate via a hydrolysis reaction [32].

In addition, possible mismatches of the DSBs were also analysed among the multiple peaks in CEX (purified from BL21) by MS. In this case the purified protein from each CEX peak was digested. However, no differences in the size of the peptides generated were identified both in native conditions and after reduction, confirming that DSB shuffling is not essentially the reason for the heterogeneous pattern in CEX (Table 2).

The stability of the two main peaks excludes a handling artefact

A reversibility analysis on the two main peaks was then performed to verify if these two main CEX forms were not an increasing modification caused by downstream operation. The eluted single peaks were therefore pooled from different purifications in N°2 and N°5 respectively, dialyzed against the equilibration buffer and each pool was loaded again on the CEX. Figure 4A shows the comparison of the elution profiles from each pool. In both cases the purification revealed a perfect reproducibility of the peaks that seem to be only in a very slow reversible

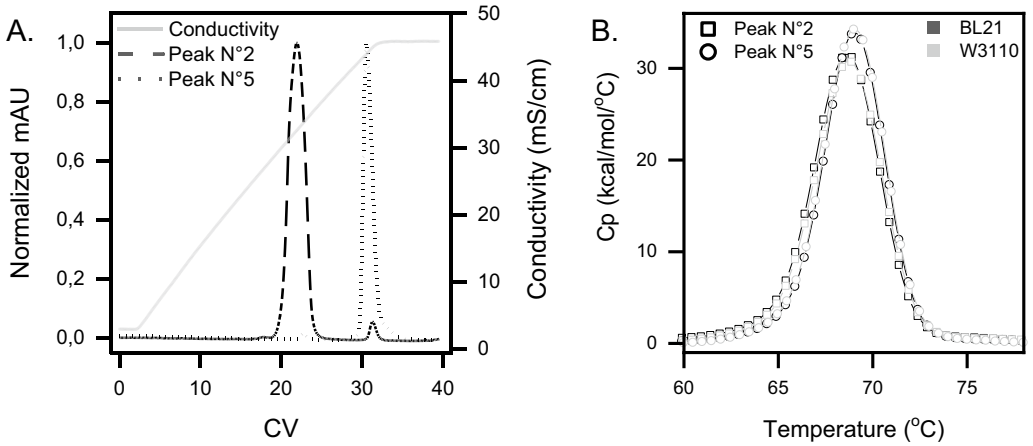


Fig. 4 Stability analysis of the two main peaks from the CEX. **A** Overlaid MonoS chromatograms of the two main peaks that were collected, dialyzed, and reloaded separately to demonstrate the stability and purity of the two forms. **B** DSC profiles of the peaks N°2 and N°5 obtained from CEX purification of scFvM expressed by BL21 and W3110. The obtained Tms are peak N°2 68.7 ℃ and peak N°5 69.1 ℃. Squares represent peak N°2 and circles peak N°5. BL21 strain is represented in black and W3110 in grey

equilibrium with the counterpart. Moreover, since with this experiment the heterogeneous pattern observed during the CEX purification of the purified material (Fig. 3B) was not visible, an artefact of the column caused by overloading of the samples could be excluded. The two peaks therefore represent scFvM isoforms that are stable under these conditions.

Biophysical characterization confirmed altered conformations of the two main charged variants

To further investigate potential conformational differences between the two main CEX forms, and the exact comparability among the two strains, a biophysical assessment was established via differential scanning calorimetry (DSC) and one-dimensional proton nuclear magnetic resonance (1D-^{1}H-NMR) spectroscopy. DSC is commonly employed to assess the thermal and conformational stability of a protein under specific buffer conditions [19]. The two separated main peaks (N°2 and N°5) were evaluated by DSC (Fig. 4B). One single transition, corresponding to the unfolding of the scFvM, was observed in all samples. The melting temperatures of the single peaks N°2 and N°5 (from both BL21 and W3110) were 68.7 ℃ and 69.1 ℃, respectively and corresponds to a temperature difference of 0.4 ℃. These results proved a high comparability, in terms of thermal stability, of the POI produced by the two *E. coli* strains.

In addition to this thermal stability difference, a modification in the state of the protein from the two main peaks from the CEX was detected via one-dimensional proton ^{1}H-NMR spectroscopy (1D-^{1}H-NMR). This technique shows signals for each hydrogen atom in the protein that is covalently bound or exchanging slowly with water (for example amide signals will be present but those from –OH and NH_3 groups will be missing). These signals resonate at different frequencies (chemical shifts in ppm; parts per million of the main field) and with different intensities based on the 3D structure, ligand binding state and dynamics of the protein all of which affect local magnetic fields in the protein. The position of peaks in the 1D-^{1}H-NMR depend at first order on the chemistry of the atoms [27], so for example CH_2 and CH_3 groups from different amino acid types (e.g., Val vs Leu) appear at different positions and also have small differences due to the primary sequence. On top of these chemical effects from residue types and the primary sequence the spectrum is also extremely sensitive to the 3D structure of the protein and very small changes in local environment and dynamics (see for example [7] can be detected, so 1D-^{1}H-NMR can be used as a fingerprint of the proteins 3D structure and to monitor small changes in the state of the protein. In this case the spectra suggest both samples contain proteins that are well-folded as indicated by the well resolved and dispersed signals in the amide region (not shown) and a series of well-resolved methyl peaks at < 0 ppm which are indicative of stable methyl aromatic packing in the protein's core (boxes in Fig. 5). On top of this, the spectra are similar enough to conclude that there have been no major changes in 3D structure and that the overall 3D fold is the same. However, while the methyl peaks at < 0 ppm are well dispersed, they also show variations between the two CEX peak samples (N°2 and N°5). For CEX peak N°5 the methyl signals that resonate at −0.736 ppm and −0.981 ppm show shifts of + 0.04 ppm and ~ −0.02 ppm and ~ 10–20% weaker peak height than CEX peak N°2 for both strains (Fig. 5B and D). As the samples were very thoroughly dialysed these differences would indicate a change in conformation within the core of the protein or the presence of a yet unidentified strongly bound ligand.

All these results combined suggested that the two strains produced a target protein with the same overall 3D fold and thermal stability. However, the difference between the two main peaks of the CEX seems to be generated by a conformational change of the protein that leaves the mass unchanged as observed by MS analysis. Two main reasons could be behind this result: the presence of a ligand tightly bound to the protein or a modification that leaves the mass unchanged. With regards to the ligand, the reversibility analysis on the two main peaks suggests that, if present, the ligand binds with high affinity to one of the two forms since it could not be removed during the dialysis process. Moreover, at least, the presence of a high molecular mass molecule as ligand could be excluded since, prior LC–MS analysis, the samples were desalted by size exclusion chromatography (SEC) and no additional peaks were observed. Concerning the second hypothesis, various types of modification can occur during protein expression, manufacturing, and storage [3, 4]. Some of these modifications can lead to mass shifts while some others can result in protein modifications leaving the mass unchanged. Examples of the latter are DSB mismatch and aspartate isomerization which can further generate aspartic acid racemization [54]. The presence of mismatched DSBs was excluded as a possible reason for multiple peaks in CEX, as described above. Aspartate (Asp) isomerization is a non-enzymatic modification that can cause conformational changes of the protein since it introduces an additional methyl group in the protein backbone [3, 4]. Furthermore, the specific structural outcome can lead to two isomeric products (L-isoAsp and D-isoAsp) where the D-amino acid can affect the peptide function [42]. This reaction occurs at an optimal pH of 5, produces succinimide (−18 Da specie) as a reaction intermediate and it is favoured on aspartate residues that are followed by a

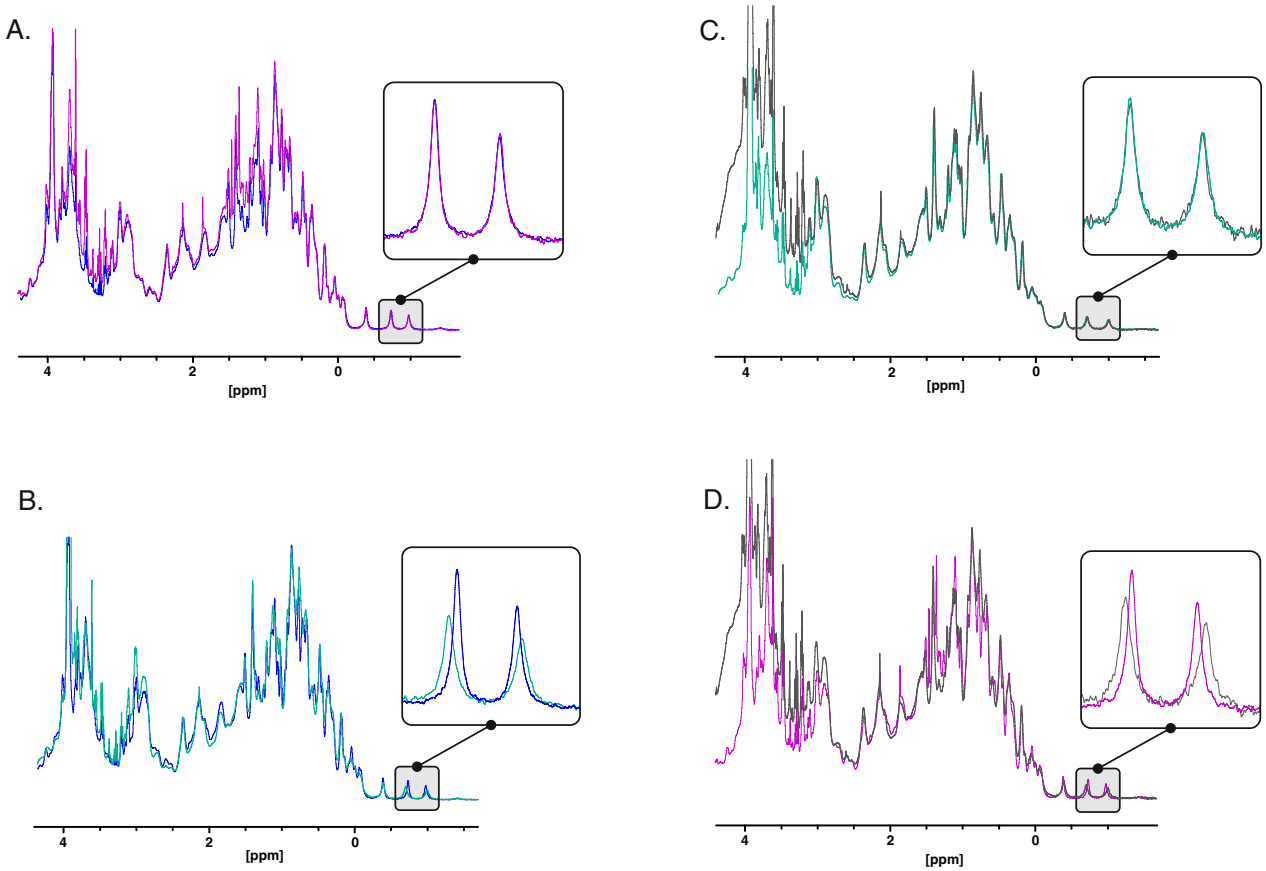


Fig. 5. ^{1}H-NMR spectra of peaks N°2 and N°5 conformers measured at 600 MHz. Spectra from basic, acidic species and control samples show the conformational and temporal stability of the species. The samples were exhaustively co-dialysed before analysis. Samples conditions: pH 5.5 and 25 ℃, sample concentration ~ 50 µM. **A** Spectra of peak N°2 from BL21 (purple) and W3110 (blue). **C** Peak N°5 from BL21 (grey) and W3110 (green) representing analysis of the same species showing reproducible conformational state as a control and spectra showing the clear difference in the fingerprint of peak N°2 and N°5 conformers from W3110 (**B)** and BL21 (**D**). The box represents the expansion of the high field methyl aromatic fingerprint region

glycine [11, 37, 38]. Moreover, it was shown in a previous study that antibody variants containing isoaspartate elute later in CEX [12]. The current analysis was carried out at pH 5.5, a second and more basic peak was present, in the protein sequence an aspartate close to a glycine is present and a species with −17/18 Da was detected. Therefore, the presence of aspartate isomerization as a possible modification cannot be excluded completely. There are some other methods to identify this modification such as LC–MS peptide mapping and 2D NMR analysis. LC–MS peptide mapping is becoming an important method for the characterization of primary sequences and PTMs in antibody products. On the other hand, this technique is labour-intensive, time-consuming and can introduce artificial PTMs, resulting in an overestimation of the target protein modifications. During sample preparation, long digestions can generate unnecessary reactions interfering with the quantitation of the peaks of interest, causing low reproducibility of the results. Shortening digestion-time can cause incomplete peptide cleavages, thus low sequence coverage and poor repeatability [17]. In this case study, where BL21 and W3110 protein product comparison was the aim, this technique had to be discarded due to reproducibility issues. On the other hand, 2D NMR analysis typically utilises labelled isotopes during the fermentation process for proteins of this mass, and therefore, this approach was not available due to experimental constraints.

Conclusion

In this case-study, we report that *E. coli* strain differences may have an influence on the final product yield but not necessarily on its heterogeneity pattern. The results showed that BL21 and W3110 have a very different productivity profile in the conditions employed, with BL21 being more industrially relevant to produce this specific scFv in terms of yield. In terms of quality, except for the B strain characteristic gluconoylation, the expressed scFvM

displays a similar heterogeneity profile. This resemblance of the produced product represents an indication of the interchangeability between the two strains, a characteristic that presents an important perspective for biosimilar and biobetter production. In our paper it is further shown that when scFvM product quality was assessed by different analytical methods, a distinctive profile of the product was obtained, suggesting that alterations in the recovered product are present independently of the host strain. These alterations appear to be stable and significant, since the two forms elute at very different salt concentrations during CEX. However, the exact identity of the cause of product heterogeneity could not be appropriately confirmed. This opens a discussion on whether MS based intact mass analysis of the POI is enough to spot the heterogeneity in a product. At the same time, we want to encourage the discoveries of new, fast and affordable methods for analysis of heterogeneity other than 2D NMR and LC–MS based peptide mapping for the identification of protein heterogeneity in biopharma, which are time-consuming.

Materials and methods

All chemicals, reagents and enzymes were of highest quality and were obtained from Sigma-Aldrich, Roth or Thermo Fisher Scientific, unless otherwise noted.

anti-c-Met scFv (scFvM) expression strain generation

Escherichia coli DH5α (Invitrogen) was used for genetic manipulations. The anti-c-Met scFv (sequence taken from Edwardraja et al. [8]), with an N-terminal wild-type OmpA SP and a C-terminal 6×His-tag was commercially synthesised (GeneArt). The synthesis construct was sub-cloned into the pFLAG-CTC vector (Sigma Aldrich) under the control of a tac promoter using NdeI and EcoRI restriction sites. This construct will be termed as OmpA-scFvM in this work. Individual clones were sequenced before transforming the expression plasmid into the expression strains *E. coli* BL21 (Novagen) and W3110 (DSMZ). In this work, the protein anti-c-Met scFv is referred to as scFvM to further correlate with Edwardraja et al. and Jones et al. [8, 20] which investigated the same protein in a different setup.

Expression in a miniaturised fermentation platform (Multifermenter, MF) and a 5 L fermentation system

The fully automated cultivation at 10 mL scale in the MF was performed as described in Janzen et al. [16] with the exception that the temperature in all reactors was set to 37 ℃ and lowered to the corresponding experiment temperature prior to induction. For the screening of conditions, a range of different temperatures (25–33.5 ℃), pH values (6.3–7.3) and isopropyl β-D-1-thioglactopyranoside (IPTG) inducer concentrations (0.5–1 mM) were tested with a design of experiment (DoE) setting in 32 MF vessels. While for BL21 a slight increase in yields and OD was observed in a screened condition set-up (refer as optimised conditions), with W3110 no optimisation could be achieved. Therefore, solely the standard conditions developed by Janzen et al. [16] and optimised conditions developed for BL21 after MFs run were tested in quadruplicates in 10 mL bioreactors for BL21 and W3110 to maintain a better comparability within the study. In the standard set-up, temperature was set to 30 ℃ prior to induction and the pH was constantly maintained at 6.8. Cultures in this case were induced with 1 mM IPTG (0.024 mL from 75 mM stock). In the optimised BL21 set-up, temperature was set to 32 ℃ prior to induction and the pH was constantly maintained at 7.3. Cultures in this set-up were induced with 0.5 mM IPTG (0.012 mL from 75 mM stock). In all cases, the DO level was maintained at ≥ 35%. In the case of the W3110 strain, a pre-culture in shake-flask was performed, since a significantly prolonged batch phase (e.g. lag phase of the cells) interfered with the fermentation protocol in the MF. The pre-culture was performed at 37 ℃ and 250 rpm until the culture reached an OD_{550} value of 2. The OD_{550} measurement of the pre-culture was manually performed (Genesys 10S UV-Vis; Thermo Fisher Scientific). In the case of BL21, bioreactors were directly inoculated from the cell bank.

Benchmark fed-batch cultivations were performed in fully controlled 5 L standard stirred-tank bioreactor systems (BIOSTAT Cplus; Sartorius Stedim) and the manufacturer provided PCS (MFCS-Win; Sartorius Stedim). Calibration and cultivation conditions and the used material and equipment are described in Janzen et al. [16]. As in MF experiments, the two set-ups (standard conditions, data shown in Results and Discussion section and optimised for BL21, data set not shown as it had a comparable trend as standard conditions) of the experiments were carried out with slight differences in the temperature, pH and induction concentration parameters (more details above). In all cases, the DO level was maintained at ≥ 35% and the pH was kept constant at the set pH ± 0.2 using 25% ammonia and 3 M phosphoric acid. Cultures were induced either with 0.5 mM (optimised conditions for BL21) or 1 mM IPTG (standard conditions) (11.8 mL or 23.6 mL from 211.9 mM stock, respectively). Samples for product quantification, PCN estimation and OD_{550} determination were manually withdrawn before induction (T0) and 4 h (T4), 7 h (T7) and 10 h (T10, end of fermentation) after the IPTG pulse. OD_{550} measurements were directly performed (Genesys 10S UV-Vis; Thermo Fisher Scientific). Samples for product quantification and

PCN estimation were stored in reaction tubes at—20 ℃ until further use.

The minimal media used for all cultivations (both 10 mL and 5 L systems) were prepared with potassium phosphate as buffering agent and source of P and K. Moreover, it contained trisodium citrate, $MgSO_4$, $CaCl_2$, glucose and the trace element solution. The trace element solution and the concentrations of additives used were the same as the ones described in Striedner et al. [50]. The medium was further supplemented with 1 mL/L antifoam agent (PPG 2000,Dow Chemical Co.) and was autoclaved for sterilization in place (SIP) prior cultivation in case of the 5 L system. In the MFs, each block was equipped with bioreactors which were aseptically filled with sterile medium supplemented with 1 mL/L antifoam agent (PPG 2000; Dow Chemical Co.) in a laminar flow hood.

Fermentation sample preparation and quantification

Quantitative analysis was performed by automated immunoassay (Gyros). The sample preparation was conducted with a liquid handling system (Tecan) in 96-well format. For fermentation suspension samples, high viscosity due to leaked nucleic acids caused by fermentation condition and sample freeze and thawing was encountered. High viscosity causes imprecise pipetting by the liquid handling robot. Therefore, as an initial step a nucleic acid hydrolysis with Benzonase® (Merck) (0.5 U/µL for ≥ 10 min, 450 rpm at room temperature (RT)) was performed. Cell lysis was performed by incubation with 1/10 v/v Lysonase (Merck) in FastBreak cell lysis reagent (Promega) for 30 min at RT with shaking at 450 rpm. The soluble fraction was analysed by the immunoassay. For this purpose, the digested cells were centrifuged at 2900 g for 10 min (Tecan centrifuge Hettich Rotanta) and the supernatant was further used. Samples were diluted in the analysis buffer (RexxipA (Gyros)) for quantification.

In the case of supernatant samples from fermentation, straight dilution in the analysis buffer (RexxipA (Gyros)) was performed.

Content quantification was performed using a Gyrolab xPlore by an automated immunoassay with an scFv-specific antibody (109-066-097 (Jackson ImmunoResearch), biotinylated for immobilization within the Gyros CD-microstructure) and an his-tag specific antibody (34670 (Qiagen), Alexa647 fluorescence labelled for detection). The Gyros protocol (200-3W-002-A) was performed according to the manufacturer's instructions. The standard curve was analysed with the Gyros Evaluator SW using a five-parameter fit. Quantification was performed in the linear range of the standard curve (15 ng/mL to 1000 ng/mL).

Plasmid copy number (PCN) estimation

For PCN estimation, $5/OD_{550}$ fermentation pellet samples were used. Fully automated plasmid extraction was performed using the QIAprep Spin Miniprep Kit (Qiagen) and the QIAcube Connect (Qiagen). DNA concentrations were measured using a NanoDropTM (Thermo Fisher Scientific). The following equation was used to estimate the PCN:

$$PCN = \frac{Plasmid\ concentration\left[\frac{ng}{\mu L}\right] * 30\,\mu L * 1.32(correction\ factor\ for\ plasmid\ loss)}{nucleotides\ of\ plasmid * 308.95\frac{Da}{nucleotide} * \left(1.66 * 10^{-15}\frac{ng}{Da}\right) * \left(7.4 * 10^{10}\frac{cells}{pellet}\right)}$$

The weight per single plasmid in ng was calculated from the number of nucleotides and the conversion factors 308.95 (mean weight per nucleotide in Da) and $1.66{\cdot}10^{-15}$ (conversion from Da to ng). DNA loss during plasmid preparation (determined by spiking experiments) and the average number of cells in $5/OD_{550}$ pellet samples ($7.4{\cdot}10^{10}$, determined in a cell counting chamber (Marienfeld Superior)) were considered to calculate the number of plasmids per single cell.

Isolation of scFvM from periplasm and supernatant and content quantification

The expressed scFvM was purified from supernatant and periplasm. The periplasmic extraction was achieved following the pureFrac protocol as described elsewhere [30]. Since a larger volume was necessary an upscaling factor of 130 was applied during all steps. The periplasmic fractions and the supernatant from BL21 and W3110 were then applied on a Ni SepharoseTM 6 Fast Flow (Cytiva) column on the ÄKTA Avant chromatography system (GE Healthcare) at RT and purified via IMAC. At a flow rate of 1 mL/min low binding impurities were washed from the column with 10 column volumes of equilibration buffer (20 mM Phosphate, 500 mM NaCl, 20 M Imidazole, pH 7.4). The protein was then eluted with 5 column volumes of 100% elution buffer (20 mM Phosphate, 500 mM NaCl, 500 mM Imidazole, pH 7.4). The protein content was monitored online by absorbance at 280 nm. After each purification the column was stripped, washed, and recharged to avoid contaminations between the different strains

and purified compartments. The eluted protein was dialyzed against 20 mM Phosphate, 20 mM NaCl, at pH 5.5 using 3.5 molecular weight cut-off (MWCO) dialysis cassettes (Thermo Scientific) and stored at −20 ℃ between purification steps.

With the eluted and rebuffered fractions (20 mM Phosphate, 20 mM NaCl, pH 5.5) from the supernatant a CEX with a MonoSTM 5/50 GL (Cytiva) column was performed on an ÄKTA Purifier chromatography system (GE Healthcare) at RT. The used buffers were 20 mM Phosphate, 20 mM NaCl, at pH 5.5 (equilibration buffer) and 20 mM Phosphate, 500 mM NaCl, at pH 5.5 (elution buffer). At a flow rate of 0.5 mL/min, a gradient from 0–100% elution was applied for 30 column volumes followed by 10 column volumes at 100% elution buffer. The protein content was monitored online by absorbance at 280 nm. The eluted protein was dialyzed against 20 mM Phosphate, 20 mM NaCl, at pH 5.5 using 3.5 MWCO dialysis cassettes (Thermo Scientific) and concentrated with 3 kDa MWCO Pierce™ Protein Concentrators PES (ThermoFisher). The fractions were stored at −20 ℃ until analysed.

Purified protein concentration was determined by measuring the absorbance at 280 nm with the Nanodrop 2000 (ThermoFisher) and a calculated molar extinction coefficient of 58′580 M^{-1} (Expasy Protparam). Affinity purified scFvM and its charged variants (CEX samples) were denatured for 5 min at 80 ℃ and run on an SDS-PAGE gel under reducing and non-reducing conditions.

DSC analysis

The DSC experiments were performed using a MicroCal VP-DSC system (Malvern). All samples were dialyzed against the same buffer (20 mM Phosphate, 20 mM NaCl, pH 5.5) prior analysis using 3.5 MWCO dialysis cassettes (Thermo Scientific). The reference cell was filled with a buffer corresponding to the sample buffer. The samples were heated from 10 ℃ to 95 ℃ at a heating rate of 60 ℃/h. The pre-scan was 3 min, the filtering period was 10 s, and the feedback mode/gain was set to passive. The midpoint of thermal transition temperature (Tm) was obtained by analysing the data using OriginTM 7 software. All experiments were performed at a protein concentration of 1 mg/mL.

Mass spectrometry analysis

For intact mass analysis, samples were injected without prior sample preparation into the UltiMate™ 3000 UHPLC system coupled to the Orbitrap Eclipse™ Tribrid™ mass spectrometer (all Thermo Fisher Scientific). Proteins were loaded on an ACQUITY UPLC BEH200 SEC, 1.7 µm, 4.6×150 mm, applying a 10 min isocratic method (20% B), with a flow rate of 0.2 mL/min and 0.1% Formic acid (FA) (Fisher Chemical, LC–MS grade) in water as mobile phase A and 0.1% FA in Acetonitrile (ACN) (Merck, Hypergrade for LC–MS) as mobile phase B. Electrospray ionisation was performed in positive ionisation mode and molecules analysed in the Orbitrap with a scan range of 500–2000 m/z and a resolution set to 240 000 (at 200 m/z) for full scan.

For DSB analysis, proteins were precipitated with $CHCl_3$/Methanol, dried at RT and subsequently dissolved in lysis buffer (7.6 M urea/50 mM Tris–HCl, pH 8), diluted with 50 mM Tris–HCl, at pH 8 and digested with Trypsin/Lys-C Mix (Promega).

Peptides were analysed on the same LC–MS system and mobile phases as for intact mass analysis. Peptides were separated on a ACQUITY UPLC Peptide CSH C18, 130 Å, 1.7 µm, 2.1×150 mm applying a 25 min gradient from 5–30% B, increasing further to 95% B within 5 min, resulting in total run time of 44 min, with a flow rate of 0.25 mL/min. Electrospray ionization was performed in positive ionization mode, the resolution was set to 120 000 (at 200 m/z), with a scan range of 200–2000 m/z for MS1 analysis. A Top N method was applied for fragmentation with CID Assisted Collision and resulting fragments analysed in the Orbitrap at a resolution of 30 000 (at 200 m/z).

The raw data files were subjected to the Byos software (v 4.2) from Protein Metrics Inc. for data processing and reporting. For intact mass evaluation, peaks found in the total ion chromatogram were integrated and full mass spectra were deconvoluted. For DSB analysis, the Byos DSB workflow was used, searching against a built-in database based on the sequence of the POI.

1D-^{1}H-NMR analysis

Samples were dissolved in a buffer containing 20 mM Phosphate, 20 mM NaCl, at pH 5.5 and were analysed by 1D-^{1}H-NMR after extensive dialysis in a common pool of buffer to reduce effects due to any systematic differences in sample preparation. Before NMR spectra were collected 5% D_2O was added as a lock solvent. Spectra were acquired at 25 ℃ using a double pulse field gradient spin echo sequence (DPFGSE) to suppress water [15] on a Bruker Advance III spectrometer operating at 600 MHz with a He cooled QCI-P cryogenic probe using Topspin 3.6.1. Spectra were measured using 32 k complex data points over a sweep width of 9615 Hz using 1024 scans with an inter-scan relaxation delay of 1 s and 4 dummy scans for equilibration. These data were processed and analysed using Topspin and an exponential window function of 3 Hz was applied to improve signal to noise.

Abbreviations

CEX	Cation exchange chromatography
DO	Dissolved oxygen
DoE	Design of experiment
DSB	Disulphide bond
DSC	Differential scanning calorimetry
IB	Inclusion body
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thioglactopyranoside
MF	Multifermenter
MS	Mass spectrometry
MWCO	Molecular weight cut-off
NMR	Nuclear magnetic resonance
OD	Optical density
PCN	Plasmid copy number
POI	Protein of interest
PTM	Post-translational modification
scFv	Single-chain variable fragment
SP	Signal peptide

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12934-023-02111-4.

Additional file 1. Figure S1. scFvM purification via two steps chromatography. **A** Representative Coomassie blue-stained gel of the scFvM produced with W3110 and purified via IMAC. The samples analysed are the total protein sample loaded (L), column flow-though (FT), eluates (E) from either the periplasm or the culture supernatant and the purified protein after dialysis (D) in non-reducing condition. Ladder (Mark12™ Unstained Standard) on the left in kDa. **B** Representative Coomassie blue-stained gel of a CEX run in non-reducing condition on the scFvM produced with the W3110 strain. peak N°2 elution fractions are in the continued box while peak N°5 ones are in the dotted box.

Acknowledgements
The authors acknowledge Dr. Laura Niederstaetter, Dr. Martin Voigtmann, Patricia Cantarell-Bargueño, and Dr. Nicole Weiner for carrying out, respectively, the mass spectrometry, immunoassay, and DSC analyses. Additionally, the authors acknowledge Alma Medimorec, Gabriela Bachleda-Kubanska and Michael Deimel for their technical support during fermentation runs.

Author contributions
KAA, LB, KK performed the research. GT performed NMR experiments and analysis. MB and CR conceived and coordinated the study. All authors read and approved the final manuscript.

Funding
This work was funded by the People Programme (Marie Skłodowska-Curie Actions) of the European Union's Horizon 2020 Programme under REA grant agreement no. 813979 (SECRETERS).

Availability of data and materials
The data used and /or analysed during the current study are available from the corresponding authors on reasonable requests.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
All authors declare that they have no competing interests.

Author details
[1]School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK. [2]Biopharma Austria, Process Science, Boehringer-Ingelheim RCV GmbH & Co KG, Dr. Boehringer-Gasse 5-11, 1121 Vienna, Austria. [3]Wellcome Trust Biological NMR Facility, School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK.

Received: 17 January 2023 Accepted: 1 May 2023
Published online: 19 May 2023

References

1. Ali S, Perez-Pardo MA, Aucamp JP, et al. Characterization and feasibility of a miniaturized stirred tank bioreactor to perform E. coli high cell density fed-batch fermentations. Biotechnol Progr. 2012;28:66–75. https://doi.org/10.1002/btpr.708.
2. Baumann P, Hubbuch J. Downstream process development strategies for effective bioprocesses: trends, progress, and combinatorial approaches. Eng Life Sci. 2017;17:1142–58. https://doi.org/10.1002/elsc.201600033.
3. Beck A, Liu H. Macro-and micro-heterogeneity of natural and recombinant IgG antibodies. Antibodies. 2019;8:18. https://doi.org/10.3390/antib8010018.
4. Beyer B, Schuster M, Jungbauer A, Lingg N. Microheterogeneity of recombinant antibodies: analytics and functional impact. Biotechnol J. 2018;13:1700476. https://doi.org/10.1002/biot.201700476.
5. Brorson K, Jia AY. Therapeutic monoclonal antibodies and consistent ends: terminal heterogeneity, detection, and impact on quality. Curr Opin Biotech. 2014;30:140–6. https://doi.org/10.1016/j.copbio.2014.06.012.
6. Carrió MM, Villaverde A. Construction and deconstruction of bacterial inclusion bodies. J Biotechnol. 2002;96:3–12. https://doi.org/10.1016/s0168-1656(02)00032-9.
7. Chen H, Huang Z, Dutta K, et al. Cracking the molecular origin of intrinsic tyrosine kinase activity through analysis of pathogenic gain-of-function mutations. Cell Rep. 2013;4:376–84. https://doi.org/10.1016/j.celrep.2013.06.025.
8. Edwardraja S, Neelamegam R, Ramadoss V, et al. Redesigning of anti-c-met single chain fv antibody for the cytoplasmic folding and its structural analysis. Biotechnol Bioeng. 2010;106:367–75. https://doi.org/10.1002/bit.22702.
9. Fekete S, Beck A, Guillarme D. Characterization of cation exchanger stationary phases applied for the separations of therapeutic monoclonal antibodies. J Pharmaceut Biomed. 2015;111:169–76. https://doi.org/10.1016/j.jpba.2015.03.041.
10. Gaza-Bulseco G, Li B, Bulseco A, Liu H. Method to differentiate asn deamidation that occurred prior to and during sample preparation of a monoclonal antibody. Anal Chem. 2008;80:9491–8. https://doi.org/10.1021/ac801617u.
11. Gupta S, Jiskoot W, Schöneich C, Rathore AS. Oxidation and deamidation of monoclonal antibody products: potential impact on stability, biological activity, and efficacy. J Pharm Sci. 2021;111:903–18. https://doi.org/10.1016/j.xphs.2021.11.024.
12. Harris RJ, Kabakoff B, Macchi FD, et al. Identification of multiple sources of charge heterogeneity in a recombinant antibody. J Chromatogr B Biomed Sci Appl. 2001;752:233–45. https://doi.org/10.1016/s0378-4347(00)00548-x.
13. Horga LG, Halliwell S, Castiñeiras TS, et al. Tuning recombinant protein expression to match secretion capacity. Microb Cell Fact. 2018;17:199. https://doi.org/10.1186/s12934-018-1047-z.
14. Huleani S, Roberts MR, Beales L, Papaioannou EH. *Escherichia coli* as an antibody expression host for the production of diagnostic proteins: significance and expression. Crit Rev Biotechnol. 2021;42:756–73. https://doi.org/10.1080/07388551.2021.1967871.
15. Hwang TL, Shaka AJ. Water suppression that works. excitation sculpting using arbitrary wave-forms and pulsed-field gradients. J Magnetic Reson Ser. 1995;112:275–9. https://doi.org/10.1006/jmra.1995.1047.
16. Janzen NH, Striedner G, Jarmer J, et al. Implementation of a fully automated microbial cultivation platform for strain and process screening. Biotechnol J. 2019;14:1800625. https://doi.org/10.1002/biot.201800625.

17. Jiang P, Li F, Ding J. Development of an efficient LC-MS peptide mapping method using accelerated sample preparation for monoclonal antibodies. J Chromatogr B. 2020;1137:121895.
18. Jing S-Y, Gou J-X, Gao D, et al. Separation of monoclonal antibody charge variants using cation exchange chromatography: resins and separation conditions optimization. Sep Purif Technol. 2020;235:116136.
19. Johnson CM. Differential scanning calorimetry as a tool for protein folding and stability. Arch Biochem Biophys. 2013;531:100–9. https://doi.org/10.1016/j.abb.2012.09.008.
20. Jones AS, Austerberry JI, Dajani R, et al. 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—Mol Cell Res. 2016;1863:3116–24. https://doi.org/10.1016/j.bbamcr.2016.09.006.
21. Kang D, Kim Y, Cha H. Comparison of green fluorescent protein expression in two industrial *Escherichia coli* strains, BL21 and W3110, under co-expression of bacterial hemoglobin. Appl Microbiol Biot. 2002;59:523–8. https://doi.org/10.1007/s00253-002-1043-3.
22. Khawli LA, Goswami S, Hutchinson R, et al. Charge variants in IgG1. MAbs. 2010;2:613–24. https://doi.org/10.4161/mabs.2.6.13333.
23. Kipriyanov SM. Antibody phage display, methods and protocols. Methods Mol Biology. 2009;562:205–14. https://doi.org/10.1007/978-1-60327-302-2_16.
24. Kleiner-Grote GRM, Risse JM, Friehs K. Secretion of recombinant proteins from E. coli. Eng Life Sci. 2018;18:532–50. https://doi.org/10.1002/elsc.201700200.
25. Koopaei NN, Khadiv-Parsi P, Khoshayand MR, et al. Optimization of rPDT fusion protein expression by *Escherichia coli* in pilot scale fermentation: a statistical experimental design approach. AMB Express. 2018;8:135. https://doi.org/10.1186/s13568-018-0667-3.
26. Kumar J, Bhat SU, Rathore AS. Slow post-induction specific growth rate enhances recombinant protein expression in *Escherichia coli*: pramlintide multimer and ranibizumab production as case studies. Process Biochem. 2022;114:21–7. https://doi.org/10.1016/j.procbio.2022.01.009.
27. Kwan AH, Mobli M, Gooley PR, et al. Macromolecular NMR spectroscopy for the non-spectroscopist. Febs J. 2011;278:687–703. https://doi.org/10.1111/j.1742-4658.2011.08004.x.
28. Lee HJ, Lee CM, Kim K, et al. Purification of antibody fragments for the reduction of charge variants using cation exchange chromatography. J Chromatogr B. 2018;1080:20–6. https://doi.org/10.1016/j.jchromb.2018.01.030.
29. Liu H, Ponniah G, Zhang H-M, et al. In vitro and in vivo modifications of recombinant and human IgG antibodies. MAbs. 2014;6:1145–54. https://doi.org/10.4161/mabs.29883.
30. Malherbe G, Humphreys DP, Dav E. A robust fractionation method for protein subcellular localization studies in *Escherichia coli*. Biotechniques. 2019;66:171–8. https://doi.org/10.2144/btn-2018-0135.
31. Marisch K, Bayer K, Scharl T, et al. A comparative analysis of industrial *Escherichia coli* K–12 and b strains in high-glucose batch cultivations on process- transcriptome- and proteome level. PLoS ONE. 2013;8:e70516.
32. Martos-Maldonado MC, Hjuler CT, Sørensen KK, et al. Selective N-terminal acylation of peptides and proteins with a Gly-His tag sequence. Nat Commun. 2018;9:3307. https://doi.org/10.1038/s41467-018-05695-3.
33. Meier S, Jensen PR, Duus JØ. Direct Observation of metabolic differences in living *Escherichia coli* strains K-12 and BL21. ChemBioChem. 2012;13:308–10. https://doi.org/10.1002/cbic.201100654.
34. Moorhouse KG, Nashabeh W, Deveney J, et al. Validation of an HPLC method for the analysis of the charge heterogeneity of the recombinant monoclonal antibody IDEC-C2B8 after papain digestion1Presented at the well characterized biotechnology pharmaceuticals meeting in San Francisco, 6–8 January 1997.1. J Pharmaceut Biomed. 1997;16:593–603. https://doi.org/10.1016/s0731-7085(97)00178-7.
35. Nascimento A, Pinto IF, Chu V, et al. Studies on the purification of antibody fragments. Sep Purif Technol. 2018;195:388–97. https://doi.org/10.1016/j.seppur.2017.12.033.
36. Noronha SB, Yeh HJC, Spande TF, Shiloach J. Investigation of the TCA cycle and the glyoxylate shunt in *Escherichia coli* BL21 and JM109 using 13C-NMR/MS. Biotechnol Bioeng. 2000;68:316–27. https://doi.org/10.1002/(sici)1097-0290(20000505)68:3%3c316::aid-bit10%3e3.0.co;2-2.
37. Nowak C, Ponniah G, Neill A, Liu H. Characterization of succinimide stability during trypsin digestion for LC-MS analysis. Anal Biochem. 2017;526:1–8. https://doi.org/10.1016/j.ab.2017.03.005.
38. Ouellette D, Alessandri L, Chin A, et al. Studies in serum support rapid formation of disulfide bond between unpaired cysteine residues in the VH domain of an immunoglobulin G1 molecule. Anal Biochem. 2010;397:37–47. https://doi.org/10.1016/j.ab.2009.09.027.
39. Panula-Perälä J, Šiurkus J, Vasala A, et al. Enzyme controlled glucose auto-delivery for high cell density cultivations in microplates and shake flasks. Microb Cell Fact. 2008;7:31–31. https://doi.org/10.1186/1475-2859-7-31.
40. Ren D, Pipes GD, Liu D, et al. An improved trypsin digestion method minimizes digestion-induced modifications on proteins. Anal Biochem. 2009;392:12–21. https://doi.org/10.1016/j.ab.2009.05.018.
41. Rettenbacher LA, Arauzo-Aguilera K, Buscajoni L, et al. Microbial protein cell factories fight back? Trends Biotechnol. 2021;40:576–90. https://doi.org/10.1016/j.tibtech.2021.10.003.
42. Riggs DL, Gomez SV, Julian RR. Sequence and Solution effects on the prevalence of d-isomers produced by deamidation. Acs Chem Biol. 2017;12:2875–82. https://doi.org/10.1021/acschembio.7b00686.
43. Rosano GL, Ceccarelli EA. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front Microbiol. 2014;5:172. https://doi.org/10.3389/fmicb.2014.00172.
44. Rudge SR, Nims RW. ICH quality guidelines: an implementation guide. In: Teasdale Andrew, Elder David, Nims Raymond W, editors. ICH quality guidelines: an implementation guide. Hoboken: John Wiley & Sons Inc; 2017.
45. Sandomenico A, Sivaccumar JP, Ruvo M. Evolution of *Escherichia coli* expression system in producing antibody recombinant fragments. Int J Mol Sci. 2020;21:6324. https://doi.org/10.3390/ijms21176324.
46. Schofield DM, Templar A, Newton J, Nesbeth DN. Promoter engineering to optimize recombinant periplasmic Fab′ fragment production in *Escherichia coli*. Biotechnol Progr. 2016;32:840–7. https://doi.org/10.1002/btpr.2273.
47. Shiloach J, Kaufman J, Guillard AS, Fass R. *Escherichia coli* BL21 (hDE3) and *Escherichia coli* JM109. Biotechnol Bioeng. 1996;49:421–8. https://doi.org/10.1002/(sici)1097-0290(19960220)49:4%3c421::aid-bit9%3e3.0.co;2-r.
48. Shiloach J, Rinas U. Glucose and acetate metabolism in E coli—system level analysis and biotechnological applications in protein production processes. In: Sang YL, editor. Systems biology and biotechnology of *Escherichia coli*. Dordrecht: Springer, Netherlands; 2009.
49. Singh SK, Kumar D, Malani H, Rathore AS. LC–MS based case-by-case analysis of the impact of acidic and basic charge variants of bevacizumab on stability and biological activity. Sci Rep-uk. 2021;11:2487. https://doi.org/10.1038/s41598-020-79541-2.
50. Striedner G, Pfaffenzeller I, Markus L, et al. Plasmid-free T7-based *Escherichia coli* expression systems. Biotechnol Bioeng. 2010;105:786–94. https://doi.org/10.1002/bit.22598.
51. Tong L, Lin Q, Wong WKR, et al. Extracellular expression, purification, and characterization of a winter flounder antifreeze polypeptide from *Escherichia coli*. Protein Expres Purif. 2000;18:175–81. https://doi.org/10.1006/prep.1999.1176.
52. Tripathi NK. Production and purification of recombinant proteins from *Escherichia coli*. Chembioeng Rev. 2016;3:116–33. https://doi.org/10.1002/cben.201600002.
53. Tripathi NK, Shrivastava A. Recent developments in bioprocessing of recombinant proteins: expression hosts and process development. Frontiers Bioeng Biotechnol. 2019;7:420. https://doi.org/10.3389/fbioe.2019.00420.
54. Wagner-Rousset E, Fekete S, Morel-Chevillet L, et al. Development of a fast workflow to screen the charge variants of therapeutic antibodies. J Chromatogr A. 2017;1498:147–54. https://doi.org/10.1016/j.chroma.2017.02.065.
55. Xia X, Han M, Lee SY, Yoo J. Comparison of the extracellular proteomes of *Escherichia coli* B and K-12 strains during high cell density cultivation. Proteomics. 2008;8:2089–103. https://doi.org/10.1002/pmic.200700826.
56. Yoon SH, Han M-J, Jeong H, et al. Comparative multi-omics systems analysis of *Escherichia coli* strains B and K-12. Genome Biol. 2012;13:R37–R37. https://doi.org/10.1186/gb-2012-13-5-r37.

57. Yüce M, Sert F, Torabfam M, et al. Fractionated charge variants of biosimilars: a review of separation methods, structural and functional analysis. Anal Chim Acta. 2021;1152:238189.
58. Zhang G, Brokx S, Weiner JH. Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in *Escherichia coli*. Nat Biotechnol. 2006;24:100–4. https://doi.org/10.1038/nbt1174.

Publisher's Note

Annex 6: Publication 2. Highly efficient export of a disulfide-bonded protein to the periplasm and medium by the Tat pathway using CyDisCo in *Escherichia coli*

Received: 27 October 2022 | Accepted: 24 February 2023

DOI: 10.1002/mbo3.1350

ORIGINAL ARTICLE

MicrobiologyOpen Open Access WILEY

Highly efficient export of a disulfide-bonded protein to the periplasm and medium by the Tat pathway using CyDisCo in *Escherichia coli*

Klaudia Arauzo-Aguilera[1] | Mirva J. Saaranen[2] | Colin Robinson[1] | Lloyd W. Ruddock[2]

[1]School of Biosciences, University of Kent, Canterbury, UK

[2]Faculty of Biochemistry and Molecular Medicine, University of Oulu, Oulu, Finland

Correspondence
Lloyd W. Ruddock, Faculty of Biochemistry and Molecular Medicine, University of Oulu, 90220 Oulu, Finland.
Email: lloyd.ruddock@oulu.fi

Funding information
H2020 Marie Sklodowska-Curie Actions

Abstract

High-value heterologous proteins produced in *Escherichia coli* that contain disulfide bonds are almost invariably targeted to the periplasm via the Sec pathway as it, among other advantages, enables disulfide bond formation and simplifies downstream processing. However, the Sec system cannot transport complex or rapidly folding proteins, as it only transports proteins in an unfolded state. The Tat system also transports proteins to the periplasm, and it has significant potential as an alternative means of recombinant protein production because it transports fully folded proteins. Most of the studies related to Tat secretion have used the well-studied TorA signal peptide that is Tat-specific, but this signal peptide also tends to induce degradation of the protein of interest, resulting in lower yields. This makes it difficult to use Tat in the industry. In this study, we show that a model disulfide bond-containing protein, YebF, can be exported to the periplasm and media at a very high level by the Tat pathway in a manner almost completely dependent on cytoplasmic disulfide formation, by other two putative Tat SPs: those of MdoD and AmiC. In contrast, the TorA SP exports YebF at a low level.

KEYWORDS
CyDisCo, disulfide bond, *Escherichia coli*, periplasm, signal peptide, Tat pathway

1 | INTRODUCTION

High-value heterologous proteins produced in *Escherichia coli* that contain disulfide bonds (DSB) are almost invariably targeted to the periplasm via the general secretory (Sec) pathway by means of a cleavable N-terminal signal peptide (SP) (Mirzadeh et al., 2020). This guides newly synthesized proteins through the SecYEG membrane channel in an unfolded state. Once across the membrane, the SP is cleaved and proteins fold in the periplasm, acquiring DSBs where appropriate (Kleiner-Grote et al., 2018). This protein export approach offers several advantages to produce therapeutic proteins, such as (i) it enables disulfide bond formation, which only occurs in the periplasm in wild-type cells, (ii) it facilitates protein isolation from the relatively small periplasmic proteome, (iii) allows control of the nature of the N-terminus of the mature protein, and (iv) minimizes exposure to cytoplasmic proteases (Karyolaimos & de Gier, 2021).

Some proteins, however, fold too rapidly for the Sec system to handle, or require co-factor insertion in the cytoplasm, thereby

precluding translocation via the Sec system. The twin-arginine translocation (Tat) pathway offers a potential alternative method of localizing proteins to the periplasm and, unlike Sec, this system transports fully folded proteins. As with Sec substrates, Tat substrates are synthesized with N-terminal SPs, but these contain Tat-specific determinants including the presence of a highly conserved twin-arginine motif (reviewed by Natale et al., 2008).

An additional, and indeed unique, feature of the Tat pathway is its in-built proofreading mechanism that can detect structurally incorrect substrates and reject them for export. This proofreading capability could allow for a more homogeneous product to be produced in the periplasm (reviewed by Frain et al., 2019). Furthermore, the recent development of TatExpress (TE) strains, that over-express the *tatABC* genes (encoding the Tat system) from the chromosome boosts the industrial potential of this pathway (Browning et al., 2017). However, this quality control poses problems for the export of disulfide-bonded proteins, because such proteins often only obtain a native conformation after the formation of DSB. CyDisCo strains (cytoplasmic disulfide bond formation in *E. coli*), express a catalyst of disulfide bond formation, usually the sulfhydryl oxidase Erv1p, and a catalyst of disulfide isomerization, usually human protein disulfide isomerase (PDI). This system facilitates DSB formation in the cytoplasm of wild-type *E. coli* and may therefore allow the efficient production of DSB-containing proteins in the cytoplasm before their export via the Tat pathway (Matos et al., 2014).

The efficiency of protein secretion varies depending on the host strain, signal sequence, and the type of protein to be secreted (Freudl, 2018). There are at least 29 SPs in the *E. coli* genome that contain a twin-arginine motif characteristic of proteins exported via the Tat pathway. However, many of these SPs are not completely Tat-specific and can lead to the secretion of the protein of interest (POI) via Sec, Tat, or both depending on the POI, strain, or media used. They are sometimes termed promiscuous SPs (Bendtsen et al., 2005; Tullman-Ercek et al., 2007). In contrast, the TorA SP, an *E. coli* Tat SP derived from pre-trimethylamine *N*-oxide (TMAO) reductase (TorA), is a very well-studied SP that is reported to be Tat-specific (Blaudeck et al., 2001; Jack et al., 2004; Lee et al., 2006). However, the use of the TorA SP also tends to induce degradation of the protein of interest, for unknown reasons. This results in low yields of some POI, as seen in the absence of precursor forms in many export studies (e.g., Alanen et al., 2015).

While it has been shown that Tat can export some DSB-ed proteins at high levels in fed-batch fermentation studies (e.g., hGH; Guerrero Montero et al., 2019a) and shake-flask studies (Alanen et al., 2015; Browning et al., 2017; DeLisa et al., 2003; Matos et al., 2014; Tullman-Ercek et al., 2007), it has yet to be determined whether it can efficiently export a protein that requires DSB formation for correct folding. In this study, we used YebF, a 10.8 kDa *E. coli* protein of unknown function that contains a single DSB. It has been previously reported that recombinant YebF is secreted by laboratory strains of *E. coli* into the extracellular medium after first being translocated into the periplasm by the Sec-system (Zhang et al., 2006). A wide variety of proteins, including N-glycosylated protein domains, are readily secreted into the growth medium via fusion with YebF (Fisher et al., 2011; Haitjema et al., 2014).

Here, we report that YebF can be exported to the periplasm and media by the Tat pathway in an almost completely CyDisCo-dependent manner. The use of TorA SP results in low yields, consistent with other POI previously examined. However, we show that two other Tat SPs, namely MdoD and AmiC, direct very high levels of protein export. While both may be capable of directing export by Sec, expression in TatExpress strains results in an increase in export flux and expression without CyDisCo or in ΔTat cells inhibits export, which all indicate that export is largely carried out by Tat.

2 | MATERIALS AND METHODS

All chemicals used in this study were supplied by Fisher Scientific (Thermo Fisher Scientific Inc.), Sigma (Sigma-Aldrich), or Formedium unless otherwise stated.

2.1 | Cloning

The desired gene fragments were obtained as synthetic genes from GeneArt or by PCR from *E. coli* genomic DNA and cloned with restriction digestion and ligation into a modified pET23-based vector with a pTac promoter replacing the T7 promoter (Gaciarz et al., 2016). The vector design allows for the incorporation of a C-terminal hexahistidine tag (-Leu-Glu-6xHis) into the expressed protein. For YebF without the signal sequence, a vector incorporating both N- (Met-6xHis-) and C-terminal hexahistidine tags was used. For constructs with signal sequences, the signal sequence plus the first four amino acids of the mature protein was added to YebF, to increase the likelihood of efficient processing of the signal sequence by the signal peptidase. The net effect of this is that each purified mature construct differs slightly from the other, resulting in mass differences by mass spectrometry and small mobility differences in SDS-PAGE. The gene inserts in the plasmids made were fully sequenced before use (see Table 1 for plasmid names and details).

2.2 | Expression

Plasmids with the gene of interest together with the plasmid containing the CyDisCo components (pMJS205) or empty plasmid (pAG82) were cotransformed into chemically competent *E. coli* cells and spread onto lysogeny broth (LB) agar plates supplemented with 35 μg/mL of chloramphenicol and 100 μg/mL of ampicillin for selection. After overnight incubation at 37°C these were used to inoculate 2–5 mL of LB media supplemented with 2 g/L of glucose, 35 μg/mL of chloramphenicol, and 100 μg/mL of

TABLE 1 Strains and constructs used in this study.

Strain/plasmid	Description	Source/reference
BL21	*E. coli* B F^- *dcm ompT lon hsdS*(r_B^- m_B^-) *gal*	Agilent
BL21 TatExpress	BL21 carrying a pTac promotor upstream of tatABCD	Browning et al. (2017)
ΔTatABCDE (ΔTat)	MC4100 strain (*Ara*^{R}, *F2 araD139 DlacU169 rpsL150 relA1 flB5301 deoC1 ptsF25 rbs*^{R}) lacking tat*ABCDE* genes, *Ara*^{R}	Wexler et al. (2000)
pMJS289	YebF (A22-R118)[a]	This study
pMJS285	AmiC (M1-Q35)-YebF (A22-R118)[b]	This study
pMJS284	MdoD (M1-D36)-YebF (A22-R118)[b]	This study
pMJS288	TorA (M1-A43)-YebF (A22-R118)[b]	This study
pMJS205	CyDisCo: Erv1p and PDI	Gaciarz et al. (2016)
pAG82	empty	Gaciarz et al. (2016)

[a]With N-terminal Met-6xHis and C-terminal -Leu-Glu-6xHis -tags.

[b]With C-terminal -Leu-Glu-6xHis -tag.

ampicillin. These starter cultures were grown 6 h, or overnight for ΔTat experiments, at 30°C, 250 rpm (2.5 cm radius of gyration), and were used to seed the cultures in a 1:100 ratio.

Expression tests to screen for optimal SP were carried out for the constructs in 24 deep well plates (DWP). The constructs were expressed alone or co-expressed with CyDisCo components in *E. coli* BL21 in 3 mL per well of terrific broth autoinduction media (AIM—Terrific Broth Base including Trace elements, Formedium) supplemented with 0.8% glycerol, 35 μg/mL of chloramphenicol and 100 μg/mL of ampicillin. The DWP was covered with air permeable membrane (Thomson) and the cultures were grown at 30°C, 250 rpm, and harvested after 24 h. The cells were collected by centrifugation at 3220*g* for 20 min at 4°C and resuspended in 3 mL of 50 mM sodium phosphate pH 7.4, 20 μg/ml DNase, 0.1 mg/mL egg white lysozyme. After 10 min incubation, the resuspended cultures were frozen at −20°C. Cells were lysed by freeze-thawing.

Main cultures with selected constructs in either BL21 or BL21 TatExpress cells were grown in 100 mL flasks with 10 mL culture in each flask. The flasks were covered with oxygen-permeable membranes (Thomson) to ensure proper oxygenation of the cultures, grown at 30°C, 250 rpm for 24 h, and harvested for fractionation.

For ΔTat experiments, cultures were grown in LB media at 30°C, 250 rpm in shake flasks until the OD_{600} of the cultures reached approximately 0.5 and were then induced with 50 μM isopropyl ß-D-1-thiogalactopyranoside for 2 h before harvesting.

2.3 | Fractionation of the cells

For the fractionation of the cells, the PureFrac fractionation protocol was used (Malherbe et al., 2019). For purification, ethylenediamine tetraacetic acid (EDTA) was not added to any of the buffers. Apart from the periplasm and cytoplasm, medium samples were also recovered (same volume in all cultures). In ΔTat experiments, the 1× phosphate-buffered saline wash of the cells and the separation of the cytoplasm and insoluble fraction was not carried out to avoid extra manipulation of this cell line due to its fragility. Samples were prepared for sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) analysis in reducing conditions.

2.4 | Purification of cytoplasmic, periplasmic, and medium samples, SDS-PAGE analysis, and western blot (WB) analysis

Purification of hexahistidine-tagged proteins was performed by standard immobilized metal affinity chromatography using HisPur Cobalt resin (Thermo Scientific) under native conditions. For 3 mL cultures from 24 DWP, IMAC was performed using 0.2 mL resin in small gravity feed columns. The resin was washed with 2 × 2 mL of water and equilibrated with 2 × 2 mL of 50 mM phosphate buffer (pH 7.4). Cell lysates on 24 DWP were cleared by centrifugation (3220*g*, 20 min, 4°C) and loaded onto the columns. The columns were rinsed with 2 mL of 50 mM phosphate buffer (pH 7.4), washed with 4 × 2 mL of wash buffer (50 mM sodium phosphate, 10 mM imidazole, 300 mM sodium chloride; pH 7.4), and then rinsed with 2 mL of 50 mM sodium phosphate (pH 7.4) before elution with 3 × 0.2 mL of 50 mM sodium phosphate, 50 mM EDTA (pH 7.4). For 10 mL cultures, the same protocol was used with the following changes: medium samples were 1:2 diluted (total volume 10 mL), periplasmic and cytoplasmic fractions were diluted in 2.5 mL of 200 mM sodium phosphate buffer and made up to 10 ml with water to reduce the salt concentration. Samples were prepared for SDS-PAGE analysis and 10 μL were loaded in 4–20% Criterion™ TGX™ Precast Midi Protein Gel, 26 well (BioRad).

For the detection of proteins by WB analysis the method as detailed in Guerrero-Montero, Dolata, et al. (2019) was performed,

TABLE 2 Theoretical oxidized and experimental molecular weights of the proteins in this study.

Construct	Location	Number of cysteine	$M_{theorOx}$ (Da)	M_{exp} (Da)	Δ mass
YebF no SP	Cytoplasm	2	12,937	12,985	48
AmiC-YebF	Periplasm		12,506	12,538	32
	Medium		12,506	12,538	32
MdoD-YebF	Periplasm		12,509	12,541	32
	Medium		12,509	12,541	32
TorA-YebF	Periplasm		12,320	12,352	32
	Medium		12,320	12,352	32

Note: The theoretical monoisotopic molecular weight for oxidized ($M_{theorOx}$) His-tag YebF constructs in dalton (Da) were calculated using the ExPaSy Compute pI/Mw tool (Gasteiger et al., 2005). The experimental molecular weight (M_{exp}) was determined by mass spectrometry. The same masses were obtained with NEM treatment. The results suggest both cysteines are in a disulfide bond in the YebF constructs analyzed.

with the exception that was transferred to the polyvinylidene fluoride-membrane (GE Healthcare) by rapid semi-dry transfer using the Invitrogen Power Blotter XL System according to the manufacturer's instructions.

2.5 | Mass spectrometry

The theoretical oxidized monoisotopic molecular weight ($M_{theorOx}$) of the His-tag YebF constructs in dalton (Da) was calculated using the ExPaSy Compute pI/Mw -tool (Gasteiger et al., 2005) (Table 2). The molecular weights of purified protein samples were measured by electrospray ionization mass spectrometry combined with liquid chromatography (LC-ESI-MS) using a Q Exactive Plus Mass Spectrometer. The protein samples were mixed with trifluoroacetic acid (TFA) to a final concentration of 0.5% before analysis. For N-ethylmaleimide (NEM)-trapped samples the protein was incubated with 20 mM NEM in 50 mM phosphate buffer pH 7.3 with 6 M guanidine-HCl for 10 min and quenched with 0.5% TFA before analysis.

3 | RESULTS

3.1 | Folding of YebF is CyDisCo dependent without its native SP and with the AmiC, MdoD, and TorA SPs

In this study, the first aim was to demonstrate the potential use of other Tat SPs in place of the well-known but non-ideal TorA SP (Alanen et al., 2015). We also sought to test whether a CyDisCo-dependent protein could be exported at high rates by the Tat system; *E. coli* TatExpress cells have been shown to export high levels of human growth hormone (hGH) but while this protein contains two DSB, they are not essential for proper folding and the Tat system can efficiently export the protein in its reduced state (Alanen et al., 2015 and references therein). Tat can only transport fully folded proteins from the cytoplasm to the periplasm, so we hypothesized that a disulfide-containing protein that required disulfide formation to reach a native state, for example, one that was dependent on CyDisCo in our system, would be a good model protein to test the true capabilities of the Tat system. Based on its reported use as a transport mechanism for secretion to the medium (Fisher et al., 2011; Haitjema et al., 2014; Zhang et al., 2006), we initially chose a small disulfide bonded *E. coli* protein as a test protein—YebF.

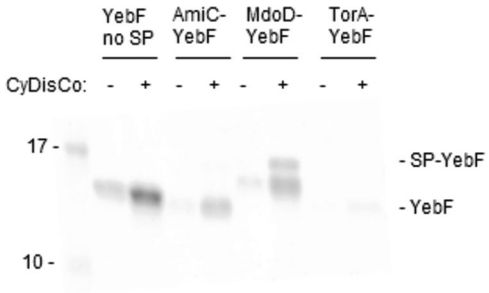


FIGURE 1 Purification of soluble YebF protein without its native SP and with AmiC, MdoD, and TorA SPs in the absence/presence of CyDisCo. Coomassie-stained Criterion™ TGX gels showing the purified soluble YebF (≈12 kDa) in BL21 wild-type in 2 mL rich-autoinduction media at 30°C from 24DWP. All the constructs show a CyDisCo dependency for the production of YebF soluble protein (marked as "-YebF"). MdoD-YebF construct shows uncleaved YebF (shown as "SP-YebF") as well as mature protein (marked as "YebF") in the presence of CyDisCo. SP, signal peptide.

In the absence of an SP (Figure 1, "YebF no SP"), the folding of YebF showed a strong CyDisCo dependence, with purified YebF being far more abundant in the +CyDisCo cells, indicating that CyDisCo improves folding to a significant extent. However, some soluble protein was present in the absence of CyDisCo indicating that the protein can fold to a low extent without CyDisCo. YebF was then tested with the widely used TorA Tat-dependent SP (construct denoted TorA-YebF) along with AmiC and MdoD SPs as potential

alternative Tat SPs (referred to as AmiC-YebF and MdoD-YebF, respectively). In all cases, YebF was expressed with a 6xHis tag on the C-terminus. To achieve efficient SP cleavage, all SPs were fused with four more N-terminal residues to YebF, resulting in differences between the mature YebF dependent on the SP used.

For both the AmiC-YebF and MdoD-YebF constructs there was a strong dependency on CyDisCo to produce a soluble protein (Figure 1), indicating that the protein was probably secreted via the Tat pathway rather than via the Sec pathway. If the proteins were exported by Sec, they would be unfolded until they reached the periplasm where they would rapidly acquire their DSBs, and CyDisCo would not influence their folding. An extra band in the gel (labeled SP-YebF) was observed for MdoD-YebF, which probably represents uncleaved SP-YebF (the POI with the SP still attached). This suggests that for this construct in these conditions, export to the periplasm was limiting. In contrast to the results with AmiC-YebF and MdoD-YebF, the TorA-YebF construct produced very low levels of soluble protein in the presence of CyDisCo and hardly detectable levels in the absence of CyDisCo. While the possibility of issues connected with messenger RNA stability linked to the sequence encoding the SP, initiation of translation initiation, and so forth cannot be excluded, it is likely that this difference reflects inefficient transport by Tat and the protein being degraded in the cytoplasm as observed with other constructs bearing the TorA SP (Alanen et al., 2015).

3.2 | TatExpress cells export much more YebF to the periplasm than standard BL21 cells with AmiC and MdoD SPs

We next assessed the export of the four constructs by fractionating cell samples into the cytoplasm (C), periplasm (P), and medium (M) to discover where and how in the cell or medium this protein was targeted in the presence of CyDisCo. The experiments were carried out in both a standard BL21 strain and the BL21 TatExpress strain which has been engineered for higher levels of Tat-dependent export (Browning et al., 2017). We reasoned that if the AmiC-YebF and MdoD-YebF constructs were exported primarily by Tat, we might observe higher levels of export in the TatExpress cells, if the export was limiting, as observed with hGH (Browning et al., 2017; Guerrero Montero, Richards, et al., 2019), and that total yields might increase if the AmiC or MdoD SP would act similar to the TorA SP and target non-exported protein for degradation.

The control experiment showed that YebF when expressed without an SP, remains in the cytoplasm when expressed in both BL21 and BL21 TatExpress (Figure 2; lanes "C" denote cytoplasmic fraction; "X" denotes TatExpress cells in this and subsequent Figures). No YebF was visible in the periplasm (Figure 2 lanes P, PX). IMAC purification of the protein from fractions confirmed the cytoplasmic localization (lanes denoted "Purification"), with a small amount of YebF being purified from the medium, most likely due to cell lysis.

Since YebF without a signal sequence could be efficiently folded by CyDisCo but was retained in the cytoplasm, we then examined if

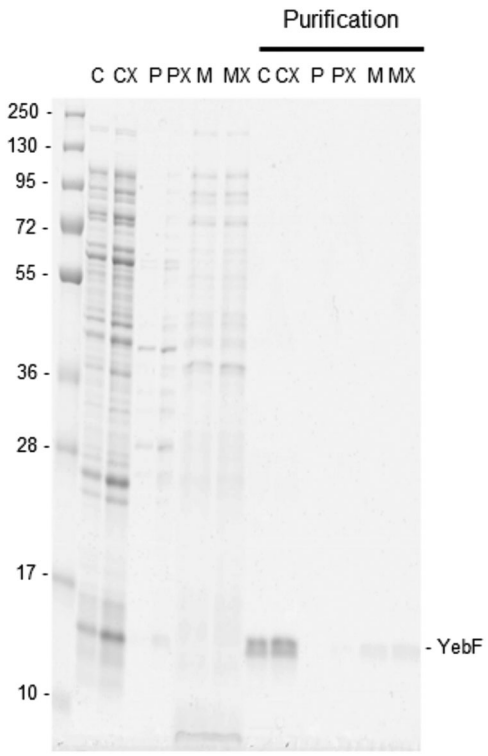


FIGURE 2 Expression of YebF control without SP in BL21 and BL21 TatExpress cells with CyDisCo. Coomassie-stained Criterion™ TGX gel of cytoplasmic, periplasmic fractions and medium samples and purifications (C, P, M) from BL21 and BL21 TatExpress cells (CX, PX, MX) expressing YebF without SP (marked as "YebF"). Samples were collected after 24 h of growth in terrific broth-based autoinduction media at 30°C in shake flasks and processed immediately. Samples from the same subcellular fraction are comparable between strains, but not between different subcellular compartments. Cytoplasm was diluted in 750 µL buffer. Periplasm was diluted in 400 µL buffer. For all medium samples, a 5 mL culture was recovered. Purifications of all cell fractions and medium are comparable among fractions and among strains. Some POI is visible in medium purification probably due to cell lysis. The representative gel from triplicate experiments is shown.

YebF could be exported to the periplasm and, if so, whether this export was increased in TatExpress cells when using the AmiC-, MdoD- and TorA SPs (Figures 3 and 4). The constructs were expressed in BL21 and BL21 TatExpress and the cells fractionated into cytoplasm, periplasm, and medium (C, P, M) with YebF purified by IMAC in the "Purification" panels.

In both constructs and strains, the periplasmic fraction is relatively clean of cytoplasm cross-contaminants as judged by the low levels or absence of major cytoplasmic proteins. Importantly, YebF was the most abundant periplasmic protein after expression of AmiC-YebF and MdoD-YebF, even in standard BL21 cells, and its abundance increased significantly in TatExpress cells (Figure 3a,b lanes PX). IMAC purification of YebF confirmed that the export of both proteins is particularly efficient in TatExpress cells. Purification from the cytoplasmic fractions showed a very faint duplex band for AmiC-YebF and a slightly more prominent

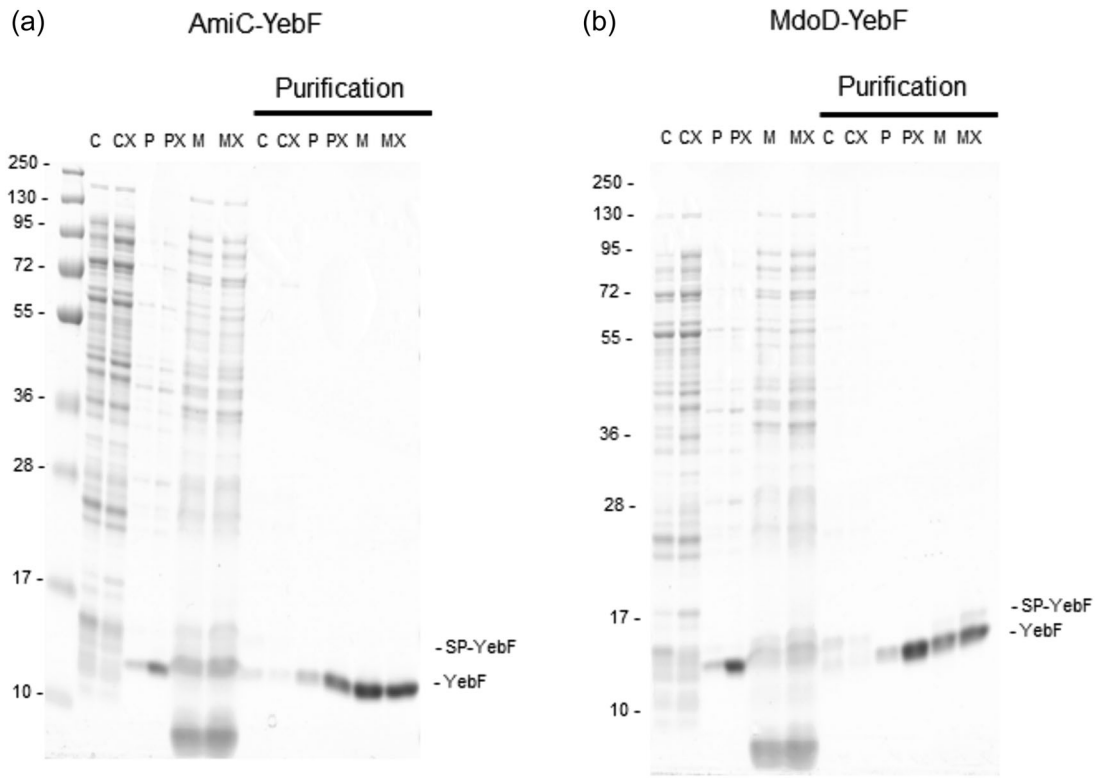


FIGURE 3 Export of AmiC- and MdoD- YebF in BL21 and BL21 TatExpress cells with CyDisCo. Coomassie-stained Criterion™ TGX gel of cytoplasmic, periplasmic fractions and medium samples and purifications (C, P, M) from BL21 and BL21 TatExpress cells (CX, PX, MX) expressing AmiC-YebF, MdoD-YebF, and TorA-YebF (marked as "-YebF" and "SP-YebF"). Samples were analyzed after 24 h of growth in terrific broth-based autoinduction media at 30°C in shake flasks. (a) AmiC-YebF fractionation and purification of BL21 and BL21 TatExpress cells for comparison. (b) MdoD-YebF fractionation and purification of BL21 and BL21 TatExpress for comparison. Samples from the same subcellular fraction are comparable between strains, but not between different subcellular compartments. Cytoplasm was diluted in 750 µL buffer. Periplasm was diluted in 400 µL buffer. Purifications of all cell fractions and medium are comparable among fractions and among strains. For all medium samples, a 5 mL culture was recovered. The representative gel from triplicate experiments is shown.

duplex for the MdoD-YebF construct, which probably represent the SP- and mature forms of YebF. In purifications from periplasm fractions, only the mature form was seen for both constructs. In purifications from the media again the duplex was observed for the MdoD-YebF construct, consistent with some of the protein in the media deriving from cell lysis as per the control YebF without a SP (Figure 2). For both proteins, but in particular, for MdoD-YebF the total amount of YebF observed increased in TatExpress cells (Figure 3b), suggesting that like the TorA SP, nonsecreted YebF retaining the SP might be targeted for degradation in the cytoplasm (a process which will probably be SP, strain, media, and protein-dependent).

To compare the efficiency of the export of these two newly proposed Tat-specific SPs to the well-characterized TorA SP, TorA-YebF was also expressed in BL21 and BL21 TatExpress, and the cells fractionated into cytoplasm, periplasm, and medium (C, P, M) with YebF again purified by IMAC in the "Purification" panels. YebF was also exported to the periplasm and to the medium (Figure 4). While YebF protein was again the most abundant protein in the periplasm, indicating efficient Tat-dependent export, it was considerably less than for the MdoD- or AmiC-SP constructs (Figure 1 and compare Figure 3 vs. Figure 4). Since the TorA SP is reported to be highly Tat-specific (Tullman-Ercek et al., 2007), this export must be by the Tat pathway, which in turn means that the DSB-ed protein is being exported.

Together, the results suggest a very efficient export by the Tat pathway to the periplasm, which was higher for AmiC-YebF and MdoD-YebF compared with TorA-YebF. In addition, purified yields of YebF from the medium fractions were high. This may come partially from lysis but probably arises mainly from the translocation of YebF from the periplasm to the medium by an unknown mechanism (Zhang et al., 2006).

3.3 | Examination of Tat-dependence

The CyDisCo dependency for AmiC-, MdoD-, and TorA-YebF folding and export (Figure 1) and the enhancement of export to the periplasm in TatExpress cells (Figure 3), both are consistent with the hypothesis that the predominant pathway used for all three constructs must be Tat, as Sec will only transport proteins in an unfolded state.

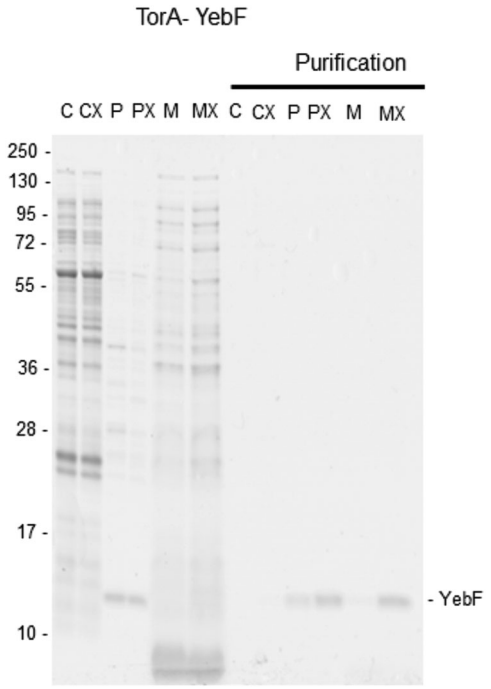


FIGURE 4 Export of TorA-YebF in BL21 and BL21 TatExpress cells with CyDisCo. Coomassie-stained Criterion™ TGX gel of cytoplasmic, periplasmic fractions and medium samples and purifications (C, P, M) from BL21 and BL21 TatExpress cells (CX, PX, MX) TorA-YebF (marked as "-YebF"). Samples were analyzed after 24 h of growth in terrific broth-based autoinduction media at 30°C in shake flasks. Samples from the same subcellular fraction are comparable between strains, but not between different subcellular compartments. Cytoplasm was diluted in 750 µL buffer. Periplasm was diluted in 400 µL buffer. Purifications of all cell fractions and medium are comparable among fractions and among strains. For all medium samples, a 5 mL culture was recovered. The representative gel from triplicate experiments is shown.

To examine this further, we expressed our constructs in a strain (ΔTat) that lacks the *tatABCDE* genes that encode the Tat apparatus. In this strain, any periplasmic export must be via Sec. ΔTat is a relatively fragile strain that tends to lyse more easily than wild-type strains in many growth conditions, has a higher level of proteases, is difficult to subfractionate, and is more sensitive to stress (Harrison et al., 2005; Sargent et al., 1998).

To minimize cross-contamination between fractions, cells were grown in a rich medium that only resulted in low-density cultivation and induced for short times before harvesting. This resulted in lower levels of protein expression and hence WB analysis, using an antibody against the C-terminal his-tag, was used to examine the subcellular localization. Analysis of medium samples suggested very low levels of lysis in these growth conditions (Figure 5).

While no protein was observed for TorA-YebF under these conditions, whether CyDisCo was expressed or not, consistent with it being a poor signal sequence, strong bands were observed for both AmiC-YebF and MdoD-YebF when CyDisCo was present. These bands were observed only in the spheroplast fractions and not in the periplasm or media fractions, consistent with our hypothesis that the export of YebF is predominantly via Tat.

While AmiC-YebF appears to be completely CyDisCo dependent under these expression conditions (similar to the strong CyDisCo-dependence observed in Figure 1), MdoD-YebF shows less CyDisCo-dependence than observed in the previous expression conditions (compare Figure 5 and Figure 1). This effect may arise from the solubilization of folding intermediates promoted by fusion partners (including potentially SPs), with different fusions resulting in different solubilization effects and/or degradation rates. The differences observed between experiments (Figure 5 and Figure 1) may arise from differences in expression conditions (media and time of

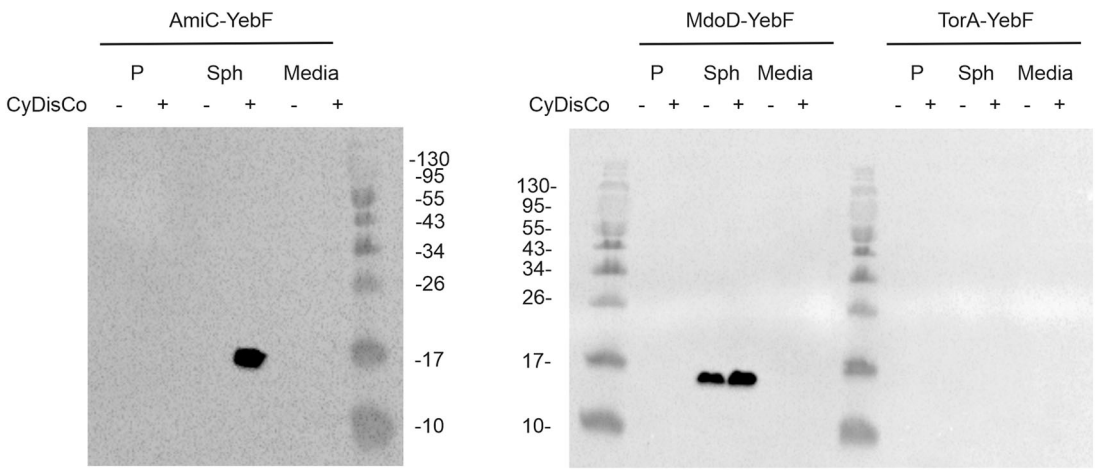


FIGURE 5 Expression of YebF SPs constructs in ΔTat strain with and without CyDisCo. western blot analysis of 15% SDS-PAGE gel of periplasmic (P), spheroplast (Sph) fractions, and medium (M) samples expressing YebF SP constructs. Samples were analyzed 2 h postinduction (50 µM IPTG) in LB media at 30°C in shake flasks. Samples from the same subcellular fraction are comparable between strains, but not between different subcellular compartments. Spheroplast was diluted in 750 µL buffer. Periplasm was diluted in 400 µL buffer. For all medium samples, 500 µL culture was recovered. Note the different SDS-PAGE gel type results in different mobilities c.f. Figures 1–4. LB, lysogeny broth; SDS-PAGE, sodium dodecyl sulphate polyacrylamide gel electrophoresis; SP, signal peptide.

induction) or cells (including higher levels of proteases in ΔTat cells) or the very different levels of protein produced.

3.4 | YebF control and SP-YebF constructs are folded and correctly cleaved from SPs in the periplasm and media

A final experiment to verify the folding and the correct cleavage of the SPs of the constructs in the periplasm and medium was carried out by subjecting the purified proteins to ESI-Mass Spectrometry. The theoretical monoisotopic molecular weight for oxidized ($M_{theorOx}$) YebF with no SP purified from the cytoplasm is 12,937 Da and we observed an experimental molecular weight (M_{exp}) of 12,985 Da, which showed a difference of 48 Da (see Table 2). The $M_{theorOx}$ of AmiC-YebF purified from the periplasm and medium is 12,506 Da and we observed an M_{exp} of 12,538 Da, which showed a difference of 32 Da. For MdoD-YebF purified from the periplasm and medium, the $M_{theorOx}$ is 12,509 Da and we observed an M_{exp} of 12,541 Da, which also showed a difference of 32 Da. For TorA-YebF purified from the periplasm and medium, the $M_{theorOx}$ is 12,320 Da and we observed an M_{exp} of 12,352 Da, which also showed a difference of 32 Da. Overall, analysis by ESI-MS confirmed that YebF purified from the cytoplasm (in the YebF no SP construct only), periplasm, and medium had the expected molecular weight, consistent with the cleavage of the respective SP when the POI is exported from the cytoplasm to the periplasm and its two cysteines form of a disulfide bond. The presence of free cysteines in the constructs was further evaluated by treating the samples with NEM before mass spectrometric analysis. NEM-trapped samples would be expected to show an addition of 125 Da in the molecular weight of the protein for each free cysteine modified. None of the samples analyzed showed any increase in the mass after NEM treatment, implying that none contained free thiols (Nguyen et al., 2011; Saaranen et al., 2010). The change of mass of 48 Da (for YebF no SP) and 32 Da for the other three constructs are due to the oxidation of methionine in the purified proteins (three and two oxygen molecules, respectively). The handling, time of storage, and/or repeatedly freeze-thawed, were responsible for this phenomenon (Grassi & Cabrele, 2019) (Table 2).

4 | DISCUSSION

E. coli production platforms are extensively used for the production of biotherapeutics, but current platforms have limitations in terms of the types of protein that they can handle. Typically, target proteins are refolded from inclusion bodies or targeted to the periplasm where they fold to a native state, the latter being the less time-consuming and labor-intensive approach (Bhatwa et al., 2021). Tat-based platforms offer significant advantages for the production of some molecules, but the system has not been validated using a wide range of proteins, especially proteins that require disulfide bonding to fold correctly, and there is a clear need to find an alternative to TorA as the commonly used Tat specific SP. The vast majority of export studies by the Tat pathway have been carried out using the TorA SP although it is known that it can cause POI degradation in the cytoplasm (Blaudeck et al., 2003) and inclusion body formation (Jong et al., 2017). The TorA SP has been shown to export a small number of proteins with high efficiency, but none of these examples required disulfide bonding to occur in the cytoplasm before translocation by Tat (Alanen et al., 2015).

Here, we set out to compare the export of YebF to the periplasm of *E. coli via* the Tat pathway with different SPs: the AmiC and MdoD SPs have been reported to be able to go through both the Sec and Tat pathway (Tullman-Ercek et al., 2007) whereas the TorA SP is reported to be Tat specific.

We examined the export to the periplasm (and medium) of YebF with different SPs in the presence and absence of CyDisCo which is required for YebF to efficiently reach a native state in the cytoplasm. In wild-type cells, YebF was exported to the periplasm and medium by the classical Tat SP TorA, but yields were relatively low—as is often reported for this SP. The use of TatExpress cells increased export to the medium, while no YebF could be observed in a ΔTat strain, as expected with an SP that cannot target proteins through the Sec pathway.

Generally, the results for AmiC SP and MdoD SP mirrored those of the TorA SP, with two exceptions. First, the periplasmic and medium yields of YebF in wild-type and TatExpress cells were far higher than for TorA, that is, they appeared to be much more efficient SPs. Second, using WB analysis in ΔTat cells protein could be detected and was only observed in the spheroplast fraction without showing any export to the periplasm or media. While this implies that both SPs are Tat-specific under these expression conditions, we do not believe that either is completely Tat-specific. Rather we believe that while Tat is the normal/predominant secretion pathway for the AmiC and MdoD SPs, Sec-dependent secretion can occur. The trigger for promiscuity for SPs is not known, nor is the potential link between promiscuity and the protein being exported (or how efficiently it folds).

The choice of the POI for this study was not arbitrary. Apart from the desired CyDisCo dependency of the protein of choice, YebF is an intriguing POI. YebF with its native SP is used as a "passenger" protein linker to export transgenic proteins to the medium by an unknown mechanism. This discovery gives the possibility of linking more disulfide-bonded difficult-to-express proteins to these two efficient and Tat-dependent SPs for easy recovery in the extracellular medium, assuring the correct folding with CyDisCo of the target protein and a maximized yield when using TatExpress cells.

In conclusion, the AmiC and MdoD SPs appeared to allow efficient secretion of a disulfide bond containing protein from the cytoplasm of *E. coli via* Tat. While these signal sequences may not be completely Tat-specific, the data suggests the majority of the YebF is exported via Tat. As they are far more efficient than the TorA SP, they are probably more suitable for large-scale protein production than the reportedly more rigorously specific Tat-SP.

AUTHOR CONTRIBUTIONS

Klaudia Arauzo-Aguilera: Formal analysis (lead); investigation (lead); visualization (lead); writing—original draft (lead). **Mirva J. Saaranen**: Formal analysis (supporting); investigation (supporting); methodology (equal); supervision (equal); writing—review and editing (equal). **Colin Robinson**: Funding acquisition (equal); supervision (equal); writing—review and editing (equal). **Lloyd W. Ruddock**: Conceptualization (equal); funding acquisition (equal); methodology (equal); supervision (equal); writing—review and editing (equal).

ACKNOWLEDGMENTS

This work was funded by the People Programme (Marie Skłodowska-Curie Actions) of the European Union's Horizon 2020 Programme under REA grant agreement no. 813979 (SECRETERS). The use of the facilities of Biocenter Oulu core facilities, a member of Biocenter Finland, is gratefully acknowledged.

CONFLICT OF INTEREST STATEMENT

A patent for the production system used to make the protein using sulfhydryl oxidases in the cytoplasm of *E. coli* is held by the University of Oulu: Method for producing natively folded proteins in a prokaryotic host (Patent number 9238817; date of patent January 19, 2016). Inventor: Lloyd W. Ruddock.

DATA AVAILABILITY STATEMENT

All data supporting the conclusions of this study are included in the article.

ETHICS STATEMENT

None required.

ORCID

Colin Robinson http://orcid.org/0000-0002-1739-032X
Lloyd W. Ruddock http://orcid.org/0000-0002-6247-686X

REFERENCES

Alanen, H. I., Walker, K. L., Lourdes Velez Suberbie, M., Matos, C. F. R. O., Bönisch, S., Freedman, R. B., Keshavarz-Moore, E., Ruddock, L. W., & Robinson, C. (2015). Efficient export of human growth hormone, interferon α2b and antibody fragments to the periplasm by the *Escherichia coli* coli Tat pathway in the absence of prior disulfide bond formation. *Biochimica et Biophysica Acta (BBA)—Molecular Cell Research*, *1853*(3), 756–763. https://doi.org/10.1016/J.BBAMCR.2014.12.027

Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T., & Brunak, S. (2005). Prediction of twinarginine signal peptides. *BMC Bioinformatics*, *6*(1), 167. https://doi.org/10.1186/1471-2105-6-167

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X. Z., & Zhou, T. (2021). Challenges associated with the formation of recombinant protein inclusion bodies in *Escherichia coli* and strategies to address them for industrial applications. *Frontiers in Bioengineering and Biotechnology*, *9*, 630551. https://doi.org/10.3389/FBIOE.2021.630551

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. *Journal of Bacteriology*, *185*(9), 2811–2819. https://doi.org/10.1128/JB.185.9.2811-2819.2003

Blaudeck, N., Sprenger, G. A., Freudl, R., & Wiegert, T. (2001). Specificity of signal peptide recognition in Tat-dependent bacterial protein translocation. *Journal of Bacteriology*, *183*(2), 604–610. https://doi.org/10.1128/JB.183.2.604-610.2001

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. *Biotechnology and Bioengineering*, *114*(12), 2828–2836. https://doi.org/10.1002/bit.26434

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), 6115–6120. https://doi.org/10.1073/PNAS.0937838100

Fisher, A. C., Haitjema, C. H., Guarino, C., Çelik, E., Endicott, C. E., Reading, C. A., Merritt, J. H., Ptak, A. C., Zhang, S., & DeLisa, M. P. (2011). Production of secretory and extracellular N-Linked glycoproteins in *Escherichia coli*. *Applied and Environmental Microbiology*, *77*(3), 871–881. https://doi.org/10.1128/AEM.01901-10

Frain, K. M., Robinson, C., & van Dijl, J. M. (2019). Transport of folded proteins by the Tat System. *The Protein Journal*, *38*(4), 377–388. https://doi.org/10.1007/S10930-019-09859-Y.

Freudl, R. (2018). Signal peptides for recombinant protein secretion in bacterial expression systems. *Microbial Cell Factories*, *17*(1), 52. https://doi.org/10.1186/S12934-018-0901-3

Gaciarz, A., Veijola, J., Uchida, Y., Saaranen, M. J., Wang, C., Hörkkö, S., & Ruddock, L. W. (2016). Systematic screening of soluble expression of antibody fragments in the cytoplasm of *E. coli*. *Microbial Cell Factories*, *15*(1), 22. https://doi.org/10.1186/S12934-016-0419-5

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Ed.), *The proteomics protocols handbook*. Humana Press. https://doi.org/10.1385/1-59259-890-0:571

Grassi, L., & Cabrele, C. (2019). Susceptibility of protein therapeutics to spontaneous chemical modifications by oxidation, cyclization, and elimination reactions. *Amino Acids*, *51*, 1409–1431. https://doi.org/10.1007/s00726-019-02787-2

Guerrero Montero, I., Richards, K. L., Jawara, C., Browning, D. F., Peswani, A. R., Labrit, M., Allen, M., Aubry, C., Davé, E., Humphreys, D. P., Busby, S. J. W., & Robinson, C. (2019). *Escherichia coli* "TatExpress" strains export several g/L human growth hormone to the periplasm by the Tat pathway. *Biotechnology and Bioengineering*, *116*(12), 3282–3291. https://doi.org/10.1002/BIT.27147

Guerrero-Montero, I., Dolata, K. M., Schlüter, R., Malherbe, G., Sievers, S., Zühlke, D., Sura, T., Dave, E., Riedel, K., & 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. *Microbial Cell Factories* *18*(1), 19. https://doi.org/10.1186/s12934-019-1071-7

Haitjema, C. H., Boock, J. T., Natarajan, A., Dominguez, M. A., Gardner, J. G., Keating, D. H., Withers, S. T., & DeLisa, M. P. (2014). Universal genetic assay for engineering extracellular protein expression. *ACS Synthetic Biology*, *3*(2), 74–82. https://doi.org/10.1021/sb400142b

Harrison, J. J., Ceri, H., Badry, E. A., Roper, N. J., Tomlin, K. L., & Turner, R. J. (2005). Effects of the twin-arginine translocase on the structure and antimicrobial susceptibility of *Escherichia coli* biofilms. *Canadian Journal of Microbiology*, *51*(8), 671–683. https://doi.org/10.1139/W05-048

Jack, R. L., Buchanan, G., Dubini, A., Hatzixanthis, K., Palmer, T., & Sargent, F. (2004). Coordinating assembly and export of complex bacterial proteins. *The EMBO Journal*, *23*(20), 3962–3972. https://doi.org/10.1038/SJ.EMBOJ.7600409

Jong, W. S. P., Vikström, D., Houben, D., van den Berg van Saparoea, H. B., de Gier, J. W., & Luirink, J. (2017). Application of an *E. coli* signal sequence as a versatile inclusion body tag. *Microbial Cell Factories*, *16*(1), 50. https://doi.org/10.1186/S12934-017-0662-4

Karyolaimos, A., & de Gier, J. W. (2021). Strategies to enhance periplasmic recombinant protein production yields in *Escherichia coli*. *Frontiers in Bioengineering and Biotechnology*, *9*, 797334. https://doi.org/10.3389/FBIOE.2021.797334

Kleiner-Grote, G. R. M., Risse, J. M., & Friehs, K. (2018). Secretion of recombinant proteins from *E. coli*. *Engineering in Life Sciences*, *18*(8), 532–550. https://doi.org/10.1002/ELSC.201700200

Lee, P. A., Tullman-Ercek, D., & Georgiou, G. (2006). The bacterial twin-arginine translocation pathway. *Annual Review of Microbiology*, *60*, 373–395. https://doi.org/10.1146/ANNUREV.MICRO.60.080805.142212

Malherbe, G., Humphreys, D. P., & Davé, E. (2019). A robust fractionation method for protein subcellular localization studies in *Escherichia coli*. *Biotechniques*, *66*(4), 171–178. https://doi.org/10.2144/BTN-2018-0135

Matos, C. F. R. O., Robinson, C., Alanen, H. I., Prus, P., Uchida, Y., Ruddock, L. W., Freedman, R. B., & Keshavarz-Moore, E. (2014). Efficient export of prefolded, disulfide-bonded recombinant proteins to the periplasm by the Tat pathway in *Escherichia coli* CyDisCo strains. *Biotechnology Progress*, *30*(2), 281–290. https://doi.org/10.1002/btpr.1858

Mirzadeh, K., Shilling, P. J., Elfageih, R., Cumming, A. J., Cui, H. L., Rennig, M., Nørholm, M. H. H., & Daley, D. O. (2020). Increased production of periplasmic proteins in *Escherichia coli* by directed evolution of the translation initiation region. *Microbial Cell Factories*, *19*(1), 85. https://doi.org/10.1186/S12934-020-01339-8

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), 1735–1756. https://doi.org/10.1016/J.BBAMEM.2007.07.015

Nguyen, V. D., Saaranen, M. J., Karala, A. R., Lappi, A. K., Wang, L., Raykhel, I. B., Alanen, H. I., Salo, K. E. H., Wang, C., & Ruddock, L. W. (2011). Two endoplasmic reticulum PDI peroxidases increase the efficiency of the use of peroxide during disulfide bond formation. *Journal of Molecular Biology*, *406*(3), 503–515. https://doi.org/10.1016/J.JMB.2010.12.039

Saaranen, M. J., Karala, A. R., Lappi, A. K., & Ruddock, L. W. (2010). The role of dehydroascorbate in disulfide bond formation. *Antioxidants & Redox Signaling*, *12*(1), 15–25. https://doi.org/10.1089/ARS.2009.2674

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

Tullman-Ercek, D., DeLisa, M. P., Kawarasaki, Y., Iranpour, P., Ribnicky, B., Palmer, T., & Georgiou, G. (2007). Export pathway selectivity of *Escherichia coli* twin arginine translocation signal peptides. *Journal of Biological Chemistry*, *282*(11), 8309–8316. https://doi.org/10.1074/JBC.M610507200

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. *Journal of Biological Chemistry*, *275*(22), 16717–16722. https://doi.org/10.1074/JBC.M000800200

Zhang, G., Brokx, S., & Weiner, J. H. (2006). Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in *Escherichia coli*. *Nature Biotechnology*, *24*(1), 100–104. https://doi.org/10.1038/NBT1174

How to cite this article: Arauzo-Aguilera, K., Saaranen, M. J., Robinson, C., & Ruddock, L. W. (2023). Highly efficient export of a disulfide-bonded protein to the periplasm and medium by the Tat pathway using CyDisCo in *Escherichia coli*. *MicrobiologyOpen*, *12*, e1350. https://doi.org/10.1002/mbo3.1350