Exploitation of the twin-arginine translocation pathway and promoter control for biopharmaceutical production in *E. coli* strains

Sarah Lena Bischoff

A thesis submitted for the degree of Doctor of Philosophy

University of Kent
Department of Biosciences
2020

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.

This research was co-funded by the Graduate Teaching Scheme and BBSRC.

Furthermore, I would like to acknowledge Dr. Douglas Browning from the Steve Busby lab at the University of Birmingham for providing the strains and promoter for the p*narG*-CC promoter work and Dr. Tobias von der Haar for providing the PETase containing plasmid.

Abstract

The industrial production of proteins in bacteria has found its limits due to the size and requirements of post-translational folding of many proteins. However, bacteria are still a preferred expression host and various tools are applied to broaden the expression spectrum of these microorganisms. We investigated methods to decrease costs and improve protein expression in *Escherichia coli* and applied tools such as genetic engineering, change of promotor and signal peptides, and aiding disulphide bond formation. Here we examined the ability of *Escherichia coli* strain MC4100 encoding the TatABC protein secretion pathway of *Agrobacterium tumefaciens* to export complex proteins into the periplasm. We have shown that the mutant strain can export native Tat proteins but rejects non-native proteins of biopharmaceutical interest which are exported over Tat in *Escherichia coli* wildtype. Furthermore, we examined a sodium nitrate inducible promoter on its expression strength and periplasmic export, and we conducted research into the expression and export to the periplasm of the novel biodegrading enzyme PETase. Additionally, we show that the *narG*-CC promoter is capable of high-level protein expression of GFP but does not facilitate periplasmic expression of hGH and an scFv. Furthermore, we show that PETase can be exported into the periplasm by the Sec pathway but not the Tat pathway. The combined research shows that *Escherichia coli* is still a key player in the production of recombinant proteins and that proteins that require cytoplasmic post-translational folding can be exported under the correct circumstances, although finding the best tools to improve the expression can be a challenge.

Table of contents

DECLARATION II

ABSTRACT III

TABLE OF CONTENTS IV

LIST OF FIGURES VIII

LIST OF TABLES X

ACKNOWLEDGEMENTS XI

ABBREVIATIONS XIII

1. INTRODUCTION 1

1.1. Biotechnology Industry 1

1.2. Protein export in bacteria 6

1.3. Protein folding in bacteria 7

1.3.1. Trigger factor 8

1.3.2. The Hsp70 chaperone system of *E. coli* 8

1.3.3. Chaperonin system GroEl-GroES 9

1.3.4. Protein degradation 10

1.4. The general secretory pathway (Sec) 10
1.4.1. Sec dependent insertion of membrane proteins 11
1.4.2. Sec-specific signal peptides 14
1.4.3. Sec specific translocation 15
1.4.4. Disulphide bond formation in the periplasm 19

1.5. The Twin-Arginine Translocase (Tat) pathway 20
1.5.1. Tat membrane proteins 21
1.5.2. Tat specific signal peptides 24
1.5.3. Disulphide bond formation in the cytoplasm 27
1.5.4. Tat specific translocation in *E. coli* 28

1.6. *Agrobacterium tumefaciens* Tat translocation system 34

1.7. Influence of the promoter on protein expression 35

1.8. Aims and objectives 39

2. MATERIALS AND METHODS 41

2.1. Supplies of chemicals, reagents and equipment used 41

2.2. DNA techniques 44
2.2.1. Preparation of plasmid DNA 44
2.2.2. Preparation of genomic DNA 44
2.2.3. Amplification of DNA via Polymerase Chain Reaction 44
2.2.4. Purification of DNA from the genome 48
2.2.5. Agarose gel electrophoresis 48
2.2.6. Purification of DNA from agarose gels 48
2.2.7. Restriction Digests of DNA 49
2.2.8. Ligation of DNA fragments into plasmid vector backbone 50
2.2.9. Gibson assembly cloning 50
2.2.10. Sequencing of plasmid DNA 51
2.2.11. Sequencing of genomic DNA 51
2.2.12. Constructs used in this study 52
2.2.13. Gene doctoring 55

2.3. Maintenance of *E. coli* cultures 55
2.3.1. Media and supplements 55
2.3.3. Storage of *E. coli* cells 59

2.4. Protein production and *E. coli* cell fractionation 59
2.4.1. Cell culture and induction of plasmids 59
2.4.2. *E. coli* fermentation 60
2.4.3. E. *coli* cell fractionation 62
2.4.4. Whole cell extract 63

2.5. Protein electrophoresis 64
2.5.1. SDS poly-acrylamide gel electrophoresis (SDS-PAGE) 64
2.5.2. Native gel/ CN-PAGE gel 64
2.5.3. Coomassie staining 65

2.6. Protein detection 65
2.6.1. Western blotting 65
2.6.2. Immunoblotting 66
2.6.3. TMAO assay 68

2.7. Protein purification 68
2.7.1. Ni-NTA Spin column purification of 6xhis tagged proteins under native conditions from *E. coli* cell lysate 68

2.8. Nitrite/ Nitrate assay, colorimetric 69
2.8.1. Nitrite/ Nitrate assay kit (Sigma Aldrich) 69

2.9. Microscopy 70
2.9.1. Leica DMR microscope 70

3. *E. COLI* GENOME MODIFICATION WITH *TATABC AGROBACTERIUM TUMEFACIENS* OPERON AND EXPORT OF TARGET PROTEINS 71

3.1. Introduction 72

3.2. Results 75
3.2.1. Expression of Tat proteins on a secondary plasmid in Δ*tat* is able to rescue the filamentous phenotype but does not improve the growth 75

3.2.2. *A. tumefaciens* TatABC is successfully inserted on the chromosome and can improve the growth 79
3.2.3. MC4100 expressing the *Agrobacterium tumefaciens tatABC* operon is able to export the native Tat substrate TorA into the periplasm 87
3.2.4. MC4100 clones expressing the *A. tumefaciens* TatABC is not capable of exporting human growth hormone into the periplasm 90
3.2.5. MC4100 *A. tumefaciens* TatABC is not capable of exporting single chain variable fragments into the periplasm 93
3.2.6. MC4100 *A. tumefaciens* TatABC is not capable of exporting GFP into the periplasm 97
3.2.7. MC4100 *A. tumefaciens* TatABC is not capable of exporting co-translationally folded proteins 101

3.3. Discussion 105

4. OPTIMISING PROTEIN EXPRESSION USING A NITRATE INDUCIBLE PROMOTER 108

4.1. Introduction 109

4.2. Results 112
4.2.1. p*narG*-CC can be used for single induction 112
4.2.2. p*narG*-CC shows tightly controlled expression in MS and SM6 media 113
4.2.3. Expression of GFP in fed-batch fermentation 118
4.2.4. p*narG*-CC is able to express comparable amount of GFP to p*tac* 124
4.2.5. p*narG*-CC can express TorA-SP-hGH in shake flask under specific conditions 129
4.2.6. TorA-hGH can be expressed in a fermenter 134
4.2.7. TorA-SP-scFv cannot be exported to the periplasm using p*narG*-CC 138

4.3. Discussion 140

5. PRODUCTION AND EXPORT OF PETASE BY THE TAT PATHWAY 143

5.1. Introduction 144

5.2. Results 147
5.2.1. PETase is not expressed at high levels using the TorA signal peptide 147
5.2.2. PETase is successfully expressed using the HyaA signal peptide 150
5.2.3. PETase could not be purified using a bench Ni-NTA spin column 159
5.2.4. PETase is being expressed using the AmiC signal peptide both with and without disulphide bond supporting components 162
5.2.5. PETase cannot be expressed using PhoD*N as a signal peptide in initial experiments 170
5.2.6. Further experiments that had been planned 173

5.3. Discussion 174

6. FINAL DISCUSSION 177

7. REFERENCES 183

8. APPENDIX 215

8.1. Alignment of TatABC *E. coli* to TatABC *A. tumefaciens* 215

8.2. Base pair and amino acid sequence of codon optimised TatABC *A. tumefaciens* operon used for pDoc-K 216

8.3. Base pair and amino acid sequence of Clone 1, containing the gene encoding the *tatABC* operon of *A. tumefaciens* 217

8.4. Base pair and amino acid sequence of Clone 2, containing the gene encoding the *tatABC* operon of *A. tumefaciens* 218

8.5. Alignment of Clone 1 insertion to the *A. tumefaciens tatABC* operon 219

8.6. Alignment of Clone 2 insertion to the *A. tumefaciens tatABC* operon 220

List of figures

FIGURE 1: CELL LINES USED FOR PHARMACEUTICAL PRODUCTION. 3
FIGURE 2: SCHEMATIC DIAGRAM OF THE TRANSMEMBRANE STRUCTURE OF SECY, SECE, AND SECG. 12
FIGURE 3: STRUCTURE OF THE SEC PATHWAY. 14
FIGURE 4: SCHEMATIC PRESENTATION OF THE SEC SIGNAL PEPTIDE MOTIF. 15
FIGURE 5: SEC-DEPENDENT TRANSLOCATION. 18
FIGURE 6: *E. COLI TATABCD*. 22
FIGURE 7: TOPOGRAPHY OF THE PROTEINS INVOLVED IN THE *E. COLI* TAT PATHWAY. 23
FIGURE 8: COMPARISON OF THE GENERAL MOTIF OF SIGNAL PEPTIDES TARGETING TO THE SEC AND THE TAT PATHWAY. 25
FIGURE 9: SIGNAL PEPTIDE INTERACTION WITH THE TATBC COMPLEX. 30
FIGURE 10: SIGNAL PEPTIDE CLEAVAGE. 31
FIGURE 11: THE FIGURE SHOWS A SCHEMATIC PRESENTATION OF THE *NARG-CC* (-40.5) PROMOTER. 39
FIGURE 12: SCHEMATIC PRESENTATIONS OF THE HERE USED FED-BATCH FERMENTER. 61
FIGURE 13: *A. TUMEFACIENS* TATABC COMPLEMENTS THE Δ*TAT* PHENOTYPE OF *E. COLI*. 77
FIGURE 14: COMPARISON OF THE GROWTH OF Δ*TAT* EXPRESSING A PLASMID CONTAINING THE *AGROBACTERIUM TUMEFACIENS TAT* OPERON TO MC4100 WILDTYPE AND Δ*TAT*. 78
FIGURE 15: PLASMID MAP SHOWING OF THE FEATURES OF THE GENE DOCTORING PLASMID PDOC-K INCLUDING THE *TATABC* OPERON OF *A. TUMEFACIENS*. 80
FIGURE 16: LOCATION OF THE HOMOLOGY REGIONS (HR1 AND HR2 ON THE GENOME). 81
FIGURE 17: FLOW DIAGRAM OF THE GENE DOCTORING PROCESS. 82
FIGURE 18: DNA INSERT CONFIRMATION OF THE MC4100 COLONIES. 83
FIGURE 19: COMPARISON OF THE DIFFERENT PHENOTYPES OF THE NEW CLONES, WILDTYPE CELLS, AND Δ*TAT*. 85
FIGURE 20: COMPARISON OF THE GROWTH OF THE TWO TAT CLONES TO MC4100 AND Δ*TAT*. 86
FIGURE 21: CLONES EXPRESSING THE *A. TUMEFACIENS TATABC* OPERON ARE CAPABLE OF EXPORTING A NATIVE TAT SUBSTRATE. 88
FIGURE 22: TATA SIGNAL IN MC4100, Δ*TAT*, AND THE NEW *A. TUMEFACIENS* TATABC. 89
FIGURE 23: THE DETECTION OF HGH IN CYTOPLASMIC, INSOLUBLE, AND PERIPLASMIC FRACTIONS FOLLOWING EXPRESSION WITH THE TAT SPECIFIC SIGNAL PEPTIDE TORA IN STRAINS ENCODING THE *E. COLI* AND *A. TUMEFACIENS TAT* OPERON. 91
FIGURE 24: TORA ACTIVITY DURING TORA-SP-HGH EXPRESSION. 92
FIGURE 25: EXPRESSION OF SCFV(COBRA) WITH THE TAT SPECIFIC SIGNAL PEPTIDE TORA IN THE NEW STRAINS. 94
FIGURE 26: EXPRESSION OF SINGLE CHAIN VARIABLE FRAGMENT (SCFV) VARIANT WITH THE TAT SPECIFIC SIGNAL PEPTIDE TORA IN THE NEW STRAINS. 96
FIGURE 27: EXPRESSION OF TORA-SP-GFP IN THE NEW STRAINS COMPARED TO MC4100 AND Δ*TAT*. 98
FIGURE 28: GFP IS EXPRESSED IN BL21 DE3. 100
FIGURE 29: EXPRESSION OF TORA-SP-R16 IN THE NEW STRAINS COMPARED TO MC4100 AND Δ*TAT*. 102
FIGURE 30: EXPRESSION OF TORA-SP-PROTEIN G IN THE NEW STRAINS COMPARED TO MC4100 AND Δ*TAT*. 103
FIGURE 31: THE FIGURE SHOWS A SCHEMATIC PRESENTATION OF THE *NARG-CC* (-40.5) PROMOTER. 110
FIGURE 32: NITRATE AND NITRITE CONCENTRATION IN THE MEDIUM OF A FED-BATCH FERMENTATION. 113
FIGURE 33: COMPARISON OF EXPRESSION OF GFP IN SM6 MEDIUM TO MS MEDIUM. 115

FIGURE 34: COMPARISON OF CONTROL OF EXPRESSION OF GFP FROM P*TAC* AND THE *NARG*-CC PROMOTERS. 117
FIGURE 35: GFP INCREASES THROUGHOUT THE FERMENTATION UNDER THE CONTROL OF THE *NARG*-CC PROMOTER. 119
FIGURE 36: FERMENTATION EXPRESSING P*NARG*-CC GFP. 120
FIGURE 37: COOMASSIE STAINED SDS-PAGE GEL SHOWING THE PRODUCTION OF GFP FROM *E. COLI* N11 EXPRESSING THE GFP (27 KDA, ARROWED) GENE FROM THE NITRATE-INDUCIBLE *NARG*-CC PROMOTER. 121
FIGURE 38: GROWTH CURVE OF THE FERMENTATION OF *E. COLI* N11 WITH 0.15 UL/MIN GLYCEROL FEED RATE. 122
FIGURE 39: COOMASSIE STAINED SDS-PAGE GEL AND WESTERN BLOT SHOWING THE PRODUCTION OF GFP FROM *E. COLI* N11 EXPRESSING THE GFP (27 KDA, ARROWED) GENE FROM THE NITRATE-INDUCIBLE *NARG*-CC PROMOTER. 123
FIGURE 40: COMPARISON OF GFP PRODUCTION FROM *E. COLI* N11USING THE *NARG*-CC PROMOTER AND THE *TAC* PROMOTER. 127
FIGURE 41: COMPARISON OF GFP CONCENTRATION EXPRESSED USING THE *NARG*-CC PROMOTER COMPARED TO THE *TAC* PROMOTER. 128
FIGURE 42: INDUCER CONCENTRATION HAS MINOR IMPACT ON TORA-SP-HGH EXPRESSION USING THE *NARG*-CC PROMOTER. 131
FIGURE 43: EXPRESSION OF TORA-SP-HGH IN THE TN3.1 STRAIN DID NOT RESULT IN EXPORT USING THE P*NARG*-CC PROMOTER. 133
FIGURE 44: TORA-SP-HGH EXPRESSED IN N11 THROUGHOUT FERMENTATION IN SM6 MEDIUM USING P*NARG*-CC. 135
FIGURE 45: TORA-SP-HGH EXPRESSED IN TN3.1 THROUGHOUT FERMENTATION IN SM6 MEDIUM USING THE *NARG*-CC PROMOTER. 137
FIGURE 46: EXPRESSION OF P*NARG*-CC TORA-SP-SCFV WITH VARYING INDUCER CONCENTRATIONS. 139
FIGURE 47: COMPARISON OF THE EXPRESSION OF TORA-SP-PETASE IN *E. COLI* STRAIN BL21 DE3 TE AT DIFFERENT TEMPERATURES AND DIFFERENT INCUBATION TIMES. 148
FIGURE 48: TORA-SP-PETASE DOES NOT EXPRESS AT 20˚C. 149
FIGURE 49: THE INFLUENCE OF TEMPERATURE ON THE PRODUCTION OF HYAA-SP-PETASE BY *E. COLI* STRAIN BL21 DE3 TE. 151
FIGURE 50: PETASE IS SYNTHESISED AT 20˚C IN LOW CONCENTRATIONS USING THE HYAA SIGNAL PEPTIDE. 153
FIGURE 51: EXPRESSION OF HYAA-SP-PETASE WITH CYDISCO COMPONENTS. 154
FIGURE 52: SYNTHESIS OF HYAA-SP-PETASE IN A MC4100 Δ*TAT* (*TATABCDE*) AND BL21. 156
FIGURE 53: COMPARISON OF EXPRESSION OF HYAA- SP-PETASE IN Δ*TAT* IN DIFFERENT MEDIA. 158
FIGURE 54: BENCH TOP PURIFICATION OF PETASE CONTAINING SAMPLES. 162
FIGURE 55: COMPARISON OF EXPRESSION OF AMIC-SP-PETASE IN BL21 DE3 TE AT DIFFERENT TEMPERATURES. 165
FIGURE 56: AMIC-SP-PETASE IS EXPRESSED AT LOWER TEMPERATURES. 166
FIGURE 57: EXPRESSION OF HYAA-SP-PETASE WITH CYDISCO COMPONENTS AND AMIC-SP-PETASE WITH CYDISCO COMPONENTS. 167
FIGURE 58: EXPRESSION IN Δ*TAT* TO DETERMINE WHETHER EXPRESSION IS TAT MEDIATED. 169
FIGURE 59: EXPRESSION OF PHOD*N-SP-PETASE. 171
FIGURE 60: EXPRESSION OF PHOD*N-SP-PETASE WITH CYDISCO COMPONENTS. 172
FIGURE 61: HERE SHOWN IS THE ALIGNMENT OF TATABC *E. COLI* (ECO) TO TATABC *A. TUMEFACIENS* (ATU). FOR ALIGNMENT WEB.EXPASY.ORG HAS BEEN USED. 215
FIGURE 62: HERE SHOWN IS THE ALIGNMENT OF THE *A. TUMEFACIENS TATABC* SEQUENCE TO THE INSERTED SEQUENCE IN CLONE 1. 219
FIGURE 63: HERE SHOWN IS THE ALIGNMENT OF THE *A. TUMEFACIENS TATABC* SEQUENCE TO THE INSERTED SEQUENCE IN CLONE 2. 220

List of tables

TABLE 1: REPRESENTATION OF THE TWIN-ARGININE MOTIF OF THE SIGNAL PEPTIDES USED IN THIS STUDY. 26
TABLE 2: PROMOTERS COMMONLY USED FOR RECOMBINANT PROTEIN EXPRESSION IN *E. COLI*. 37
TABLE 3: CLONING PRIMERS USED IN THIS WORK. 45
TABLE 4: RESTRICTION ENZYMES USED IN THIS WORK. 49
TABLE 5: GIBSON ASSEMBLY COMPOSITION. 50
TABLE 6: SEQUENCING PRIMERS USED IN THIS WORK. 51
TABLE 7: PRIMERS USED FOR SEQUENCING THE *E. COLI* GENOME. 52
TABLE 8: PLASMID CONSTRUCT USED IN THIS WORK. 52
TABLE 9: THE IN THIS STUDY USED MEDIA AND THEIR COMPONENTS. 56
TABLE 10: REPRESENTATION OF THE *E. COLI* STRAINS USED IN THIS WORK AND THEIR GENOTYPES. 57
TABLE 11: ANTIBODIES USED IN THIS WORK. 67
TABLE 12: BUFFERS REQUIRED FOR THE NI-NTA SPIN COLUMN PURIFICATION 69
TABLE 13: COMPARISON BETWEEN THE *A. TUMEFACIENS* TAT PROTEINS AND THE *E. COLI* TAT PROTEINS. 74

Acknowledgements

First of all, I would like to thank my supervisor **Prof. Colin Robinson**. Thank you for giving me this opportunity and for always having an open door for me. I am truly grateful you gave me this opportunity and guided me during my research. Especially, through the difficult times such as the lockdown.

I also want to thank the Robinson lab, all the current and old members, for the uncountable coffee breaks, memes and the good experiences throughout this journey. Especially, I want to take **Dr. Alison Walters** who has never got tired to give me feedback for my experiments and keeping my motivation up. Your guidance throughout the PhD has been more than just helpful and deserves a special thank you!

To **Dr. Isabel Guerrero Montero** (it is still exciting writing Dr. for you) and **Chillel Jawara** for simply being great and always being there for a laugh or to listen to me about empty western blots or the fermenters not working (AGAIN!).

Thank you to **Dr. Douglas Browning** for his countless advice and motivating words to the genomic engineering work and for allowing me to use his promoter in my experiments.

Furthermore, I would like to thank the **KFG** and **Jack**, **Dan**, and **Mireya** in particular. Thank you for feeding me coffee and biscuits whenever I needed it and for always making me laugh. You made my third year a lot of fun and I hope we keep our coffee meetings up.

Without my family and especially my parents, I would have never started my PhD. My parents were supportive the whole time and always took time to listen to me. Thank you, **Patricia and Jürgen**, for everything you have done and for always believing in me.

A special thank you goes to my partner. This work would not have been possible without **Adrian**. You are always understanding no matter in what mood I was in. Thank you for creating an environment where I could write my thesis and for taking me on all our weekend adventures. I still cannot believe how patient you have been throughout our quarantine time when you couldn’t escape my moods whilst I was writing my introduction.

Thank you!

Abbreviations

°C	Degrees Celsius
Å	Ångströms
A. tumefaciens	*Agrobacterium tumefaciens*
APS	Ammonium persulphate
BBSRC	Biotechnology and Biological Sciences Research Council
BSA	Bovine serum albumim
C	Cytoplasmic fraction
C-	Carboxy terminus
CHO	Chinese Hamster Ovary
CM	Cytoplasmic membrane
CyDisCo	Cytoplasmic disulphide bond formation in *E. coli*
dH_2O	Distilled water
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ErV1p	Sulfhydryl oxidase 1
FDA	U. S. Food and Drug Administration
g	Grams
GFP	Green fluorescent protein
hGH	Human growth hormone
His	Hexa-histidine tag
Hr(s)	Hour(s)
HSP70	Heat shock protein 70

I. sakaiensis	*Ideonella sakaiensis*
I	Insoluble fraction
IPTG	Isopropylthiogalactoside
kDa	Kilodaltons
L	Litre
LB	Lysogeny broth
LBA	Lysogeny broth agar
M	Molar
mg	Milligram
MHET	mono(2-hydroxyethyl) terephthalic acid
min	Minutes
mL	Millilitre
mM	Millimolar
N-	Amino terminus
NEB	New England Biolabs
nm	Nanometre
O/N	Overnight
OD	Optical density
P	Periplasmic fraction
PAGE	Polyacrylamide gel electrophoresis
PBS	Phoshate buffered saline
PBS-T	Phosphate buffered saline and tween
PCR	Polymerase Chain Reaction
PDI	Protein disulphide-isomerase
PET	Poly (ethylene) terephtalate

PMF	Proton-motive force
PVDF	Polyvinylidene fluoride membrane
Rpm	Revolutions per minute
RR	Twin-arginine motif
RT	Room temperature (20 °C)
s	Seconds
scFV	Single-chain variable-fragment from an antibody
SDS	Sodium dodecylsuphate
Sec	General secretory pathway
SEM	Scanning electron microscopy
SM6	Synthetic medium 6
SOD1	Superoxidase dismutase 1
sp	Signal peptide
SRP	Signal recognition particle
Strep-tag	Strep II tag
Tat	Twin-arginine translocation
TB	Terrific broth
TE	TatExpress
TEMED	Tetramethylethylenediamine
TM	transmembrane
Tm	Melting temperature
TMAO	Trimethylamin-N-oxide
TorA	Trimethylamine-N-oxide reductase
tPA	Tissue plasminogen activator
Tris	Tris (hydroxymethyl) aminomethane

Tween20	Polyoxyethylenesorbitan monolaurate
V	Volts
w/v	Weight per volume
WT	Wild type
Δ (Delta)	Gene deletion
µL	Microlitre
µM	Micromolar

1. Introduction

1.1. Biotechnology Industry

We understand the term biotechnology as "the use of living organism/s or their product/s to modify or improve human health and human environment". With this definition, we recognise that biotechnology is as old as human history and an example for early use of biotechnology is fermentation (Verma *et al.*, 2011). The oldest fermentation products that have been found date back to 7000 to 6000 BCE and it is a fermented drink made out of fruit, honey and rice found in China (McGovern *et al.*, 2004). It took many centuries from accidental biotechnology to modern/ directed biotechnology. The construction of bacterial plasmids and DNA cloning in 1973 has paved the path of genetic engineering and modern biotechnology (Cohen *et al.*, 1973).

The origins of biotechnology as we know it nowadays in relation to industry are strongly connected with the history and discovery of insulin. Human insulin is the first recombinant protein drug to have been licensed. The so called Humulin, a recombinant human insulin, was also the first drug produced using recombinant DNA. Humulin was approved by the US Food and Drug Administration (FDA) in 1982. Insulin is expressed from a small, double-strand DNA ring called a plasmid which is inserted in the bacterial cell. Insulin is composed of two peptide chains, insulin A and insulin B chain, which are linked together by two disulphide bonds. At the time, insulin was expressed in a two-plasmid system. The two peptide chains were purified and then disulphide bond formation was promoted to produce bioactive insulin (Johnson, 1983; Baeshen *et al.*, 2014). Insulin accumulates within the cell and after the growth process

the bacteria are broken open to purify the protein (Nilsson *et al.*, 1996; Baeshen *et al.*, 2014).

Before insulin was produced in *E. coli*, insulin was purified from pancreatic tissue of cows and pigs (Sanger, 1958; Vecchio *et al.*, 2018). At that time, the quality of insulin was variable and the yields were low (Lens and Evertzen, 1952). Additionally, challenges were provided by the transport of the pancreatic tissue which had to be kept frozen. Otherwise, the pancreatic tissue would no longer have been suitable for purification. Up until today, *E. coli* is used as host for the expression of insulin using expression plasmids and purification from the bacterial cells. In the early 90s yields of 100g dry cell weight per litre could be achieved which was a major success and is still applicable today (Yee and Blanch, 1992; Lee, 1996; Yamanè and Shimizu, 2005).

The landscape of products changed towards more complex molecules like antibodies with complex folding requirements in the mid-90s. *E. coli* was the preferred host for biopharmaceutical production because of its economic ease of culture and cost-effectiveness, but could not fulfil the folding requirements of the new substrates (Swartz, 2001). Nowadays, *E. coli* is still a key player for protein production with about 24% of the approved therapeutics being produced in this host organism (Baeshen *et al.*, 2014; Mohamed N. Baeshen, 2014). Examples for those therapeutics are various interferons and cancer drugs such as Endostatin, Tasonermin, and Filgrastim (Sanchez-Garcia *et al.*, 2016). However, mammalian cell based protein production dominates the market, see Figure 1, as mammalian cells are able to express complex proteins with several disulphide bonds and sizes over 150 kDa, such as antibodies (Ranade, 2010). In 2010 about 40% of approved therapeutics were glycoproteins (proteins requiring glycosylation to be functional) (Lingg *et al.*, 2012).

The oligosaccharide structures differ depending on the host machineries of glycosyltransferases and the culture conditions. Therefore, the same amino acid sequence can have different glycoforms (Higgins, 2010). The glycan profile of proteins expressed in many mammalian cells is similar to that of the native human proteins which makes mammalian cells a preferred system.

In comparison, in 2011 30% of biopharmaceuticals, typically defined as a drug produced from a biological source, were produced in *E. coli* and only 39% in mammalian cells (Waegeman and Soetaert, 2011).

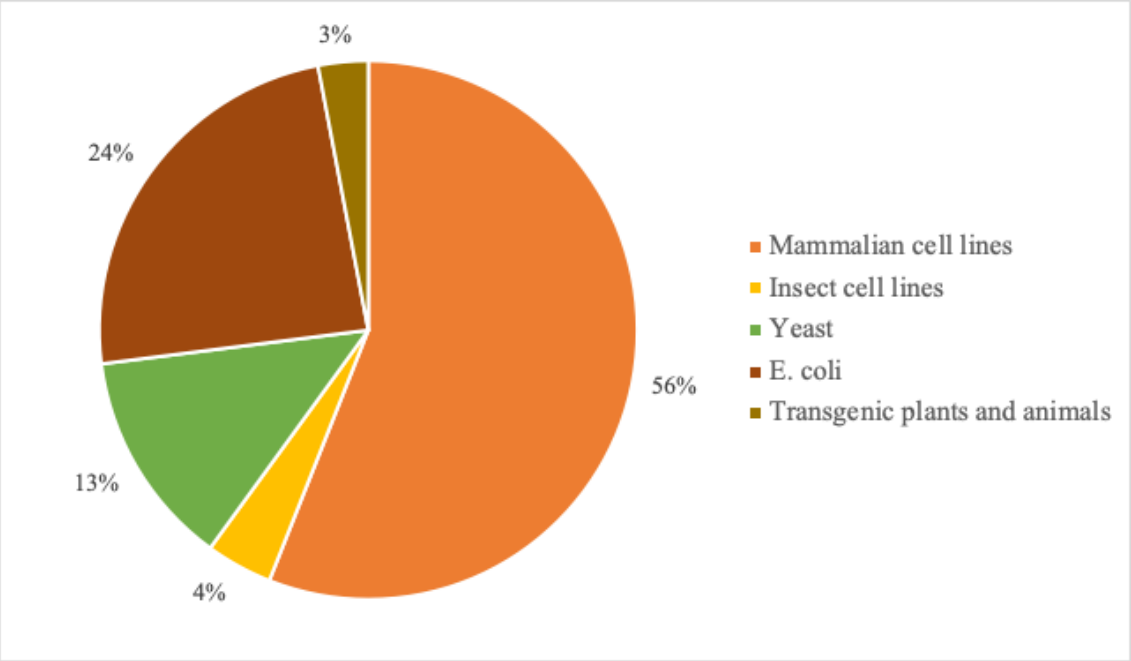


Figure 1: **Cell lines used for pharmaceutical production.** Here presented are the estimated percentages of cell lines used for biopharmaceutical production, based on numbers from 2014. Mammalian cell lines are dominating the market with about 56% of biopharmaceuticals produced in mammalian cell lines, most commonly used are Chinese Hamster Ovary (CHO) cells (Baeshen *et al.*, 2014).

Compared to mammalian host systems *E. coli* is a desirable expression system because of its rapid growth rate, capability of high-level expression, and cost-effectiveness (Waegeman and Soetaert, 2011). An initial challenge to overcome when using mammalian cells was that the cells prefer to be adhered to a surface while *E. coli* can grow in suspension. However, nowadays most processes use cell lines growing in suspension and only certain mammalian some used cell lines are anchorage-dependent (Flickinger *et al.*, 2010; Bielser *et al.*, 2018). Additionally, *E.coli* can be grown at various different temperatures and is, compared to mammalian cell cultures, flexible towards temperature and pH changes within the culture (Werner and Noé, 1993; Bielser *et al.*, 2018). Therefore, the fermenter systems used have less technical requirements, are cheaper, and allow a higher cell density. The cost of recombinant protein production in *E. coli* can be further decreased with targeting the desired protein into the growth medium or the periplasm to simplify purification processes and therefore reduce downstream costs (Quax, 1997; Burdette *et al.*, 2018).

Many systems have been developed to perform genetic manipulations to optimize *E. coli* further for higher yields and optimized protein expression (Lee *et al.*, 2009; Thomason *et al.*, 2014) and our understanding of protein folding and folding factors has improved (Swartz, 2001). *E. coli* has limitations for protein production such as limits in post-translational folding. The folding pathway in *E. coli* will be described at a later stage. These limits are a major factor driving in the expansion of mammalian cell production systems. Protein yields can be low even for protein expression optimized bacterial strains, plasmids and promoters (short DNA sequence that regulates expression of a plasmid) but with genetic manipulations the so far known limits can be changed. Scientists are continuously on the search for new tools for the biotech industry to overcome current limitations and optimise protein production

processes for new targets. In industry most proteins are produced in a fermentation process. Fermentation has been defined as ‘a chemical process by which molecules are broken down anaerobically by yeast or bacteria’ (Fuchs and Schlegel, 2006). Fermentation can increase the yields, optimise conditions and in some cases can reach high-cell-density cultures (Luli and Strohl, 1990). One form of fermentation is fed-batch fermentation where essential substrates are supplied to prevent starvation of cells or overfeeding throughout the fermentation process (Money, 2016). Some substrates such as 2,3-butanediol are not suited for a batch process due to an initial high substrate concentration which leads to a substrate inhibition and low culture yields (Sabra *et al.*, 2011). Increasing substrate availability has the benefit, that productivity of the cells can be maintained over a longer period of time and substrate inhibition can be prevented (Sabra *et al.*, 2011; Srivastava and Gupta, 2011).

To improve protein production for biotechnological purposes one can choose several strategies. Aspects that are commonly changed to improve protein expression are the plasmids and the hosts (Waegeman and Soetaert, 2011). Changes on the plasmid can be the signal peptide targeting the protein to a specific pathway or the promoter used in the expression cassette. Another method can be a genetic engineering of the host strain with changing aspects on the genome (Makrides, 1996). To develop a further understanding of recombinant protein production and the pathways that can be genetically modified, it is important to know the two main protein export pathways in *E. coli*.

1.2. Protein export in bacteria

An essential function of prokaryotic cells is the export of proteins through the cytoplasmic membrane as all major polypeptides are synthesised in the cytoplasm (Clark and Pazdernik, 2013). The translocation has to maintain the structural integrity or function of the cytoplasmic membrane (CM) whilst exporting the protein. This can happen via several pathways. There have been in total sixteen different pathways found in bacteria (Economou *et al.*, 2006). Some of those pathways only occur in either Gram-negative or Gram-positive prokaryotes (Economou *et al.*, 2006; Green and Mecsas, 2015). The most common pathways for the transport of soluble proteins, both found in Gram-negative and Gram-positive bacteria, are the general secretory pathway (Sec) and the Twin-Arginine Translocase (Tat) pathway (Green and Mecsas, 2015). A third pathway, the co-translational Signal Recognition Particle (SRP)-mediated pathway is more commonly used for the insertion of membrane proteins, although SRP is also used for the export of some soluble proteins. The most important difference between the Sec and the Tat those two pathways to point out is that the Sec pathway translocates unfolded proteins while and the Tat pathway translocates proteins that have been folded in the cytoplasm (Green and Mecsas, 2015).

Both pathways have been identified with minor variations in archaea, bacteria, and eukarya (Robinson and Bolhuis, 2004; Papanikou, Karamanou and Economou, 2007). Tat is found in the thylakoids of plants and green algae and in the plasma membranes of bacteria and archaea. The Sec pathway in *E. coli* exports the majority of proteins (Natale, Brüser and Driessen, 2008) and the Tat pathway exports about 30 native proteins, some of which are crucial for cell division and cell vitality (Palmer, Sargent

and Berks, 2010).The current work focusses on the exploitation of the Tat pathway for the secretion of fully folded proteins.

1.3. Protein folding in bacteria

The majority of proteins need to be folded into a three-dimensional structure which is an error prone process because it relies on the interaction of amino acids far apart in the sequence (Balchin, Hayer-Hartl and Hartl, 2016). Furthermore, misfolded proteins can accumulate and can, because of their exposed hydrophobic amino-acid residues, form toxic aggregates (Kim *et al.*, 2013). To avoid the accumulation of misfolded proteins, cellular pathways have evolved to assist in protein folding and refolding, and to degrade misfolded proteins that are refractile to refolding (Hartl, 2017). The network that maintains proteins in correct concentration is called protein homeostasis or proteostasis network (Powers, Powers and Gierasch, 2012) Chaperones (proteins that interact with other proteins to stabilise or aid to acquire its functionally active conformation) are responsible for maintaining proteostasis (Hartl, Bracher and Hayer-Hartl, 2011). A variety of chaperones have been identified such as foldase chaperones (mediating folding), unfoldases (correcting folding mistakes), Sec-avoidance chaperones (preventing targeting to the Sec pathway), proofreading chaperones (suppressing further transport till folding is finished), and chaperones that protect the substrate from cytoplasmic proteases (Turner, Papish and Sargent, 2004).

1.3.1. Trigger factor

Several chaperones are involved in maintaining proteostasis. One of these chaperones is trigger factor (TF) which binds to the ribosome exit tunnel and initiates folding of newly synthesized proteins (Merz *et al.*, 2008). TF interacts with the unfolded protein which is passing through the interior of TF and then assists the folding of the nascent chain. The contact to TF is lost when the nascent chain changes its conformation due to folding and the hydrophobic regions of the nascent chain are buried within the nascent polypeptide(Merz *et al.*, 2008; Hartl, Bracher and Hayer-Hartl, 2011). The release from TF is ATP independent and transfers proteins to the downstream chaperones. Such a downstream chaperone is Hsp70 (Morán Luengo, Mayer and Rüdiger, 2019).

1.3.2. The Hsp70 chaperone system of *E. coli*

The heat shock protein 70 (Hsp70) has been found in many organisms, from bacteria to human. Genome sequencing has shown that many organisms encode multiple members of the Hsp70 family. In *E. coli* three members of the Hsp70 family have been found (Genevaux, Georgopoulos and Kelley, 2007; Powers, Powers and Gierasch, 2012). One of the Hsp70 chaperones in *E. coli* is DnaK which is involved in the DnaK-DnaJ-GrpE cascade. The DnaK machinery is usually referred to as KJE and is the best studied Hsp70 system in *E. coli*. This chaperone machine consists of DnaK, the cochaperone DnaJ (Hsp40), and the nucleotide exchange factor GrpE. (Genevaux, Georgopoulos and Kelley, 2007; Imamoglu *et al.*, 2020).

The role of the KJE system is to bind misfolded proteins and unfold them (Powers, Powers and Gierasch, 2012). DnaK can be in in a high and low affinity state and is in an open state when ATP binds to the nucleotide-binding domain (Powers, Powers and Gierasch, 2012; Imamoglu *et al.*, 2020). A substrate can bind directly to DnaK or it can bind to the cochaperone DnaJ which is transferring the substrate to DnaK. When ATP is hydrolysed, DnaK undergoes a conformational change, causing the bound substrate to change conformation (Powers, Powers and Gierasch, 2012). However, the process how DnaK folds the substrate is not yet understood (Imamoglu *et al.*, 2020).

1.3.3. Chaperonin system GroEl-GroES

Another downstream chaperone system is the chaperonin GroEL-GroEs in the bacterial cytosol (Kim *et al.*, 2013). Chaperonins enclose the substrate for folding which has the benefit that this protects the substrate of aggregation or rebinding to upstream chaperones (Hartl, Bracher and Hayer-Hartl, 2011). The substrate binds to GroEL which has hydrophobic amino acid residues facilitating the binding. In the next step, ATP binds to GroEL triggering a conformational change which allows GroES to bind to the substrate-GroEL complex. The binding of GroES leads to the enclosure of the substrate in the chaperonin cavity (Clare *et al.*, 2012). Proteins up to 60 kDa can be encapsulated in this process (Tang *et al.*, 2006). The enclosure takes approximately 10 seconds, then GroES dissociates from the complex, by ATP binding, and the folded protein leaves GroEL. If the protein is incompletely folded it gets recaptures for another folding attempt (Tang *et al.*, 2006; Clare *et al.*, 2012; Kim *et al.*, 2013).

1.3.4. Protein degradation

ATP dependent proteases (AAA+) are responsible for degradation of proteins. Five different types of AAA+ proteases are found in *E. coli*. These are ClpXP, ClpAP, HslUV, Lon and FtsH (Sauer and Baker, 2011; Nyquist and Martin, 2014). These proteases have their oligomeric barrel like complex in common. The complex contains subunits with two functional domains, the ATPase and the proteoase domain (Bittner, Arends and Narberhaus, 2017). When a substrate binds to the AAA+ protease, the substrate is unfolded, translocated, and finally degraded. Binding can occur in a direct manner or indirect mediated by adaptor proteins. The next step is ATP-dependent. The ATPase domain of the AAA+ protease unfolds the substrate and translocate the substrate into the proteolytic chamber of the protease domain. The final step, the protease domain degrades the substrate in approximately 5 to 25 amino acid long peptide fragments (Gur, Biran and Ron, 2011; Gur, Ottofueling and Dougan, 2013).

1.4. The general secretory pathway (Sec)

The Sec pathway is the best understood pathway in bacteria and all the genes involved have been identified (Driessen and Nouwen, 2008). The majority of exported proteins in *E. coli* are translocated by the Sec pathway (Pugsley, 1993; Papanikou, Karamanou and Economou, 2007). Proteins are transported in an unfolded state across the CM and are called pre-proteins because of the removable amino (N)- terminal signal peptide which is cleaved off after translocation. The Sec pathway can be found in different domains of life and some of its components, such as SecY, are also present like the *secY* gene can be found in archaea and eubacteria. The Sec61 complex, built by Sec61α, Sec61β, and Sec61γ, is found in eukaryotes (Cao and Saier, 2003). The Sec

pathway is essential for cell viability in all organisms(Berg *et al.*, 2004; Du Plessis, Nouwen and Driessen, 2011).

1.4.1. Sec dependent insertion of membrane proteins

The Sec translocon in most bacteria is built up by eight different proteins embedded into the cytoplasmic membrane (Nouwen *et al.*, 2005). Those proteins can be classified as essential for translocation or essential for the efficiency of translocation, and the proteins include integral membrane proteins, protein export chaperone (SecB) and ATPase translocation motor (SecA) (Pogliano and Beckwith, 1994; Sachelaru *et al.*, 2017). Three of the integral membrane proteins that are essential for translocation are the protein translocase subunit SecY, the protein translocase subunit SecE, and the protein-export membrane unit SecG. Those three proteins together form SecYEG complex (Brundage *et al.*, 1990). The largest component of the SecYEG complex is SecY with 48 kDa and 10 transmembrane α-helical segments (TMS), followed by SecE with 14 kDa and 3 TMS, and SecG the smallest membrane protein of the complex has a mass of 12 kDa and 2 TMSs as schematically presented in Figure 2 (Cao and Saier, 2003; Du Plessis, Nouwen and Driessen, 2011).

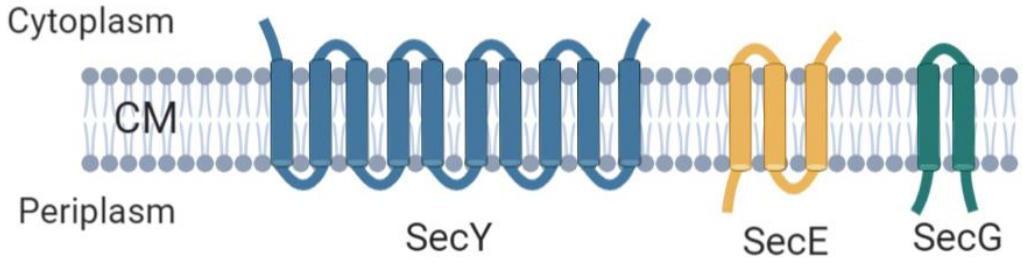


Figure 2: **Schematic diagram of the transmembrane structure of SecY, SecE, and SecG**. SecY is the largest component with 10 transmembrane segments. SecE has three transmembrane segments, and SecG is the smallest protein unit with two TMSs.

Additional components of the translocon are the protein translocase subunit SecA and the membrane protein insertase YidC. YidC is 61 kDa and SecA is even larger with 102 kDa. SecA, which has not been found in archaea, is functions as an ATPase and functions as the Sec pathway translocase motor (Oliver, 1993; Müller *et al.*, 2001).YidC, SecA, and the SecYEG complex are understood to be the main components for the functionality of the Sec pathway (Cao and Saier, 2003). SecA has the ability to interact with all components of the translocation system and is involved in the initial binding as well as translocation across the structure (Oliver, 1993) and YidC is involved in membrane biogenesis and further assists the transfer of proteins by interacting with the surface of the lateral gate of SecY (Sachelaru *et al.*, 2013, 2017).

The essential components of the *E. coli* Sec pathway are presented in Figure 3. Also shown in Figure 3 is the SecDF (yajC) complex built by the protein translocase subunit

SecD, the protein translocase subunit SecF, and the Sec translocon accessory complex subunit yajC. SecD and SecF have large periplasmic domains that make up about one third of their size (Pogliano and Beckwith, 1994; Nouwen *et al.*, 2005). The SecDF (yajC) complex does not occur in all prokaryotes and has not been found in eukaryotes. Furthermore, the complex is not essential for the functionality of Sec related export (Tseng *et al.*, 1999). However, SecDF is essential for efficient export by the Sec pathway. In *E. coli*, the deletion of *secD* and *secF*, and therefore the SecDF (yajC) complex, has resulted in a 10-fold decrease of exported protein compared to wildtype (WT) cells (Pogliano and Beckwith, 1994). The detailed function of the SecDF(yajC) complex remains unclear, however, it has been proposed that SecDF(yajC) is regulating the activity of SecA and SecG (Tsukazaki, 2018). Another working model suggests that SecDF interacts with an unfolded protein and is using the PMF to drive the forward movement of the protein and prevent backward movements (Tsukazaki and Nureki, 2011).

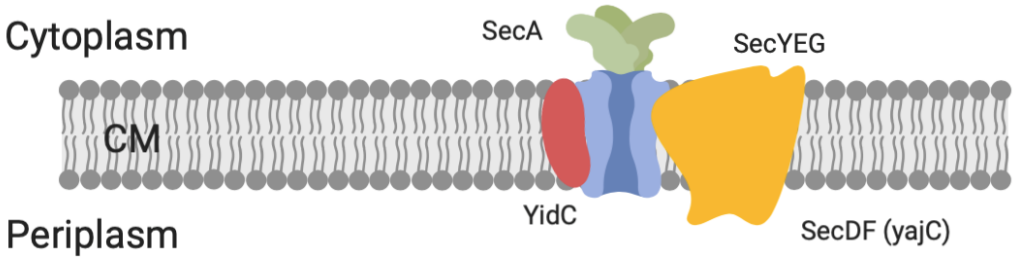


Figure 3: **Structure of the Sec pathway.** The Sec pathway is located in the cytoplasmic membrane (CM). The major parts are the protein-conducting channel SecYEG (blue), a cytoplasmic based motor domain SecA (green), and the two heterotrimeric complexes SecDF (orange) and YidC (red). SecYEG and YidC support pre-protein insertion. The pre-protein carries a N-terminal signal sequence which targets them to the Sec translocase.

The export of proteins to the periplasm requires signal peptides at the N-terminus of the protein to be successfully translocated by the Sec pathway and these are subsequently removed to generate the mature protein (Makrides, 1996).

1.4.2. Sec-specific signal peptides

The so far known signal peptides targeting proteins to the Sec pathway do not share sequence similarities. However, they share similarities in their overall structure. They have a positively charged n-region, a hydrophobic h-region of approximately ten amino acids, and the c-region contains a consensus cleavage motif of A-X-A as shown in Figure 4 (Dalbey and Kuhn, 2000; Rusch and Kendall, 2007). The Sec-specific signal sequences are typically about 20 amino acids long (Papanikou, Karamanou and Economou, 2007; Green and Mecsas, 2015).

Figure 4: **Schematic presentation of the Sec signal peptide motif.** Identified Sec specific signal peptides have a positively charged n-region, a hydrophobic h-region and share the A-X-A motif in the c-region.

Proteins with well characterised Sec signal peptides are for example OmpA (Movva, Nakamura and Inouye, 1980) and PhoA (Kikuchi *et al.*, 1981; Inouye *et al.*, 1982; Sjöström *et al.*, 1987). OmpA is an outer membrane protein with an N-terminal signal peptide which has been known since 1978 when its sequence was first determined (Endermann, Hindenach and Henning, 1978; Movva, Nakamuran and Inouye, 1980). The signal peptide together with the mature protein build the precursor protein which are recognised either by the signal recognition particle (SRP) or the SecB chaperone. Some proteins, such as OmpA, are SecB dependent whereas others like integral membrane proteins are targeted by SRP (Lee and Bernstein, 2001).

1.4.3. Sec specific translocation

The Sec translocon is located in the cytoplasmic membrane and the process can be divided into three steps. In the first step, pre-proteins are targeted to the Sec pathway

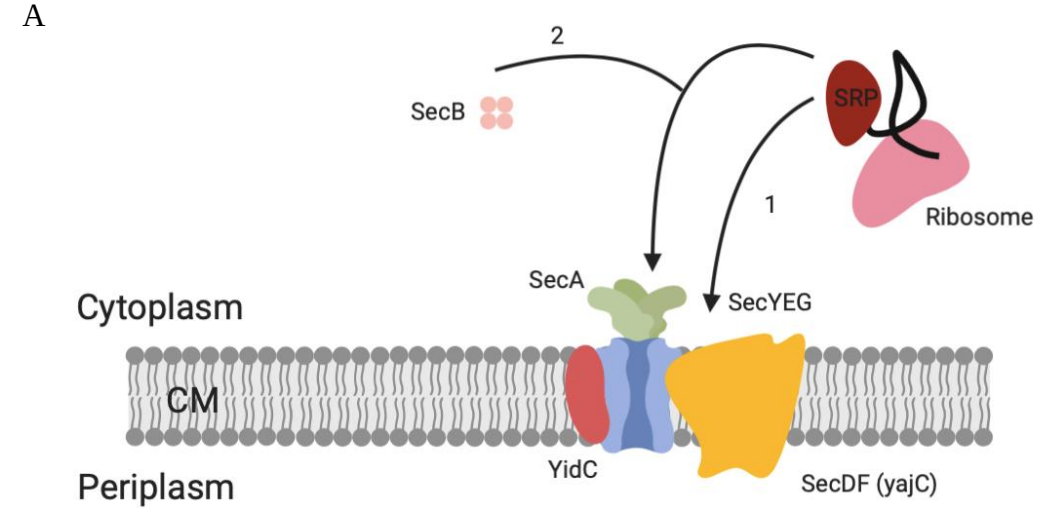
and are recognised either by the SRP (Figure 5A, targeting route 1) or the SecB chaperone (Figure 5A, targeting route 2). Sometimes the two different recognition ways are differentiated into two different Sec-related pathways referred to as the SRP pathway and the Sec pathway (Luirink and Sinning, 2004; Lee *et al.*, 2016).

SecY, SecE and SecG build the protein-conducting channel (SecYEG) and has a cytoplasmic based motor domain (SecA). The SecYEG complex is a narrow protein-conducting channel with a size of 5 to 8 Å and contains a SecA specific binding domain (Berg *et al.*, 2004). When the pre-protein is targeted to the membrane by the SecB chaperone, SecB binds to SecA. SecYEG can either interact with SRP for co-translation of mostly highly hydrophobic, integral membranes for lipid-phase integration or it interacts with the SecB/ SecA/ preprotein complex over the SecA binding domain for complete translation of less hydrophobic proteins as shown in Figure 5A (Randall and Hardy, 2002; Holland, 2004; Ito and Mori, 2009; Sachelaru *et al.*, 2017). SecA is known to drive translocation using ATP for its conformation changes during translocation and different models have been proposed for the translocation process itself (Cranford-Smith and Huber, 2018; Komarudin and Driessen, 2019).

SecYEG is also flanked by SecDF and YidC. SecDF and YidC support protein insertion and YidC interacts with the lateral gate of SecY in relation to the substrate availability. For SecA to work as a motor domain ATP binding is required and whilst undergoing conformational transition, SecA interacts with SecY (Kostakioti *et al.*, 2005). The third step of the Sec translocation is the release and maturation of the pre-protein (see Figure 5B). Throughout the process the hydrophilic, positively charged N-terminal region remains membrane located. The pre-protein is converted into a mature protein since the signal peptide, having targeted the pre-protein to the Sec

pathway, is cleaved by the signal peptidase. In a final step, the protein is folded into its native configuration on the trans side of the membrane, sometimes with the aid of folding factors (Paetzel *et al.*, 2002).

A



B

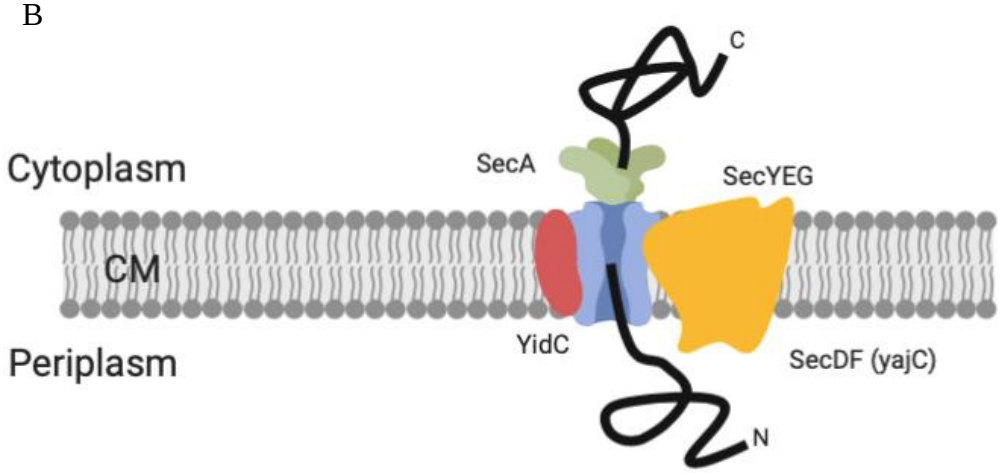


Figure 5: **Sec-dependent translocation. A** shows the two different ways pre-proteins (black line) can be targeted to the Sec pathway by the signal-recognition particle (SRP) (dark red) or the SecB chaperone (light red). The pathway in the image labelled with 1, is mediated by SRP and is used for co-translation. The pathway labelled with 2 is SecB mediated targeting used for co- and post-translational insertion of membrane proteins. This targeting is driven by an N-terminal signal sequence. In **B**, the protein gets pushed through the membrane driven by the SecA ATPase which is step 2 in the secretory process. Not shown here is step 3 where the signal peptide gets cleaved off the pre-protein and the protein is folded.

1.4.4. Disulphide bond formation in the periplasm

There are two systems that are responsible for periplasmic protein folding (Bardwell, McGovern and Beckwith, 1991; Miot and Betton, 2004). The first catalysts are the Dsb proteins, which form disulphide bonds and the second are peptidyl-prolyl isomerases (PPIase).

To create a disulphide bond with Dsb proteins, DsbA donates its own disulphide bond. In the first step, DsbA reacts with the target protein and creates an unstable mixed disulphide. In a second step, the mixed disulphide is targeted by another thiol group of the target protein and replaces the mixed the disulphide (Kadokura, Katzen and Beckwith, 2003; Nakamoto and Bardwell, 2004).

PPIases assists protein folding by catalysing a *cis-trans* isomerization of proline residue. The four identified PPIases are SurA, FkpA PpiA and PpiD (Mogensen and Otzen, 2005). Whilst it is understood which family of PPiases they belong to, SurA, FkpA, PpiA and PpiD do not appear to be essential for bacteria and their cellular role has not been entirely understood (Miot and Betton, 2004).

Furthermore, quality control proteases, such as the Deg protease family, are involved in the quality control process. Deg (in plant or bacteria) or HtrA (in mammalian cells) proteases are found in almost every organism (Kieselbach and Funk, 2003). In *E. coli* DegP is involved in degrading irreversibly damaged proteins and as chaperone assisting in the refolding of denatured or misfolded proteins (Schuhmann and Adamska, 2012).

An interesting example for successful expression with periplasmic folding in *E. coli* is the expression of tissue plasminogen activator (tPA) which requires 17 disulphide

bonds for its active state. The research group, overexpressed DsbA and DsbC to promote periplasmic disulphide bond formation. Through the overexpression they could improve the protein yield of correctly folded protein by 25% (Qiu, Swartz and Georgiou, 1998).

If the proteins are misfolded, they will be degraded by proteases (Strauch, Johnson and Beckwith, 1989).

1.5. The Twin-Arginine Translocase (Tat) pathway

The second major export system is the Tat system which is found in most bacteria (e.g. Gram-positive bacteria such as *B. subtilis* and *S. aureus* and Gram-negative bacteria such as *E. coli*) and in the chloroplasts of higher plants and algae (Sargent, 2007; Patel, Smith and Robinson, 2014). The Tat pathway transports proteins in a folded state and could therefore be capable of transporting more complex biopharmaceuticals that require cofactor insertion and disulphide bond formation. In previous experiments, Tat has been exporting successfully single chain variable fragment (scFv) and human growth hormone (hGH) (Browning *et al.*, 2017). So far the protein yields have been too low for commercial use and various efforts have been done to improve Tat export (Robinson *et al.*, 2011). However, it has been shown that the Tat pathway transports correctly folded proteins and heterologous proteins into the periplasm, suggesting that it has potential as a powerful tool for recombinant protein production (Matos *et al.*, 2014; Alanen *et al.*, 2015). Tat specific proofreading is discussed in more detail in section 1.5.5 Proofreading and quality control of the Tat translocon.

In extremophile bacteria (e. g. *Haloferax volcanii* and *Natrialba magadii*) approximately 50% of the proteins are transported over the Tat pathway. In contrast, in *E. coli* only about 30 proteins in total are transported over the Tat machinery (Rose *et al.*, 2002; Dilks *et al.*, 2003; Ghosh *et al.*, 2019). When the Tat pathway in *E. coli* is deleted the bacterium is still viable. However, the cells are unable to divide and a filamentous chain phenotype occurs due to the lack of a cell wall amidase required for daughter cell separation (Stanley, Findlay and Berks, 2001; Bernhardt and De Boer, 2003; Ize *et al.*, 2003). Interestingly, an overexpression of the *tat* genes does not lead to a change in phenotype and the cells are capable of expressing high concentrations of recombinant protein even throughout fermentation processes (Matos *et al.*, 2012). Native Tat substrates range in size from 30 kDa up to 150 kDa and the pathway has been named after the Arg-Arg motif found near the N-terminus of the signal peptide (Delisa, Tullman and Georgiou, 2003; Palmer and Berks, 2012; Berks, Lea and Stansfeld, 2014). The Arg-Arg motif is found in almost all signal peptides targeting the proteins to the Tat pathway and some examples are shown in Table 1 on page 26.

1.5.1. Tat membrane proteins

Many Tat systems utilise three membrane proteins, TatA, TatB, and TatC (Müller and Klösgen, 2005; Kanehisa Laboratories, 2010). The three integral membrane proteins, and additionally TatD, are encoded by the *tatABCD* operon which is shown in Figure 6. *tatD* is not involved in the Tat pathway and encodes a DNase (Wexler *et al.*, 2000). The *E. coli* genome additionally contains a *tatE* gene which is found further downstream of the *tatABCD* operon.

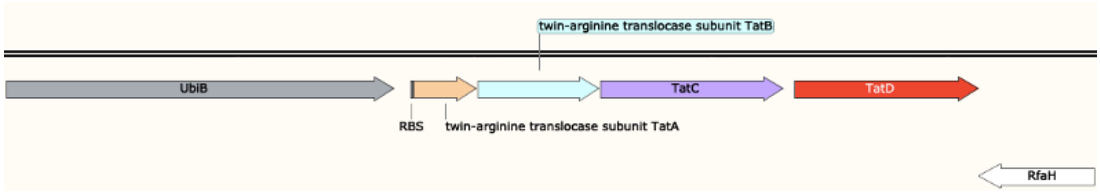


Figure 6: ***E. coli tatABCD.*** *tatABCD* operon is flanked by the genes *ubiB* and *rfaH* genes. TatE is not shown here as it is located elsewhere in the genome.

TatA is 89 amino acids long with a mass of 9.6 kDa, TatB is a 171 amino acid long protein (18.5 kDa), and TatC consists of 258 amino acids with a molecular weight of 28.9 kDa (Patel, Smith and Robinson, 2014). TatA and TatB share about 20% sequence similarity but have distinct functions (Lee *et al.*, 2002; Hicks *et al.*, 2003). TatC, as the largest membrane protein in the complex, traverses the membrane six times. TatA and TatB span the membrane once and each has an unstructured C-terminal region and a α-helix at their N-terminus. C- and N-terminal of all three proteins are located in the cytoplasm. Figure 7 shows the topography of TatA, TatB, and TatC in the cytoplasmic membrane. TatE is a paralogue for TatA and shares 57% sequence similarity and according to Eimer *et al.* TatA and TatE share a 'striking functional and structural similarity'(Sargent *et al.*, 1998; Eimer *et al.*, 2015). TatE seems to be *E. coli* specific as this gene could not be found in many Gram-negative bacteria. (Berks, Lea and Stansfeld, 2014; Patel, Smith and Robinson, 2014) It has been shown that TatE can fulfil the role of TatA in Δ*tatA* mutant strains. *tatE* is expressed in such low levels that only basic functions can be maintained (Robinson and Bolhuis, 2004).

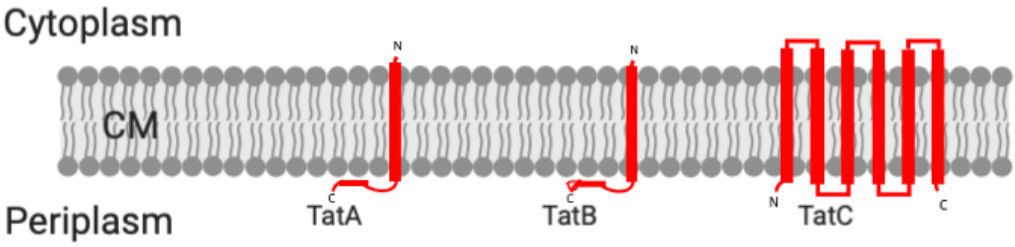


Figure 7: **Topography of the proteins involved in the *E. coli* Tat pathway.** Typically, three integral membrane proteins are encoded by the tatABC operon. TatA is 89 amino acids long, TatB is a 171 amino acid protein and shares a 20% similarity with TatA. In contrast, TatC consists of 258 amino acids and traverses the membrane 6 times. Both, C- and N-termini are located in the cytoplasm.

TatA, TatB, and TatC are expressed in different amounts. Their ratio is approximately 30:1:0.4 for TatA:TatB:TatC (Jack *et al.*, 2001; Berks, Palmer and Sargent, 2003). Jack *et al.* report that *TatA* is being expressed the most and *TatB* is being expressed 26-fold lower than *TatA* (Jack *et al.*, 2001).

Depending on whether the bacteria is Gram-positive or Gram-negative, differences can be found in the Tat system. In Gram-negative bacteria TatA, TatB, and TatC build a complex which is essential for the functionality of the Tat pathway. Gram-positive bacteria with a high GC-content and plants contain a TatABC based Tat system (Dilks *et al.*, 2003; Joshi *et al.*, 2010). However, in gram-positive bacteria with a low-GC-content in the genome and in archaea often only TatAC subunits exist (Barnett *et al.*, 2009; Palmer and Berks, 2012). The Gram-positive bacterium *Bacillus subtilis* (low-GC-content genome) uses two Tat systems which are both build of TatAC subunits

(TatAdCd and TatAyCy) (Barnett *et al.*, 2009; Goosens, Monteferrante and Van Dijl, 2014).

1.5.2. Tat specific signal peptides

As previously noted, the Tat specific signal peptides all contain a twin-arginine motif (SRRxFLK) near the N-terminus which gave the pathway its name (Hinsley *et al.*, 2001; Sargent, 2007). After the RR follows usually a polar amino acid (Müller and Klösgen, 2005). Furthermore, the signal peptides possess a positively charged n-region and a hydrophobic h-region like Sec signal peptides (Natale, Brüser and Driessen, 2008). The main difference between Tat and Sec signal peptides is that the former are generally longer. On average Tat signal peptides are about 24 to 38 amino acids long compared to typically 20 amino acid long signal peptides targeting the Sec pathway (Patel, Smith and Robinson, 2014). Also, the Tat signal peptide h-region is less hydrophobic compared to the Sec signal peptide h-regions, shown in Figure 8 (Cristóbal *et al.*, 1999). Furthermore, some signal peptides contain at the c-region an increased amount of positive charged amino acids which has been identified as a Sec-avoidance motif (Blaudeck *et al.*, 2003). For *E. coli* 22 to 34 Tat targeting signal peptides have been predicted. Some of those target the proteins to the Tat pathway and some signal peptides are promiscuous and can mediate export to Tat and Sec. Such a promiscuous signal peptide is AmiC (Blaudeck *et al.*, 2003; Dilks *et al.*, 2003; Tullman-Ercek *et al.*, 2007). It is unclear why some Tat signal peptides are ‘ambiguous’ and whether the signal peptides have an influence on the folding of the mature protein is unclear. This might be possible since Sec-dependent proteins tend to remain unfolded in the cytoplasm.

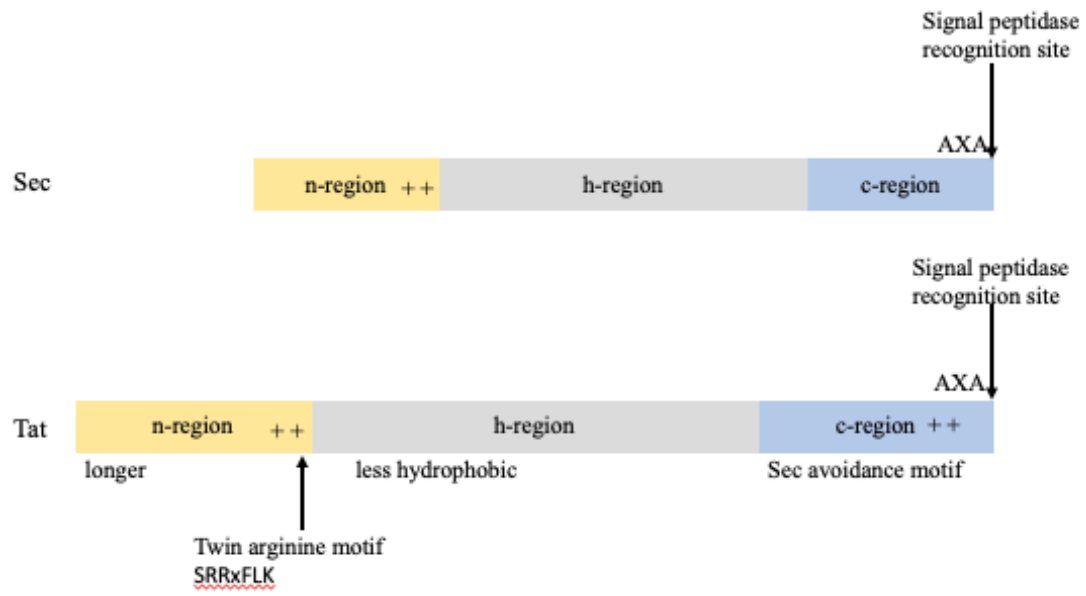


Figure 8: **Comparison of the general motif of signal peptides targeting to the Sec and the Tat pathway.** Tat signal peptides are in comparison 4 to 18 amino acids longer, their h-region is less hydrophobic, they also contain a twin-arginine motif in the n-region and a sec avoidance motif in the c-region. Adapted from Cristóbal *et al.*, 1999.

The signal peptides used in this study are AmiC, TorA, PhoD, and HyaA. Their signal sequence and shared motif is shown in Table 1. HyaA, TorA, and PhoD are Tat unambiguous signal peptides and AmiC, whilst being a signal peptide for the Tat pathway, can additionally direct proteins to the Sec pathway(Tullman-Ercek *et al.*, 2007).

Table 1: **Representation of the twin-arginine motif of the signal peptides used in this study.** The Tat signal peptides share a common structure with a hydrophobic h-region and a positively charged n-region. The Arg-Arg motif is underlined.

	n-region	h-region	c-region
TorA	MNNNDLFQASRRRFLA	QLGGLTVAGMLGPSLL	TPRRATA
PhoD	MAYDSRFDEWVQKLKEESFQNNTFDRRKFIQ	GAGKIAGLSLGLTIAQSVGNFEVN	AAPNFH
HyaA	MNNEETFYQAMRRQGVTRRSFLK	YCSLAATSLGLGAGMA	PKIAWA
AmiC	MSGSNTAISRRRLLQ	GAGAMWLLSV	SQVSLA

The AmiC signal peptide was shown to be a Tat signal peptide in 2003. GFP cannot be exported in a functional state via the Sec pathway. However, when fused to AmiC, GFP gets exported to the periplasm (Bernhardt and De Boer, 2003). Trimethylamine N-oxide (TMAO) reductase (TorA) has been shown to have a N-terminal Tat signal peptide (Barrett *et al.*, 2003; Palmer, Berks and Sargent, 2010; Browning *et al.*, 2017). AmiC and TorA have been found in *E. coli*. PhoD was originally discovered as a signal peptide a Tat substrate in *Bacillus subtilis* and the version of the PhoD signal peptide used in this study has a mutation within the h-region (amino acid underlined in Table 1) (Pop *et al.*, 2002; Gerlach, Pop and P., 2004). The last presented signal peptide is that of HyaA. HyaA is also found natively in *E. coli* and the HyaA signal peptide has

successfully transported proteins by Tat (Mori and Cline, 1998; Palmer, Sargent and Berks, 2005).

1.5.3. Disulphide bond formation in the cytoplasm

As previously discussed, proteins may require disulphide bond formation to be folded correctly and maintain stability. It has been suggested that many proteins exported by the Tat pathway have to be correctly folded before transport because they contain cofactors that are enzymatically inserted in the cytoplasm (Rose *et al.*, 2002; Dilks *et al.*, 2003; Palmer and Berks, 2012; Frain, Robinson and van Dijl, 2019).

It was thought that the reducing environment in the cytoplasm was incompatible with disulphide bond formation (Mallick *et al.*, 2002). However, deleting the two main reducing pathways (Δ*gor*/Δ*trxB*) allows disulphide bonds to form in the cytoplasm (Prinz *et al.*, 1997). The pathways involved in cytoplasmic disulphide bond formation are still poorly understood (Hatahet, Boyd and Beckwith, 2014). In this study, we support disulphide bond formation by providing eukaryotic protein disulphide-isomerase (PDI) and sulfhydryl oxidase (ErV1p) on the expression plasmid instead of using a Δ*gor*/Δ*trxB* double mutation. The system has been called CyDisCo after 'cytoplasmic disulphide bond formation in *E. coli*' and efficient export of a disulphide-bonded protein in CyDisCo strains has been presented by the Colin Robinson lab (Matos *et al.*, 2014; Alanen *et al.*, 2015).

It has been shown that chaperones are involved in the cytoplasmic folding process, such as in integrating cofactors and folding chaperones that promote assembly of enzymes (Turner, Papish and Sargent, 2004) and their role in the Tat pathway will be

further discussed at a later stage. After the disulphide bond formation, proteins are directed to the Tat translocase embedded in the cytoplasmic membrane.

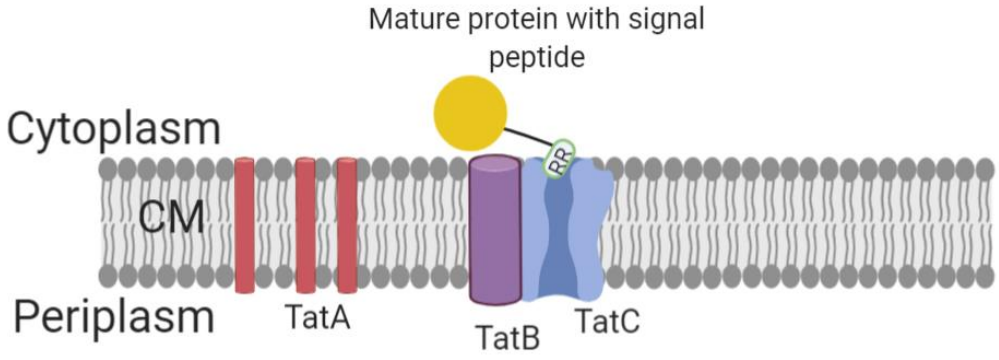
1.5.4. Tat specific translocation in *E. coli*

E. coli is a Gram-negative bacterium and its Tat translocon is built up of the three cytoplasmic membrane proteins, TatA, TatB, TatC (Ize *et al.*, 2003). TatB and TatC build a complex where each of the membrane proteins occurs six to eight times building sub-domains (Bolhuis *et al.*, 2001; Tarry *et al.*, 2009). The predicted molecular weight of the TatBC complex is approximately 600 kDa and the TatA complex has been found in variable sizes. This suggests the possibility of variable sizes in the full TatABC complex but that has not been confirmed yet (Palmer, Sargent and Berks, 2005; Leake *et al.*, 2008; Baglieri *et al.*, 2012).

In the first step, the Tat signal peptide interacts with a conserved surface patch on the cytoplasmic face of TatC and therefore interacts with the TatBC complex (Rollauer *et al.*, 2012). This step does not require energy. The proofreading mechanism of the Tat pathway might be involved in the targeting of the substrate to the TatBC complex (Kostecki *et al.*, 2010). The proofreading mechanism will be discussed at a later point. The binding of the signal peptide promotes conformational changes in TatB, displacing TatB from the TatC binding site, allowing TatA to bind to TatC (Gérard and Cline, 2007; Alcock *et al.*, 2016; Habersetzer *et al.*, 2017). TatA forms the protein conducting channel (Jack *et al.*, 2004). Researchers could show that the twin-arginine motif of the signal peptide is interacting with TatC (Alami *et al.*, 2003; Gérard and Cline, 2006), shown in Figure 9A. The assembly of the TatABC complex is proton-

motive force (PMF) dependent (Mould and Robinson, 1991; Cline, 2015). The TatABC complex is presented in Figure 9B. It is not known how the substrate passes through the membrane. A recent model suggests that the TatA oligomer facilitates translocation by weaking the cytoplasmic membrane. This is based on fluorescence data showing that the TatA hinge region and N-terminal part of the TatA amphipathic helix move toward the membrane surface which correlated with membrane weakening. Further data shows that TatA oligomers smaller than 25 TatA subunits are unstable in the lipid bilayer creating bilayer ruptures. Crosslinking shows that oligomers are build-up of eight to 16 TatA subunits. The existence of oligomers with less than 25 TatA subunits suggests that the translocase creates an environment permitting the assembly of bilayer-destabilizing TatA oligomers (Berks, 2015; Cline, 2015; Habersetzer *et al.*, 2017; Hou *et al.*, 2018).

A



B

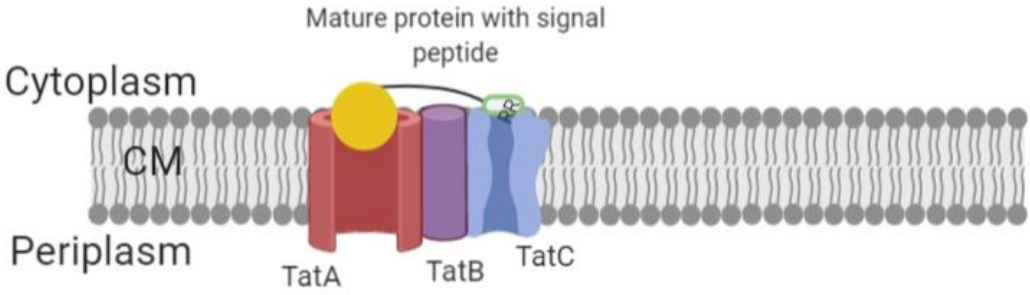
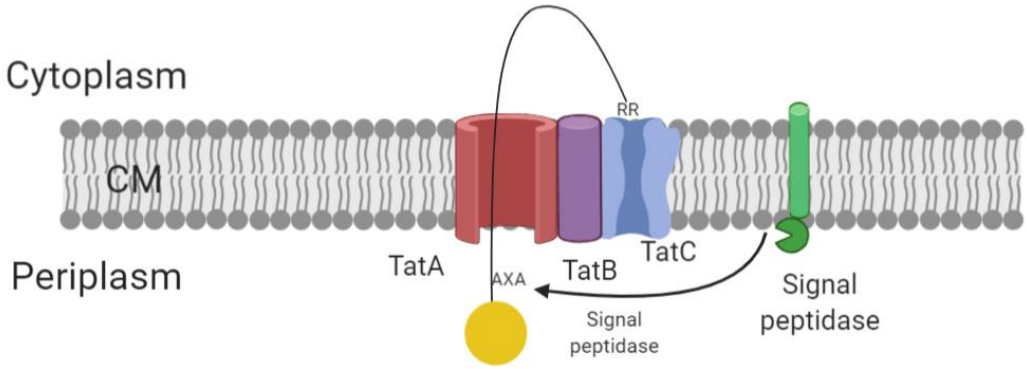


Figure 9: **Signal peptide interaction with the TatBC complex. A** The signal peptide interacts with the TatBC complex. Throughout that interaction, the Arg-Arg motif binds to TatC whilst the substrate is near TatB. The proton-motive force causes the substrate – signal peptide compound to bind more tightly. **B** Then, TatA interacts with the TatBC complex building a oligomeric ring structure and the substrate crosses the membrane.

In the next step, the substrate presumably passes through the TatABC complex and the signal peptide is cleaved of by signal peptidase (Figure 10A). It is not known whether in this step of the translocation process further energy from the PMF is required (Berks, Palmer and Sargent, 2003).

A



B

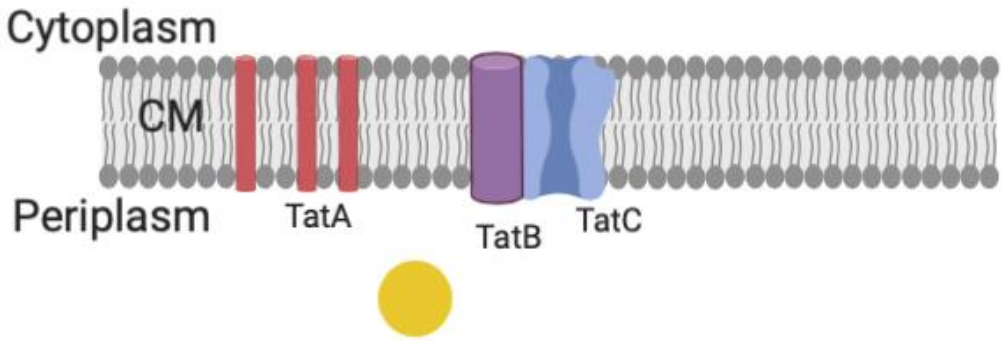


Figure 10: **Signal peptide cleavage. A** When the mature protein (yellow) passes through the passenger domain, the signal peptide will stay bound to the TatBC complex (purple and blue) which maintains it in the cytoplasm. The cleavage site of the signal peptide is in the periplasm and the signal peptide gets cleaved off by the signal peptidase (green). It is not known whether the signal peptide follows the substrate or whether it is degraded in the cytoplasm. **B** After the mature protein has been cleaved off the signal peptide, TatA (red) changes back to its original state.

Once the signal peptide is cleaved off and the mature protein is released into the periplasm, TatA depolymerases from the TatBC complex, as shown in Figure 10B, and the membrane proteins return to their original state.

1.5.5. Proofreading and quality control of the Tat translocon

Proofreading of proteins is essential for the organism. Most proteins exported over the Tat pathway are essential for the viability of the cell and many of those proteins contain cofactors. 18 out of 30 of the *E. coli* native Tat substrates require the insertion of redox cofactors (Palmer, Sargent and Berks, 2005). The proteins need to be correctly folded to maintain viability of the cells but also, the Tat translocase needs to prevent export of proteins that are not fully folded (Miot and Betton, 2004). A number of the complex Tat substrates, such as TorA, are dependent on the insertion of cofactors. TorD is a proofreading chaperone for TorA which is essential for the assembly and stability of TorA. Furthermore, TorA maturation improves by two to three fold when TorD is present (Genest *et al.*, 2005; Dow *et al.*, 2013).

As previously discussed, Tat exports folded proteins and appears to reject misfolded proteins. The Tat system appears to use two stages of proofreading activity. On the first stage proofreading is carried out by the proteostasis network. Additionally, there appears to be a second stage of proofreading which occurs at the Tat translocase itself. The Tat translocase rejects substrates that are not correctly folded or are not recognised as a Tat substrate (Delisa, Tullman and Georgiou, 2003; Turner, Papish and Sargent, 2004).

The possibility of a Tat specific proofreading at the Tat translocon has been suggested in some studies (Delisa, Tullman and Georgiou, 2003; Richter and Brüser, 2005; Lim *et al.*, 2009; Robinson *et al.*, 2011). It has been shown that PhoA (a Sec substrate) that requires periplasmic disulphide bond formation can be exported via the Tat pathway if a Tat signal peptide is attached and disulphide bond formation in the cytoplasm is

enabled (Richter and Brüser, 2005). The protein is not exported in wild type strains, showing that a correct 3D structure is required for transport by Tat. Another research group shows that alkaline phosphatase, scFv, and F_{AB} cannot be exported by the Tat pathway in the absence of disulphide bonds (Delisa, Tullman and Georgiou, 2003). Furthermore, used the proofreading mechanism has been used as a Tat-mediated genetic selection technique to identify soluble protein domains of mammalian cells (Lim *et al.*, 2009).

In the first step, the chaperones interact with the substrate and suppress transport until folding has occurred. This may be achieved by chaperone shielding the Arg-Arg sequence in the signal peptide to prevent a prematurely targeting to the Tat pathway (Santini *et al.*, 1998). In the second step, the substrate is folded, and cofactors are integrated where necessary. Typical cofactors found in Tat substrates are iron-sulphur and iron-nickel clusters, copper but also molybdopterin and FAD (Berks, Palmer and Sargent, 2003). In the third step, the substrate is released and is escorted to the Tat pathway (Jack *et al.*, 2004; Hatzixanthis *et al.*, 2005; Robinson, Cristina F. R. O. Matos, *et al.*, 2011).

As previously discussed in the section 1.3 Protein folding in bacteria, chaperones are a crucial factor for the proteostasis network. Some chaperones, such as TorD, are specifically involved in the translocation of Tat substrates. Additionally, work from different research groups suggests a Tat specific proofreading located at the translocon itself. However, the mechanism behind that is not understood. Research shows that the deletion of the last 40 residues of TatA does not impact the export capability and are dispensable for the translocation of substrates of the Tat translocon. The same has been found for the last 70 residues of TatB (Lee *et al.*, 2002). The function of the

truncated sequences is not known and the question arises how the truncations influence the recognition and rejection of Tat substrates.

Both, the Sec and the Tat pathway, have advantages and disadvantages for expressing biopharmaceuticals. The Sec pathway exports a variety of proteins; however, it is not capable of exporting complex proteins that require cofactor binding in the cytoplasm, or which fold too quickly for the Sec system to handle. In comparison, the Tat pathway is capable of exporting complex proteins but is limited in the substrates that are exported. This makes the Tat pathway a fascinating tool for biotechnology.

Another tool for biotechnology can be manipulating the protein export mechanisms. Therefore, we have additionally been investigating the *Agrobacterium tumefaciens* Tat mechanisms and its proofreading ability.

1.6. *Agrobacterium tumefaciens* Tat translocation system

Agrobacterium tumefaciens (*A. tumefaciens*) is a Gram-negative soil bacterium that is pathogenic to plants. It is known for its ability to infect the plant with its own DNA and integrates the DNA into the plant chromosome from where it induces non-necrotising tumours (Alfano and Collmer, 2001; Hohn, 2001).Three different protein secretion systems are found in *A. tumefaciens* and the Tat system is one of them (Wood *et al.*, 2001). It has been shown that the Tat system is essential for the infection of plant cells and the assembly of the flagellum (Ding and Christie, 2003). *A. tumefaciens tat* operon encodes TatABC, additionally the genome contains two *tatD* genes. One of the *tatD* genes is located 200 kb away from the operon. The other *tatD* gene is located on an accessory plasmid called pATC58 (Wood *et al.*, 2001; Ding and Christie, 2003).

It has been predicted, that TatA, TatB, and TatC are similarly integrated into the cytoplasmic membrane as they are in *E. coli* and TatA and TatB are crossing the membrane once and TatC crosses the membrane six times (Ding and Christie, 2003). *A. tumefaciens* is mostly known for its infectious behaviour towards plants and the translocation of proteins by the Tat system has not been researched extensively. Ding and Christie showed that *A. tumefaciens* is also able to export TorA-GFP expressed from a plasmid into the periplasm (Ding and Christie, 2003). Due to the ability to export TorA-GFP and a similar cytoplasmic membrane topography, *A. tumefaciens tatABC* is an interesting target operon for replacing the *E. coli* t*atABC* operon.

After looking into how to manipulate the host organism for protein production a further tool for optimising protein expression for biopharmaceutical production can be changing the promoter of expression plasmid.

1.7. Influence of the promoter on protein expression

Bacterial protein expression is naturally controlled by promoters (Haugen, Ross and Gourse, 2008). As the previous example of the cold-inducible *cspA* promoter shows, promoters are a powerful tool for recombinant protein production (Miroux and Walker, 1996; Mujacic and Cooper, 1999). The main requirements for a promoter in industry are its strength, its transcriptional activity needs to be tightly controlled, and the induction needs to be independent of components of common culture media (Makrides, 1996; Terpe, 2006). The strength of a promoter in regard to the target protein production is typically measured in the ability to express a high amount of total cellular protein. About 15 to 30% of total cellular protein is considered strong and in

rare occasions up to 50% of total cellular protein can be achieved (Miroux and Walker, 1996; Terpe, 2006).

Control of the transcriptional activity has a huge impact on protein production, especially when the desired protein has toxic effects for the cell and prevents further cell growth. Therefore, the protein yield is being affected. An example for such a protein is bovine pancreatic DNaseI. DNaseI cleaves double-stranded DNA when it is expressed in an active state and leads to cell death when active DNaseI accumulates inside the bacterial cell. DNaseI could not be expressed in *E. coli* using a *tac* promoter, presumably because the basal expression level of the tac promoter in absence of an inducer leads to the production of fully active DNaseI. The research group had to use the pDOC55 plasmid under the control of the λp_L promoter making the promoter tightly controlled in absence of an inducer (Doherty, Connolly and Worrall, 1993). Therefore, using the right promoter is essential for the quality and efficiency of the export of the target protein and can reduce cell stress (Kawe, Horn and Plückthun, 2009).

An additional requirement on promoters has evolved in recent years. With the developing recombinant protein production market in South East Asia, focus has shifted to the inducer required for the promoter (Wang, Chen and Tsai, 2012). Common inducers have been arabinose and Isopropyl-β-D-thiogalactopyranosid (IPTG). IPTG is an expensive reagent to use and it can be toxic to the cells therefore there is a clear need for alternative promoters that are tightly regulated (Figge *et al.*, 1988; Thermo Fisher Scientific, 2020). Promoters that are commonly used for industrial purposes include, for example, the *lac* or *tac* promoters, which are both IPTG induced, see Table 2 (Gronenborn, 1976; Brosius, Erfle and Storella, 1985; Rosano and Ceccarelli, 2014).

Table 2: **Promoters commonly used for recombinant protein expression in *E. coli*.** The majority of promoters are induced by adding IPTG and the promoter strength can be varied by the concentration of IPTG added.

Espression system the promoter is based on	Inducer (concentration used)	Reference
lac promoter	IPTG (0.05 – 2.0 mM)	(Gronenborn, 1976)
grac promoter	IPTG (0.05 – 2.0 mM)	(Phan, Nguyen and Schumann, 2012)
trc and *tac* promoter	IPTG (0.05 – 2.0 mM)	(Brosius, Erfle and Storella, 1985)
araBAD promoter (p_{BAD})	L-arabinose (0.001 – 1.0%)	(Guzman *et al.*, 1995)
cspA promoter	Thermal induced	(Mujacic and Cooper, 1999)
T7	IPTG	(Studier, 1991)

To design a novel promoter system, promoter sequences in microorganism are identified. Common promoter sources are *E. coli*, λ, and T7 (Makrides, 1996). The assembly of a promoter is essential for its effectiveness. Moving the operator of the *lac* promoter within the promoter has shown that the level of expression is majorly impacted (up to 70-fold difference in the repression) which shows how challenging and tedious promoter design can be (Lanzer and Bujard, 1988; Browning, Butala and Busby, 2019).

In this work, a novel promoter is tested for its ability to control recombinant protein production. The promoter is based on the *E. coli* K-12 *narG* operon promoter which controls the expression of the Nar nitrate reductase. Nitrate and nitrite are used as electron acceptors for respiration, the nitrate is reduced to nitrite and then in a secondary step reduced to ammonium (Trivedi and Ju, 1994; Lin and Stewart, 1997). This is shown in the following to equations:

(1) $$NO_3^- + NADH + H^+ \rightarrow NO_2^- + NAD^+ + H_2O$$

(2) $$NO_2^- + 3NADH + 5H^+ \rightarrow NH_4^+ + 3NAD^+ + 2H_2O$$

E. coli can only utilise nitrate under anaerobic conditions. The genes encoding the nitrate reductase are not expressed when O_2 is present (Richardson and Watmough, 1999; Merrimack Pharmaceuticals, 2016). The native promoter is additionally controlled by two transcription factors FNR and NarL. FNR is inactivated by oxygen and when oxygen levels are low or oxygen is absent, FNR binds to the *narG* promoter to activate transcription. However, the *narG* promoter also requires activation by NarL which only occurs when nitrate is present in the medium (Walker and DeMoss, 1992). The *narG*-CC promoter used in the current studies is a fusion of the *narG* promoter and the *CC* (-40.5) promoter, the upstream of the promoter carries NarL, and integration host factor (IHF) sites from p*narG* and the downstream region cyclic AMP receptor protein (CRP) site from CC (Figure 11). This promoter has been designed by the Steve Busby lab at the University of Birmingham. CRP functions as an activator within the promoter system (Browning, Butala and Busby, 2019). NarL fulfils the role of a regulator and activator (Andrew J. Darwin and Stewart, 1996). CRP is a strong activator when it is at position -41.5 upstream of the transcriptional start site. When it

is at position -40.5, however, CRP is less active and the promoter requires another transcription factor (Rossiter *et al.*, 2011, 2015). At this promoter, the CRP binding site is located at position -40.5 and, therefore, CRP is less active and requires support from NarL. This ensures that the promoter is responsive to nitrate.

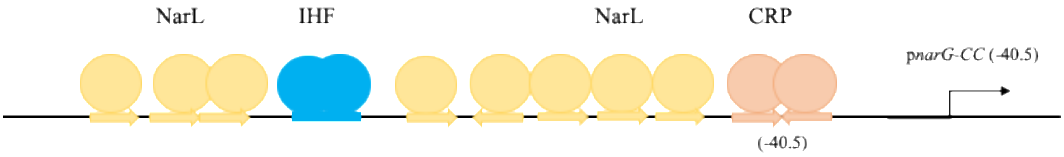


Figure 11: **The figure shows a schematic presentation of the *narG-CC* (-40.5) promoter.** The location of the binding sites for NarL, IHF and CRP are indicated. The start site of transcription is shown by a bent arrow. (Adapted from a presentation given by Douglas Browning at the Recombinant Protein Production 10 conference).

1.8. Aims and objectives

The primary aim of this project is to develop systems to improve recombinant protein export using genetic engineering and promoter change. *E. coli* is a strong protein expression host with many benefits. However, the established protein expression pathway (Sec pathway) has its limit. The Tat pathway could expand the landscape of protein expression in bacteria and purification of proteins from the periplasm has benefits in the downstream process (Quax, 1997). Therefore, the preferred pathway for protein translocation is the Tat pathway and we aim to improve the understanding of Tat export and exploit its limits as well as establishing a new tool for recombinant protein production.

In this study, we have successfully replaced the *E. coli tatABC* operon with the *tatABC* operon from *A. tumefaciens*. We aim to test the limits of the developed strain and investigate whether the *tat* operon of *A. tumefaciens* allows us to avoid proofreading mechanisms in place.

Furthermore, the research objectives have been to establish whether TatABC from *A. tumefaciens* can complement the *E. coli* Δ*tat* phenotype and changes *E. coli* TatABC selectiveness.

We also aim to test a promoter based on the nitrate pathway in *E. coli* and see whether the *narG*-CC promoter can improve export in general and supports Tat export.

Additionally, we are interested in the novel enzyme PETase and aim to investigate whether the enzyme can be expressed in an active state into the periplasm by the Tat pathway using different signal peptides and strains for expression.

2. Materials and Methods

2.1. Supplies of chemicals, reagents and equipment used

All reagents, materials and equipment were obtained from the companies listed below.

Acros Organics: Trimethylamine N-Oxide dihydrate 98%,Methyl violgen 98%

Beckman Coulter Inc (USA): TL-100 Ultracentrifuge TLA 100.3 rotor; Avanti J-25, JA10 rotor; DU 730 UV/Vis Spectrophotometer

Bio-Rad (USA): Agarose; Bio-Rad Chemidoc XRS+; Bio-Rad Mini-PROTEAN

Cayman Chemical: Nitrate/ Nitrite colorimetric assay

Eppendorf: 1.5mL centrifuge tubes; PCR tubes; Centrifuge 5417R; pipette tips

Eurofins (UK): Oligo's/Primers

Fisher Scientific (UK): Anti-his (C-term) unconjugated antibody; dNTP nucleotide mix; GeneRuler (100 bp and 1 kb); Mircobiological spreaders; Inoculation loops; Nanodrop 2000c

Formedium (UK): PBS

GE healthcare (UK): ECL^{TM} detection reagents; PVDF and nitrocellulose membrane

GenPure: UV/UF water purification system for MilliQ H_2O

Eppendorf: Pipettes, P2, P20, P200, P1000

Grant: Heat block; SUB Aqua 12 Plus water bath

Greiner Bio-one: Universals; Easy load tips

IBA Lifescienes: Anti-strep HRP conjugate

IDT: Oligo's/Primers

Infors (Switzerland): Multitron Pro shaking incubators; Minifors 2 fermenters

Invitrogen (USA): Anti-hexahistidine (C-term) conjugated

Leica: Type F immersion liquid, DFC9000 GT, DMR

Life technologies (USA): BSA

MSE: Soniprep 150 plus

New England Biolabs (UK): Phusion high fidelity DNA polymerase (2U/ μL); Restriction enzymes; NEB® Turbo compotent cells; Monarch PCR & DNA Cleanup kit, NEB® BL21 DE3 competent cells; Coloured protein standard broad range(11-245kDa) Monarch Gel Extraction Kit; Antarctic Phosphatase

Panasonic: Heated static incubator

Promega (UK): dNTP mix (200μM); Anti-mouse HRP conjugate; Anti-rabbit HRP conjugate

Quiagen (Germany): QIAprep Spin Miniprep Kit, Ni-NTA Spin column purification of 6xhis tagged proteins under native conditions from *E. coli* cell lysate

Roche applied science (UK): T4 DNA ligase

Sarstedt: Serological pipettes; 15mL and 50mL falcon tubes; Petri-dishes 92x16 mm; Filter tips; 1 mL cuvettes

Sigma-Aldrich (UK): Lysozyme; Anti-flag HRP conjugate, LB Broth, Nitrite/ Nitrate colorimetric

Starlab: Starguard comfort gloves

Thermo Fisher Scientific Inc (USA): Nanodrop 2000c; Megafuge 16R Centrifuge, 75003181 rotor; MaxQ 800 Shaking incubators, GeneJET Genomic DNA Purification Kit

VWR: Microcentrifuge

All images used in this work were created using bioRENDER, powerpoint, and SnapGene.

2.2. DNA techniques

2.2.1. Preparation of plasmid DNA

For preparing the plasmid DNA, Qiagen QIAprep Spin Miniprep Kit was used per manufacturer's instruction. The plasmids were eluted in 50µL elution buffer (10mM Tris-Cl pH 8.5). Finally, the concentrations were determined using the Nanodrop.

2.2.2. Preparation of genomic DNA

For preparing genomic DNA, the GeneJET Genomic DNA Purification Kit was used per manufacturer's instruction. The DNA was eluted in 200µL elution buffer (10mM Tris-Cl, pH 9.0, 0.1mM EDTA). The concentrations were determined using the Nanodrop.

2.2.3. Amplification of DNA via Polymerase Chain Reaction

PCR's were performed using Bio-Rad T100^{TM} Thermal Cycler. Each contained, 1.5µL template DNA (80–100ng) or 1.5µL of colony mix (colony diluted in 20µL dH_2O). Furthermore 0.5µL Phusion high fidelity DNA polymerase, 1µL dNTP mix, 10µL 5X Phusion HF buffer, and one 1µL each of forward and reverse primer (0.5µM) were added and the mix was made up with 50µL with autoclaved dH_2O). The primers used in this study are presented in Table 3.

An example cycle is given below:

Initial denaturation	98°C	2mins	
Denaturation	98°C	10sec	x35 cycles
Annealing	58°C	30sec	
Elongation	72°C	2mins 30sec	
Final extension	72°C	10min	
Hold	4°C	∞	

Table 3: **Cloning primers used in this work.** Primers have been supplied either by Eurofins or integrated DNA technologies (IDT) as stated in the table.

Primer name	Sequence 5' → 3'	Supplier
pMA-RQ_F	GAAGGCCGTCAAGGCCGCAT	Eurofins
pMA-RQ_R	CGGAAGGCCCATGAGGCCCAG	Eurofins
TorAR	TTTAAGAAGGAGATATACATATGAACA ATAACGATCTCTTTC	Eurofins
scFvF	CAGTGGTGGTGGTGGTGGTGTGCGGCCC CATTC	Eurofins
pET22bF	AAGAGATCGTTATTGTTCATATGTATAT CTCCTTCTTAAAGTTAAAC	Eurofins
pET22bR	AGGATCTGAATGGGGCCGCACACCACC ACCACCA	Eurofins
AmiC_PETase_F	TAAGTCAGGTCAGTCTGGCTATGAACTT CCCCCGT	IDT

AmiC_vector_F	GAGGCACGGGGGAAGTTCATAGCCAGACTGACCTG	IDT
PETase_vector_R	TTCGCACCGCGAACTGTTCCCACCATCACCATCACC	IDT
HyaA_PETase_F	CACCAAAGATTGCCTGGGCGATGAACTTCCCCCGT	IDT
PETase_insert_R	TAATGGTGATGGTGATGGTGGGAACAGTTCGCGG	IDT
HyaA_vector_F	GAGGCACGGGGGAAGTTCATCGCCCAGGCAATC	IDT
TorA_PETase_F	TAATGGTGATGGTGATGGTGGGAACAGTTCGCGG	IDT
TorA_vector_F	GAGGCACGGGGGAAGTTCATCGCAGTCGCACG	IDT
GFP_F	CGCCGCGACGTGCGACTGCGATGGTGAGCAAGGGC	IDT
pYU49R_GFPhis	TCCTCGCCCTTGCTCACCATCGCAGTCGCACG	IDT
pYU49F_GFPhis	GCATGGACGAGCTGTACAAGCACCACCACCACCACCACTGAGATCCGA	IDT
GFP_R	CAGTGGTGGTGGTGGTGGTGCTTGTACAGCTCGTCC	IDT
R16_ins_F	CAACCGCAGCACAGGCAGCAAAACTGAATGAAAGCCATCGTC	IDT

R16_vec_F	CGATGGCTTTCATTCAGTTTTGCTGCCT GTGCTGC	IDT
R16_ins_R	ATCTCCTTCCACCTAAACGGTTAATGAT GATGGTGATGATGTTCC	IDT
R16_vec_R	ATCATCACCATCATCATTAACCGTTTAG GTGGAAGGAGAT	IDT
GB1_ins_F	CAACCGCAGCACAGGCAGCAATGACCT ATAAACTGATTCTGAATG	IDT
GB1_vec_F	AGAATCAGTTTATAGGTCATTGCTGCCT GTGCTGC	IDT
GB1_ins_R	ATCTCCTTCCACCTAAACGGTTAGTGAT GGTGATGATGATGTTCG	IDT
GB1_vec_R	ATCATCATCACCATCACTAACCGTTTAG GTGGAAGGAGAT	IDT
PhoDn_F	AATTCCACAGAGGAGGATCCATGGCCT ATGATAGCCGC	IDT
PhoDn_R	GAGGCACGGGGGAAGTTCATATGAAAG TTCGGTGCTGC	IDT
pPETase_F	ATGCAGCACCGAACTTTCATATGAACTT CCCCCGTGC	IDT
pPETase_R	AAGCGGCTATCATAGGCCATGGATCCTC CTCTGTGGAAT	IDT

2.2.4. Purification of DNA from the genome

PCR products were purified for sequencing using NEB Monarch PCR & DNA Cleanup Kit as per manufactuer's instructions. The PCR product was eluted in 20µL elution buffer (10mM Tris, 0.1mM EDTA, pH 8.5 elution buffer) and the concentration was measured using the Nanodrop.

2.2.5. Agarose gel electrophoresis

DNA was run on agarose gels containing 1% (w/v) agarose diluted 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 pH8). The 1x TAE and agarose were in a beaker boiled in the microwave for 1min until the agarose has dissolved. Depending on the size of the agarose gel electrophoresis either 50mL 1x TAE buffer and 0.5g agarose or 150mL 1x TAE buffer and 1.5g agarose were used. 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 120V for 30 to 50 mins depending on the expected sizes. For the visualisation of the agarose gels a Bio-Rad Gel doc was used.

2.2.6. Purification of DNA from agarose gels

Under UV transilluminescence the bands are visualized to cut out the desired DNA fragments from the agarose gel using a scalpel blade. Using the Monarch DNA Gel Extraction Kit the DNA was cleaned up. DNA was eluted in 20µL elution buffer and the concentration was determined using the Nanodrop.

2.2.7. Restriction Digests of DNA

For restriction digests, typically 1µg of DNA was used and mixed up with 5µL 10x Cutsmart NEB buffer and 0.25µL of each enzyme (Table 4). The total reaction volume was made up to 50µL with MilliQ water. The reaction was incubated for 2hours in a 37°C heat block. To the backbone restriction digest, 2.5µL antarctic phosphatase and 5µL 10x Antarctic Phosphatase Reaction Buffer was added after 1hour and 30mins. After incubation, the backbone reaction mix was heat shocked at 80°C for 2min to inactivate the antarctic phosphatase. Then, the reaction mix was run as described in 2.2.5 on an agarose gel electrophoresis and purified.

Table 4: **Restriction enzymes used in this work.**

Enzyme name	Sequence 5' → 3'
NdeI	CA-TATG
XhoI	C-TCGAG
KpnI-HF	GGTAC-C
HindIII-HF	A-AGCTT
SalI	G-TCGAC

2.2.8. Ligation of DNA fragments into plasmid vector backbone

For a 10µL ligation mix 1µL T4 DNA ligase, 1µL 10x T4 ligase buffer was used. Furthermore, the used insert to vector ratio used was typically 3:1. The ligation was incubated for 1 to 3 days at 4°C and then transformed.

2.2.9. Gibson assembly cloning

For constructs designed by Gibson cloning a 20 µL reaction mix was prepared using a 1:10 dilution of the primers. PCR's were performed using Bio-Rad T100^{TM} Thermal Cycler. Each contained, 0.3µL template DNA (80–100ng). Furthermore 0.1µL Phusion high fidelity DNA polymerase, 0.4µL dNTP mix, 4µL 5X Phusion HF buffer, and one 0.8µL each of forward and reverse primer were added and the mix was made up with 13.6µL with autoclaved dH_2O. As previously described the samples were visualized using agarose electrophoresis. After adding 0.5µL 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 5 and the assembly mix was incubated for an hour at 50˚C before transformation into NEB Turbo competent cells.

Table 5: **Gibson assembly composition.**

	Concentration
Amount of each DNA fragments	0.02 – 0.5 pmols
Gibson Assembly Master Mix (2x)	10 µL
Add deionized H_2O to a total volume of	20µL

2.2.10. Sequencing of plasmid DNA

For this study, the sequencing service of GATC and GENEWIZ was used. Typically, 80 to 100ng of DNA was mixed with 5µL of 5mM sequencing primer (Table 6).

Table 6: **Sequencing primers used in this work.** Primers have been supplied either by Eurofins or integrated DNA technologies (IDT) as stated in the table.

Primer name	Sequence 5' → 3'	Supplier
HR1_pDocK	CAGCTTCAAAAGCGCTCT	Eurofins
HR2_pDocK	CAATTTCACACAGGAAACAGC	Eurofins
pDocInsertF	CCTCCAGCGCGGGGATCTCATG	Eurofins
pDocInsertR	GGCAGGTGGTCTGATCGCCTGG	Eurofins
Tat7R	GCAATAACCAGCAGTTCGCTCC	Eurofins
Tat13R	AGGTACGCAGCATATTCGGC	Eurofins
pET22b insR	TTATTGCTCAGCGGTGGCAG	Eurofins
ptacF_seq	GAGCGGATAACAATTTCACACAGG	IDT
t7R	GCTAGTTATTGCTCAGCGG	Eurofins
CM_pET23_SEQF	AGCTGTTGACAATTAATCATCGGCT	Eurofins

2.2.11. Sequencing of genomic DNA

For this study, the sequencing service of GATC was used. 80 to 100ng of DNA was mixed with 5 µL of 5 mM sequencing primer. The genomic DNA pieces send for

sequence were about 2000 bp large and were cloned out of the genome using the primers presented in Table 7.

Table 7: **Primers used for sequencing the *E. coli* genome.**

Primer name	Sequence 5' → 3'	Supplier
UbiB_genome F	GCCTTTCCTGGAGTCGTGG	Eurofins
rfaH_genome R	CGAAAGACAGGCTGTGAATTGCC	Eurofins

2.2.12. Constructs used in this study

A variety of constructs has been used in this study, which are presented in Table 8. Genes that have been synthesises for cloning have been ordered from ThermoFisher.

Table 8: **Plasmid construct used in this work.**

Plasmid name	Function	Tag	Reference
pACBSR	pACBSR		(Lee *et al.*, 2009)
pDoc-K	pDocK		(Lee *et al.*, 2009)
pSB15	pDOcK_ΔTatABCagro (new)		This study

pSB02	pDocK_HR1_HR2Δ_Tat13 Meso.		This study
pSB18	pET15b narG his6-GFP	His6	Douglas Browning, University of Birmingham
pSB19	pET20b narG TorA-hGH-his6	His6	Douglas Browning, University of Birmingham
pSB23	pET22b ptac TorA-hGH-his6	His6	Douglas Browning, University of Birmingham
pSB24	pET22b ptac his6-GFP	His6	Douglas Browning, University of Birmingham
pSB25	pET-21b(+)-Is-PETase-W185A	His6	(Austin *et al.*, 2018)
pSB26	pET23 TorA-ScFv(Cobra)-6his	His6	(Alanen *et al.*, 2015)
pSB27	pET23 TorA-hGH-6His+CyDisco	His6	(Alanen *et al.*, 2015)

pSB28	pET20b narG scFv(cobra) his6	His6	This study
pSB29	pET22b ptac scFv(cobra) his6	His6	This study
pSB30	pYU49-TorA-GFP-his6	His6	This study
pSB31	pEXTII-TorA-PETase-his6	His6	This study
pSB33	pEXTII-HyaA-PETase-his6	His6	This study
pSB34	pEXTII-HyaA-PETase-his6+CyDisCo	His6	This study
pSB35	pEXTII-AmiC-PETase-his6	His6	This study
pSB36	pEXTII-AmiC-PETase-his6+CyDisCo	His6	This study
pSB42	pYU49-TorA-R16-his6+CyDisCo	His6	This study
pSB43	pYU49-TorA-ProteinG-his6+CyDisCo	His6	This study

2.2.13. Gene doctoring

To construct strains containing TatABC homologues the gene doctoring method described in Lee et al. was used (Lee *et al.*, 2009). The *ubiB* region and *rfaH* region were synthesized and cloned into pDoc-K plasmid with SalI – NdeI, and KpnI – HindIII were used to clone the Tat inserts into the plasmid.

The pDoC-K plasmid containing the homology regions and Tat inserts (200 μg/ml ampicillin and 50 μg/ml kanamycin) were transformed with pACBSCE (35 μg/ml chloramphenicol) into MC4100 and plated onto LBA plates containing 5 % sucrose. One colony was used to inoculate 1 mL LB containing appropriate antibiotics and 0.5 % glucose at 37 ºC for 2 hours whilst shaking. The cells were spun down for 2 mins at 18000 x g and resuspended in 1 mL LB with 0.5 % L-arabinose and incubated at 37 ºC for 4 hours at 250 rpm. Cultures were plated on LB plates containing 50 μg/ml kanamycin and 5% sucrose and incubated O/N at 30 °C. For testing for positive strains the colonies were patched onto plates containing ampicillin and chloramphenicol and positive clones were send to with primers homologous to chromosomal DNA.

2.3. Maintenance of *E. coli* cultures

2.3.1. Media and supplements

Unless otherwise stated, Leuria Bertani (LB) medium was used as a culture medium. Table 9 describes all media used in this study. For some of the work, alternative media have been used, see Table 9. LBA (LB with the addition of 10 g/L bacto-agar) was used for growth on plates. All were supplemented with appropriate antibiotic. Ampicillin (100 μg/mL) and Kanamycin (50 μg/mL) were dissolved in water whereas

Chloramphenicol (35 µg/mL) was dissolved in ethanol. Additionally, plates with 5% sucrose content were used for selection.

Table 9: **The in this study used media and their components.**

Media	**Components**
LB	10 g/L sodium chloride, 10 g/L tryptone, 5 g/L yeast extract
LB-GT	10 g/L sodium chloride, 10 g/L tryptone, 5 g/L yeast extract, 0.5 % v/v glycerol, 0.4 % w/v TMAO
TB	24g/L yeast extract, 12g/L tryptone, 4mL/L glycerol Potassium phosphate buffer: K_2PO_4 12.5g/100mL and KH_2PO_4g/100mL
SM6	10mL 10X SM6 trace elements, 95g/L glycerol, 5.2g/L $(NH_4)SO_4$, 4.4g/L $NaH_2PO_4.H_2O$, 4.16g/L citric acid, 4.03g/L KCl, 1.04g/L $MgSO_4.7H_2O$, 0.25g/L $CaCl_2.H_2O$
10X SM6 trace elements	104g/L citric acid, 10.06g/L $FeCl_3.6H_2O$, 5.22g/L $CaCl_2.H_2O$, 2.72g/L $MnSO_4.7H_2O$, 2.06g/L $ZnSO_4.4H_2O$, 0.81g/L $CuSO_4.5H_2O$, 0.42g/L $CoSO_4.7H_2O$, 0.03g/L H_3BO_3, 0.02g/L $Na_2MoO_4.2H_2O$
MS medium	4.5g/L KH_2PO_4, 10.5g/L K_2HPO4, 1g/L $(NH_4)_2SO_4$, 0.5g/L $Na_3C_6H_5O_7$, 0.05g/L $MgSO_4$, $7H_2O$, 1mL/L *E. coli* sulphur salts, 1mL/L 1mM Sodium Selenate,

	1mL/L 1mM Ammonium molybdate tetrahydrate, LB Broth 4g/L, 4mL/L glycerol, 6.4g/L $C_4H_2Na_2O_4$
E. coli sulphur free salts	82g/L $MgCl_2$, $7H_2O$, 10g/L $MnCl_2$, $4H_2O$, 4g/L $FeCl_2$, $6H_2O$, 1g/L $CaCl_2$, 20mL concentrated HCl
1mM Sodium selenate	0.19g/L Na_2O_4Se
1mM Ammonium molybdate tetrahydrate	1.24g/L $(NH_4)_6Mo_7O_{24}$, $4H_2O$

2.3.2. Transformation of competent *E. coli* cells

Throughout the here presented work a variety of strains has been used. These are shown in Table 10.

Table 10: **Representation of the *E. coli* strains used in this work and their genotypes.**

Strain	Genotype	Reference/ source
MC4100	*Ara^R, F2 arD139 DlacU169 rspL150 relA1 flB5301 deoC1 ptsF25 rbs^R*	
NEB® Turbo Competent *E. coli*	*F' $proA^+B^+$ $lacI^q$ Δ*lacZM15* / fhuA2 Δ(lac-proAB) glnV galK16 galE15 R(zgb-*	NEB®

	*210::Tn10)*Tet^{S} *endA1 thi-1* Δ*(hsdS-mcrB)5*	
MC4100 Δ*tatABCDE*	MC4100 with Δ*tatABCDE* deletion	Tracy Palmer – Newcastle University
N11	K-12 with Δ*narG* deletion	Douglas Browning – University of Birmingham
Tn3.1	W3110 carrying a *Tac* promoter upstream of *tatABCD* and Δ*narG* deletion	Douglas Browning – University of Birmingham
BL21 TE	BL21 carrying a *pTac* promotor upstream of *tatABCD*	(Browning *et al.*, 2017)

The *E. coli* strains used in this study (Table 10) were typically inoculated in 5 mL LB medium and cultured at 37 °C, 220 rpm. After minimum of 18 hours, 100 µL of the pre-culture was inoculated into 10 mL of fresh LB medium and incubated at 37 °C, 220 rpm until an OD_{600} of approximately 0.5 was reached. The cells were spun down at 1690 x g, 4 °C for 10 mins and the supernatant was discarded. The cell pellet was re-suspended in 10 mL 100 mM $MgCl_2$ (ice-cold) and were incubated for 5 mins on ice. Cells were spun down as previously and re-suspended in 1 mL ice-cold 100 mM $CaCl_2$ and were incubated on ice for 1 hr.

The transformation was performed on ice, where 50 to 100 µL *E. coli* cells were incubated with 100 ng DNA for 30 mins. In a 42 °C water bath, the cells were heat shocked for 25 to 30 secs and immediately placed back on ice for 2 mins. The cells were recovered in LB, in 3:1 LB to cells, for 1 hr at 37 °C and plated on LBA with appropriate antibiotics. Plates were inverted and incubated O/N in the plate incubator at 37 °C or left at RT for 2 days.

2.3.3. Storage of *E. coli* cells

Cells were stored at -80 °C as a glycerol stock. Glycerol stocks were prepared using 750 µL 50% glycerol and 750 µL stationary phase culture.

2.4. Protein production and *E. coli* cell fractionation

2.4.1. Cell culture and induction of plasmids

For cell culture assays with induced plasmids, a 5 mL pre-culture was set up with *E. coli* cells containing the required plasmid and were grown O/N at 37 ºC, 220 rpm. The following morning, the cultures were diluted to OD_{600} 0.1 with fresh medium and appropriate antibiotic in 50 mL media in 250 mL Erlenmeyer flasks. Depending on the used strains and plasmids different media were used and the cells were grown at either 30 ºC or 37 ºC. The cultures were grown until OD_{600} 0.4 – 0.6 (mid-log phase) was reached. When the stationary phase was reached plasmids were induced for typically 3 hours. The here used inducers are IPTG, Arabinose, and sodium nitrate.

2.4.2. *E. coli* fermentation

Typically, a pre-culture with 5 mL MS media and appropriate antibiotic was set up and grown for 6 hours at 37 ºC, 250 rpm. 2 mL of pre-inoculate were transferred to 200 mL of SM6 media and grown O/N at 37 ºC, 220 rpm. For the inoculation of the fermenter, a volume of 300 $\frac{OD}{L}$ was used.

At the University of Kent, a 1.5L fed-batch fermenter from Infors, called Minifors2, are used. The fermenter have a fermentation volume of 0.5L is used and their components are presented schematically in Figure 12. The fed-batch fermenter is inoculated through the sample port on top of the fermenter. Furthermore, the pO_2, temperature, and pH are monitored by sensors. Acid and base are added in relation the measured pH. The temperature is controlled over a cooling/ heating sleeve. The pO_2 measurements influence throughout the fermentation process the stirrer speed. Throughout the fermentation process, glycerol is added as a carbon source and samples are taken through the sample port. Furthermore, additives like the appropriate antibiotic or anti-foaming agent is added through a septum collar on top of the fermenter. The septum collar is not shown in the image.

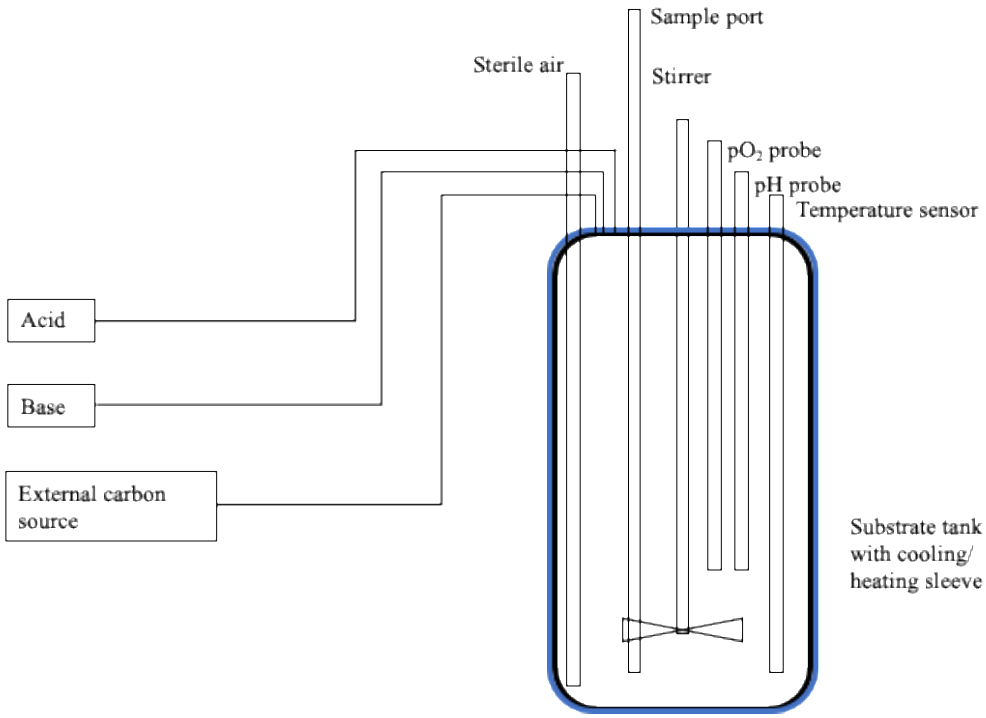


Figure 12: **Schematic presentations of the here used fed-batch fermenter.** The fermenter has a 1.5 L vessel and the fermentation volume is 0.5L. The fermenter is innoculted through the sample port with 300 $\frac{OD}{L}$. The culture is continuously monitored for its pO_2 and pH. Throughout the fermentation process acid and base are fed in relation to the measured pH. The stirrer is related to the pO_2 measured. Furthermore, a carbon source is being supplied. Throughout the here presented experiments, glycerol has been used as carbon source. Additionally, the temperature is measured continuously, and the fed-batch fermenters temperature is being controlled by a cooling/ heating sleeve. The fed-batch fermenter is inoculated through the sample port and samples are taken through the same port. Not shown in the picture is the septum collar on the top of the fermenter through which the appropriate antibiotic and anti-foam agent is added.

To prepare the fermenters for a run, they were autoclaved with 500 mL SM6 media and pH and pO_2 probe inserted. The temperature was set up to 37 ºC and temperature was maintained via a chilling mantel. Throughout the run, the pH was maintained at

pH 7 with 25 % sulphuric acid and 25 % ammonia. Before inoculation the pO_2 probe was calibrated at 100% saturation.

After inoculation the pO2 was set to cascade and the stirrers and airflow were set to maximum (stirring of 900 to 1600 rpm; total flow of 1.5 to 2 $\frac{L}{min}$). When stirrer and airflow reached their maximum, the temperature of the culture was dropped to 25 ºC. Throughout the run, the OD's were monitored, and following supplements were added to the fermentation:

OD 38 to 42: add 8 $\frac{mL}{L}$ 1 M $MgSO_4$, 7 H_2O

OD 54 to 58: add 5 $\frac{mL}{L}$ 1 M NaHPO, H_2O

OD 66 to 70: add 7 $\frac{mL}{L}$ 1 NaHPO, H_2O and begin glycerol feed (80% w/w glycerol at a rate of 0.3 of pump capacity)

OD > 75: Induction with appropriate inducer.

Induction usually was done in the late evening and samples were taken from the next morning onwards (approximately 15 hours post-induction). Cell samples were either fractioned according to the *E. coli* cell fractionation protocol or prepared as whole cell extracts.

2.4.3. E. *coli* cell fractionation

Cells were fractionated into their periplasmic (P), cytoplasmic (C) and insoluble (I) fraction using the osmotic/ cold shock method as described by Randall and Hardy,

1986 (Randall and Hardy, 1986). Cells were harvested and equalised to the same OD using the equation 10/ OD_{600} and were centrifuged at 1690 x *g*, 4 ºC for 10 mins. Samples were placed on ice and the supernatant was discarded. Then, cells were resuspended in 0.5 mL ice cold Buffer 1 (100 mM Tris-acetate pH 8.2, 500 mM sucrose, 5mM EDTA). 0.5 mL ice cold mqH_2O and 40 µL lysozyme (1 mg/ mL stock) were added. The eppendorf was inverted 3 times and cells were incubated on ice. After 5 mins 20 µL 1M $MgCl_2$ was added to stabilise the cytoplasmic membrane, and cells were centrifuged at 20800 x *g*, 4 ºC for 2 mins. The supernatant 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 750 µL ice cold Buffer 2 (50 mM Tris-acetate pH8.2, 250 mM sucrose, 10 mM $MgSO_4$) and centrifuged at 20800 x *g*, 4 ºC for 5 mins. The supernatant was discarded, and the pellet was resuspended in 750 µL 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 x 4) and then ultra-centrifuged at 264,360 x *g*, 4 ºC for 30 mins. This guarantees the separation of the membrane and insoluble material. The supernatant was collected as cytoplasmic fraction in a 1.5 mL Eppendorf without disturbing the cell pellet. The remaining supernatant was discarded, and the pellet was resuspended in 750 µL Buffer 3 (50 mM Tris-acetate pH 8.2, 2.5 mM EDTA). All fractions were stored at -20 ºC for further sample preparation.

2.4.4. Whole cell extract

For sample preparation without separating the cytoplasm, insoluble fraction, and periplasm, the cells were harvested and equalised to the same OD using the equation 10/ OD_{600} were centrifuged at 1690 x *g*, 4 ºC for 10 mins. The supernatant was poured

off and the pellet was diluted in 1 mL Buffer 1. To lyse the cells, they were sonicated on ice for 1 min.

2.5. Protein electrophoresis

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

For protein separation SDS-PAGE was used in the Bio-Rad Mini-PROTEAN® Tetra System. The gels are containing a separation and a stacking gel composed out of 15% acrylamide, 0.3% bis-acrylamide (37:5:1), 375 mM Tris-HCl pH 8.85, 0.1% SDS, 0.1% APS and 0.06% TEMED for the separation gel. For the stacking gel 5% acrylamide, 0.0375 % bis-acrylamide, 125 mM Tris-HCL pH 6.8, 0.001% SDS, 0.6% APS and 0.06% TEMED was used. The gels were cast in the Bio-Rad 0.75 gel plates. For running the samples on SDS-PAGE they were mixed with protein gel loading buffer (125 mM Tris-HCl pH 6.8, 20% glycerol, 4% SDS, 0.02% bromophenol blue and 5% β- mercaptoethanol) and heated for 5 mins at 95 ºC. Gels were run at 60 mA for 40 mins in protein gel running buffer (25 mM Tris, 192 mM Glycine and 0.1% SDS at pH 8.3).

2.5.2. Native gel/ CN-PAGE gel

For native protein separation PAGE was used in the Bio-Rad Mini-PROTEAN® Tetra System as described in 2.5.1. The gels are containing a separation and a stacking gel composed out of 15 % acrylamide, 0.3 % bis-acrylamide (37:5:1), 375 mM Tris-HCl pH 8.85, 0.1 % APS and 0.06 % TEMED for the separation gel. For the stacking gel

5 % acrylamide, 0.0375 % bis-acrylamide, 125 mM Tris-HCL pH 6.8, 0.6 % APS and 0.06 % TEMED was used. Gels were run at 60 mA for 40 mins in protein gel running buffer (25 mM Tris and 192 mM Glycine at pH 8.3).

2.5.3. Coomassie staining

For visualisation of the protein on SDS-PAGE the gels were incubated in 40 mL Coomassie stain (10 % acetic acid, 40 % methanol and 1 g/L Coomassie Brilliant Blue R-250 blue) for 1 hr at RT on a shaker. The stain was removed by washing the gel in D-stain (5 % ethanol, 7.5 % acetic acid) and replacing the D-stain 3 times over the period of 4 hours.

2.6. Protein detection

2.6.1. Western blotting

Proteins were also visualised using electrophoresis for wet-western blotting. Proteins were transferred onto a PVDF or nitrocellulose membrane depending on the tag to be visualised. Proteins were transferred using the Bio-Rad system and 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). The sandwich was placed in a cassette between the electrodes and the tank was filled with western blot transfer buffer. Proteins were transferred at 90 V for 1 hr.

2.6.2. Immunoblotting

For proteins immunoblotted on PVDF membrane, the membrane was stored overnight in 1 x PBS-T containing 2.5 % skimmed milk powder at 4ºC. On the next morning the membrane was washed 3 times for 5 mins in PBS-T at RT and then incubated for 1 hr in primary antibody. Followed by 3 more washing steps for 5 mins in PBS-T and incubation for 1 hr in the secondary antibody. The here used antibodies are listed in Table 11. The remaining secondary antibody is washed off in 6 washes for 5 mins in PBS-T. The membrane was treated using the ECL™ kit according to the manufacture's guide and the bands were visualised using BioRad Gel-doc chemiluminescence imager.

Table 11: **Antibodies used in this work.**

Antibody	Concentration	Source
Anti-TatA	1:2000	Laboratoy stock
Anti-strep	1:32000	IBA life sciences
Anti-GFP	1:1000	ThermoFisher
Anti-mouse HRP conjugate	1:5000	Promega
Anti-rabbit HRP conjugate	1:5000	Promega
Anti-strep HRP conjugate	1:32000	IBA life sciences
Anti-hGH	1:10000	Laboratoy stock (Alanen *et al.*, 2015)
Anti-hexastidine (C-term)	1:8000	Invitrogen

For proteins transferred onto nitrocellulose membrane, the membranes were incubated overnight at 4ºC in 20 mL PBS-T containing 3 % BSA. At the next morning, the membrane was washed with PBS-T 3 times for 5 mins on the rocker. Then, the membrane was incubated for 10 mins in 20 mL PBS-T with 20 µL Biotin blocking buffer. Afterwards the membrane was incubated in 20 mL PBS-T, 20 µL anti-strep antibody and 3 % BSA for 1 hr shaking. This is followed by 2 1 min wash steps in PBS-T and 2 1 min was steps in PBS. The membrane was treated using the ECLTM kit according to the manufacture's guide and the bands were visualised using BioRad Gel-doc chemiluminescence imager. All western blots were visualised for 60 seconds taking 100 images.

The here presented western blot images have been exposed between 5 and 10 seconds.

2.6.3. TMAO assay

For testing whether the Tat systems are functional a Trimethylamine N-Oxide assay was done. Cell cultures were grown and fractioned as described previously and LB-GT media was used. The samples were run on native gel as described in 2.5.2. For visualising the TMAO in the sample fractions the native gel was incubated for 15 mins in 100 mM potassium phosphate buffer containing 0.25 % w/v methyl violgen mixed with 10 mM sodium hydroxide buffer with 1 % sodium dithonite. After 15 mins the gel should be stained blue. To visualise the TMAO the gel is transferred into 100 mM potassium phosphate buffer containing 0.4 % w/v TMAO.

2.7. Protein purification

2.7.1. Ni-NTA Spin column purification of 6xhis tagged proteins under native conditions from *E. coli* cell lysate

For the bench top purification the spin columns have been equilibrated with 600 µL NPI-10 buffer. Then the columns were centrifuged at 890 x *g* for 2 minutes at 4˚C. Then 100 µL sample has been loaded onto the column and the columns was centrifuged for 5 min at 270 x *g* at 4˚C. and the flow through was collected. Then the column was twice washed using 600 µL NPI-20 buffer and the column was centrifuged at 890 x *g* for 2 min at 4˚C. Both wash fractions were collected. This was

followed by two elution steps using 300 µL NPI-500 buffer. The column was centrifuged at 890 x *g* for 2min at 4˚C and the elution fractions were collected. The buffers required for the bench top purification are presented in Table 12.

Table 12: **Buffers required for the Ni-NTA spin column purification**

Buffer	**Components**
NPI-10	6.9 g/L NaH_2PO_4, 17.5 g/L NaCl, 0.68 g/L $C_3H_4N_2$
NPI-20	6.9 g/L NaH_2PO_4, 175.4 g/L NaCl, 1.36 g/L $C_3H_4N_2$
NPI-500	6.9 g/L NaH_2PO_4, 175.4 g/L NaCl, 34 g/L $C_3H_4N_2$

2.8. Nitrite/ Nitrate assay, colorimetric

2.8.1. Nitrite/ Nitrate assay kit (Sigma Aldrich)

After samples were taken from the fermenter and spun down, 2.5 µL of the fermenter sample were topped up with the buffer solution to a final volume of 100 µL for the nitrite assay and a final volume of 80 µL for the nitrate assay. Standard curve was prepared according to the manufacturer's instruction. Each sample has been prepared in a triplicate.

For the nitrate assay 10 µL nitrate reductase solution and 10 µL enzyme cofactor solution were added to each well and the samples were mixed thoroughly and incubated at 25˚C, shaking. After 2 hours incubation 50 µL Griess reagent A were added to all samples including the nitrite ones followed by 5 minutes incubation. Then 50 µL Griess reagent B were added to each well and after 10 minutes incubation the absorbance (540nm) was measured.

Once the absorbance has been measured, the standard curve has been plotted and the relation of absorbance to nitrite/ nitrite and nitrate has been calculated. Then the concentration of nitrite (1) and nitrite and nitrate (2) has been calculated. From there the concentration of the nitrate has been determined (3).

(1) $\frac{A_{NO2}}{V_{NO2}} = C_{NO2}$

(2) $\frac{A_{[NO3-+NO2]}}{V_{[NO3-+NO2]}} = C_{[NO3-+NO2]}$

(3) $C_{NO3-} = C_{[NO3-+NO2]} - C_{NO2}$

2.9. Microscopy

2.9.1. Leica DMR microscope

For imaging of the *E. coli* cells a Leica DMR microscope with attached DFC9000 GT camera and a x63 lens was used.

3. *E. coli* genome modification with *tatABC Agrobacterium tumefaciens* operon and export of target proteins

3.1. Introduction

The Tat pathway exists in archaea, bacteria and plant chloroplasts and the different organisms utilise Tat to various ends. (Berks, 1996; Dilks *et al.*, 2003) It is known that the Tat pathway is capable of transporting fully folded proteins through the membrane into the periplasm (Matos *et al.*, 2014; Alanen *et al.*, 2015). Because of its unique ability of exporting correctly folded proteins and the ability to transport proteins up to a size of 150 kDa, Tat is interesting in regard to biopharmaceutical protein production (Jones *et al.*, 2016). The reason why certain proteins are only transported in a folded state using the Tat apparatus is poorly understood. However, Richter *et al*. show that introducing six hydrophobic residues to different Tat substrates (HiPIP and FG5) prevents translocation. This suggests that Tat pathway tends to reject proteins with hydrophobic residues exposed at their surface and that translocation is restricted to substrates that are adequately hydrophilic on the surface. A change of surface charge leads to the rejection of Tat substrates (Richter *et al.*, 2007). The proteins translocated by the Tat pathway are crucial for cell division, cell motility and processes like organophosphate metabolism which shows how important this pathway for bacteria is (Stanley, Findlay and Berks, 2001; Palmer and Berks, 2012).

Possible reasons for the transport of Tat substrates in a folded state could be that the tools (e.g. chaperones and redox factors) required for folding are not available in the periplasm and therefore the protein must be transported in a folded state (Natale, Brüser and Driessen, 2008). This hypothesis is supported by the fact that Tat transports almost exclusively cofactor containing redox proteins in most organisms (Ding and Christie, 2003; Dilks and Giménez, 2005). Some extremophile bacteria tend to transport more proteins over the Tat pathway which might be related to challenges in

folding the proteins in the periplasm due to, for example, high extracellular salt concentrations (Palmer and Berks, 2012; Green and Mecsas, 2015). This might explain why more proteins are transported over the Tat pathway in, for example, *Haloferax volcanii* compared to *E. coli* (Rose *et al.*, 2002). However, overcoming its proofreading mechanism is challenging (Mendel and Robinson, 2007; Jones *et al.*, 2016).

In this study we aim to modify the *E. coli* genome to investigate if an alternative *tat* operon is capable of expressing a larger variety of proteins and to overcome the proofreading mechanism of *E. coli* Tat and test the export ability. We chose to work with the *tat* operon of *Agrobacterium tumefaciens*. It has been shown that the *A. tumefaciens* TatA, TatB, and TatC proteins are similarly configured compared to the *E. coli* proteins (Oates *et al.*, 2003). TatA and TatB each transverse the membrane once, while TatC transverses the membrane six times. However, *A. tumefaciens tat* genes are not closely related to *E. coli tat* genes (see Table 13) and an alignment is shown in the appendix (Ding and Christie, 2003).

Table 13: **Comparison between the *A. tumefaciens* Tat proteins and the *E. coli* Tat proteins.** Whilst the proteins show a high similarity in the way they are embedded in the membrane the actual amino acid sequence varies considerably. Ding and Christie showed in 2003 that TatA shows the highest similarity with only 42%. TatB shows 27% identity and TatC 35%.

	TatA	TatB	TatC
Amino acid length *A. tumefaciens*	68	239	267
Amino acid length *E. coli*	89	171	258
Amino acid identity in %	44	31	35

The questions addressed here are whether an alternative *tat* operon integrated into the *E. coli* genome can complement the Δ*tat* specific chaining phenotype. Furthermore, we were testing the modified strain for its ability to export Tat native substrates and we were investigating which heterologous substrates the system can export additionally.

3.2. Results

3.2.1. Expression of Tat proteins on a secondary plasmid in Δ*tat* is able to rescue the filamentous phenotype but does not improve the growth

In the first experiment, we evaluated which *tat* operons could be possible targets for an insertion onto the *E. coli* chromosome. We transformed a plasmid containing *tat* operon sequences into a Δ*tat* strain on plasmid pWM and compared the growth and phenotype. pWM is a derivative of pLysSBAD with a Δ*Lys* deletion. In this plasmid the *tat* operon is induced with L-arabinose. An empty plasmid would have been useful but a version of the pWM lacking the *tat* operon was not available for this work. However, Δ*tat* has been transformed with a variety of plasmids in previous studies, none of which complemented the phenotype (Frain *et al.*, 2017). Various operons were investigated, such as the *tat* operons of *Agrobacterium tumefaciens, Mesorhizobium loti,* and *Synechocystis sp.*. These were selected because they contain a *tat* operon on the genome and show varying degrees of similarity to the *E. coli* system. However, in the work presented here we focus on *A. tumefaciens* TatABC.

Cells lacking a Tat system are not able to export proteins that are vital for the ability to separate during cell division. Therefore, these cells show a 'chaining' phenotype, Figure 13B. Under the microscope we could observe that cells expressing TatABC of *A. tumefaciens* (Figure 13C and Figure 13D) are capable of rescuing the phenotype of the Δ*tat* strain and the cells are showing a similar phenotype to MC4100 wildtype (Figure 13A). Cells not expressing the TatABC *A. tumefaciens tat* operon show the same phenotype as Δ*tat* cells not containing a plasmid, Figure 13C. The phenotype changes are only confirmed after the synthesis of *A. tumefaciens* TatABC once induction of expression has occurred.

As the *A. tumefaciens tat* operon is able to improve the phenotype, we expect the cells to show a similar growth pattern as the wildtype.

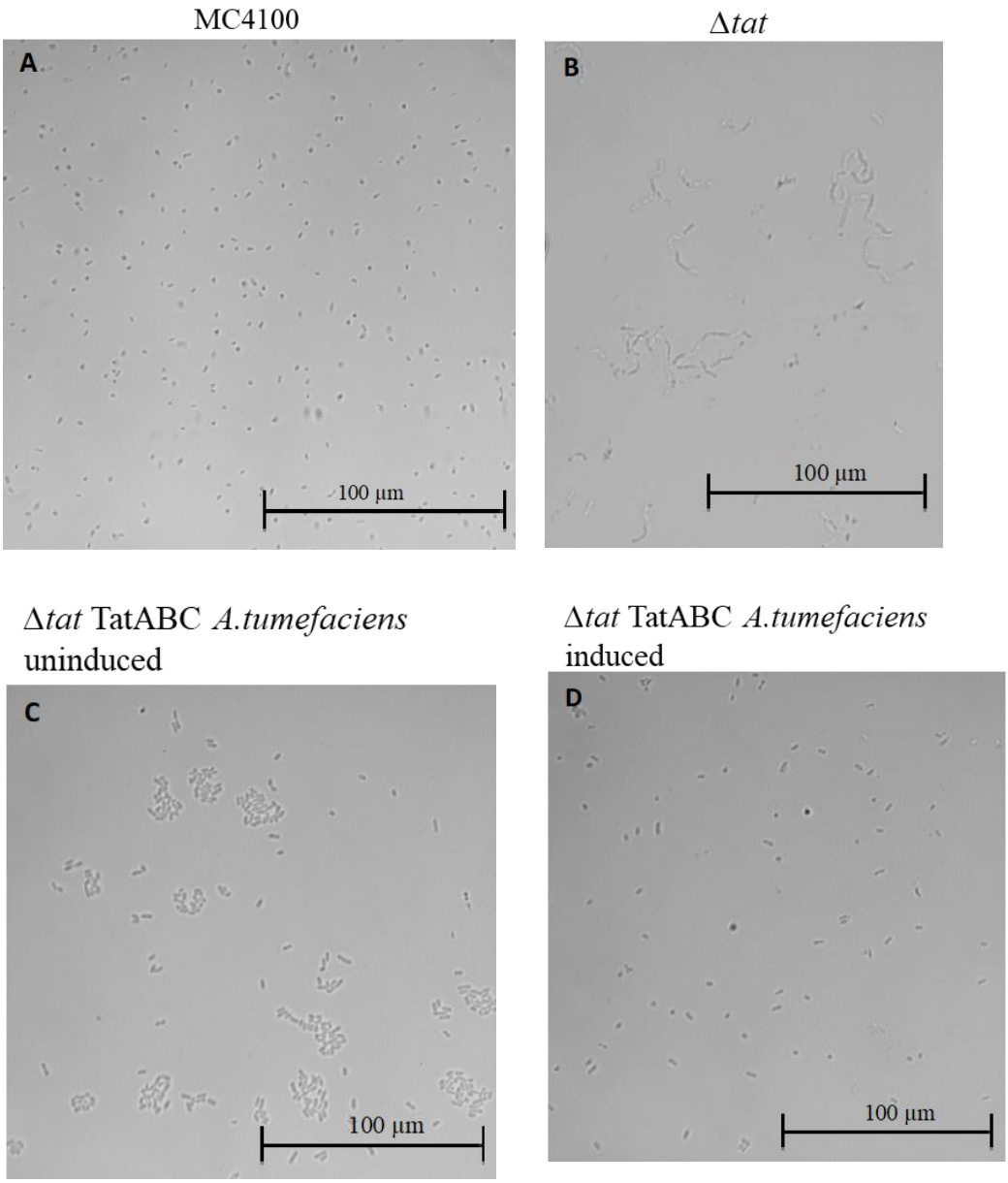


Figure 13: ***A. tumefaciens* TatABC complements the Δ*tat* phenotype of *E. coli*.** The figure shows microscope (Leica DMR microscope) images of the different phenotypes of **A** MC4100, **B** Δ*tat* which lacks proteins important for the separation of the cell membrane and shows a 'chaining' phenotype, **C** of Δ*tat* cells expressing *A. tumefaciens* TatABC from a plasmid pre-induction, and **D** of Δ*tat* cells expressing *A. tumefaciens* TatABC following induction. Cells were grown in 10mL LB in shake flask and visualized on a Leica DMR microscope. The *tat* operon was expressed of the plasmid (pWM) and expression was induced with 20µM L-arabinose.

Furthermore, we transformed a plasmid containing the *A. tumefaciens tat* operon sequence into a Δ*tat* strain and compared the growth and phenotype.

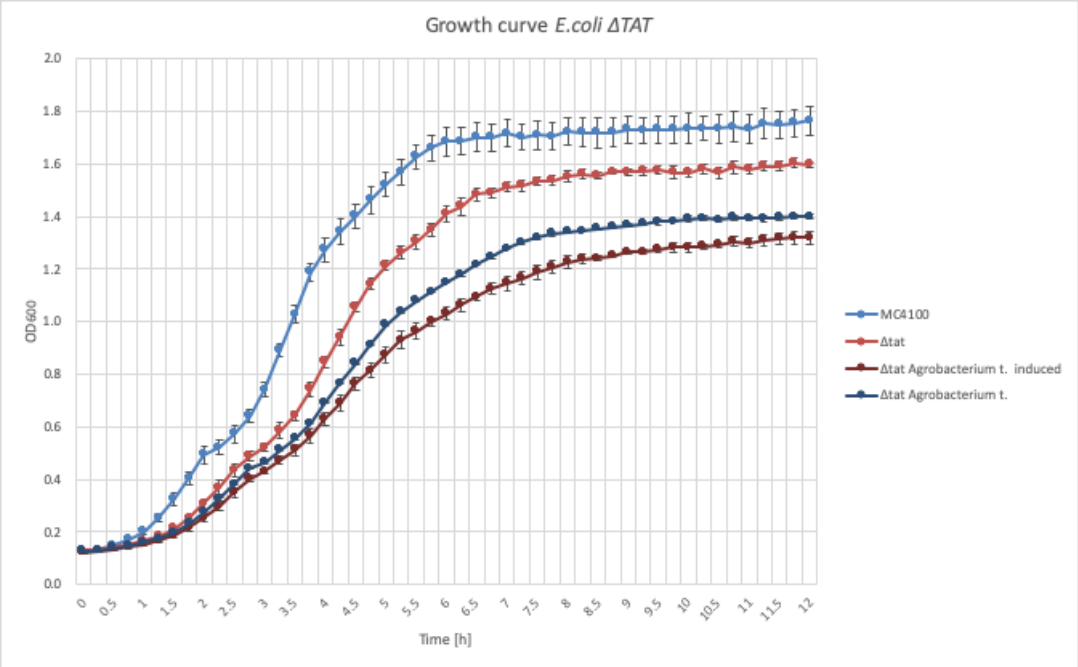


Figure 14: **Comparison of the growth of Δ*tat* expressing a plasmid containing the *Agrobacterium tumefaciens tat* operon to MC4100 wildtype and Δ*tat*.** The growth of MC4100 wildtype, Δ*tat* cells, and Δ*tat* cells containing the *A. tumefaciens* TatABC plasmid pWM (with and without induction) were monitored in a plate reader in 1mL LB cultures. Time points were taken every 10 minutes for 12 hours. The values presented here are averaged values from quadruplicates. One sample of Δ*tat* cells containing the *A. tumefaciens* TatABC pWM (pLysSBAD Δ*Lys*) plasmid was induced with 20µM L-arabinose after 2 hours.

Agrobacterium tumefaciens tatABC is not able to improve growth when expressed from a plasmid instead it decreases the growth rate as shown in Figure 14. An empty vector control could have been used as a control in this experiment. However, only

when the plasmid containing *A. tumefaciens tatABC* was induced were the cells complemented. Cells without a functioning Tat system are prone to lysis and show a reduced growth rate. Here the observed decrease in growth rate may be related to the combination of stress that the cells experience due to the *tat* deletion affecting their viability and the expression of a foreign Tat system. Also shown in Figure 14, the plasmid inhibits growth and the cells containing *A. tumefaciens tatABC* had a lower yield than the Δ*tat* strain. The decreased cell growth yield could be related to overexpression of the foreign Tat systems and challenges integrating this system into the cytoplasmic membrane. Here we used the pWM plasmid which is a pLysSBAD derivative with a Δ*Lys* deletion induced with 20μM L-Arabinose.

3.2.2. *A. tumefaciens* TatABC is successfully inserted on the chromosome and can improve the growth

The Δ*tat* strain shows changes in the phenotype when a *tat* operon is expressed from a plasmid, but the growth does not improve. It had been planned to test the export ability of foreign Tat systems in *E. coli*. However, further expression experiments will be challenging as two-plasmid expression increases cell stress on already sick cells and in the previous experiment the cells expressing the *tat* operon showed a decreased growth rate. Therefore, we have chosen gene doctoring as a method to replace the *E. coli* TatABC genes on the chromosome of the strain MC4100.

Gene doctoring consists of using the λ red genes and a transfer plasmid called pDoc-K. The pDoc-K plasmid contains the *sacB* gene which is a counterselection gene that makes the cells sensitive to sucrose, an ampR as selection markers, a kanR as selection

marker, and two Sce1 restriction sites flanking the target DNA which are cut in the gene doctoring process (Figure 15).

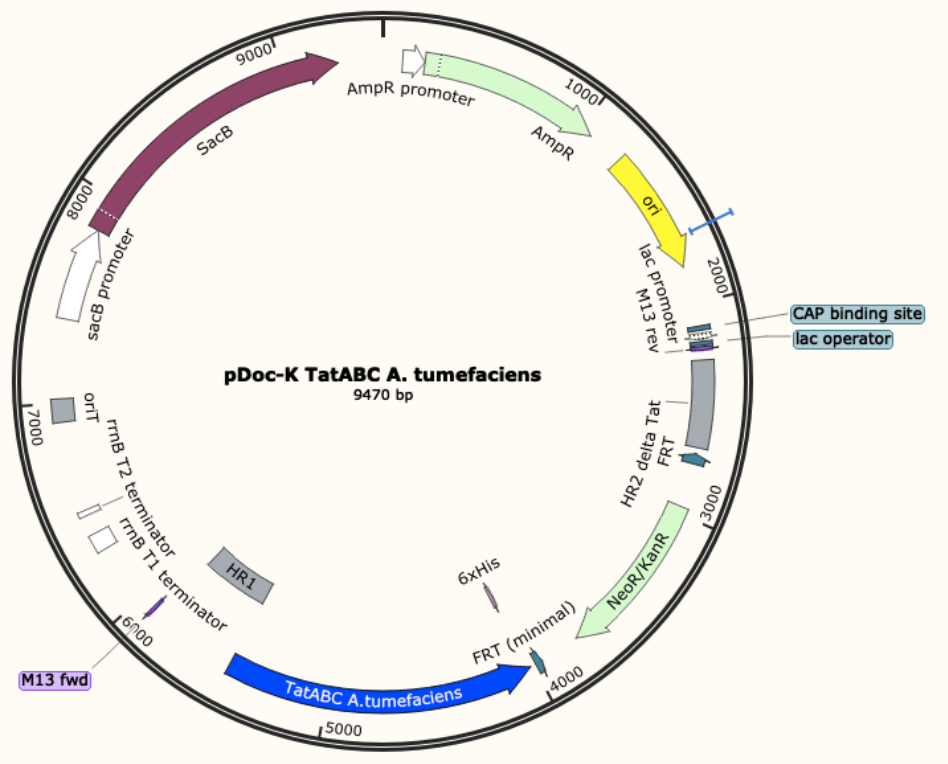


Figure 15: Plasmid map showing of the features of the gene doctoring plasmid pDoc-K including the *tatABC* operon of *A. tumefaciens*.

To apply the pDoc-K vector for replacing the genes encoding *E. coli* TatABC with the genes encoding *Agrobacterium tumefaciens* TatABC, two homology regions (HR) have to be designed via gene synthesis. Homology regions have to align to the flanking gene sequence where the deletion or insertion is planned. The HR regions are about 400 bp long and align to the UbiB and TatD region next to the *E. coli* TatABC genes. The *A. tumefaciens tatABC* operon had been codon optimised and the sequence is shown in the appendix on page 216. To express the *A. tumefaciens tatABC* operon the native promoter on the genome has been used.

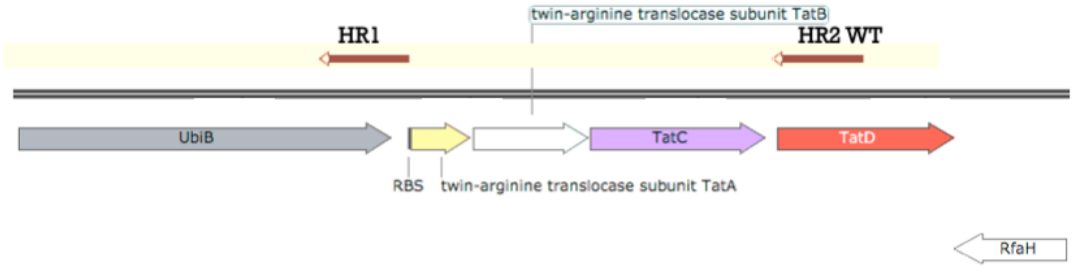


Figure 16: **Location of the homology regions (HR1 and HR2 on the genome).** For the gene doctoring process homology regions have to be designed which align with the target site on the chromosome. The homology regions are 400 bp long and in this case align to the *ubiB* and *tatD* genes. Throughout the process the sequence between the homology regions is going to be replaced.

After the pDoc-K plasmid containing the HR regions, shown in Figure 16 and the operon of interest was clones, the gene doctoring process was performed as shown in Figure 17.

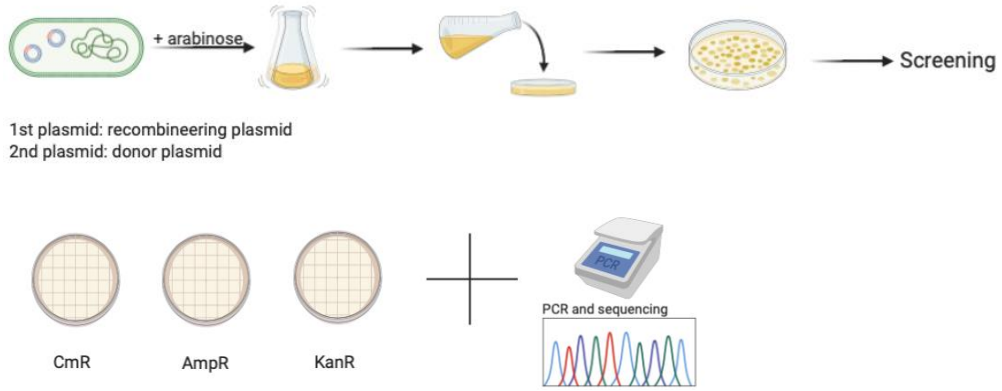


Figure 17: **Flow diagram of the gene doctoring process.** In the first step, two recombineering plasmids are transformed into the cells. The 1st plasmid contains the λ red genes for the recombineering process and the 2nd plasmid is the donor plasmid of the operon of interest. Once the plasmid is induced with arabinose, the λ red genes are expressed. After a 3-hour incubation the cells are plated out on agarose plates. To confirm that the operon of interest has been introduced into the genome, the colonies are screened for the loss of the two plasmids (ampicillin resistance and chloramphenicol resistance) and the insertion of operon of interest by kanamycin resistance and PCR.

The cells were plated out and were re-streaked to confirm a loss of the gene doctoring plasmid and a positive insertion on the chromosome. Cells growing on agar plates with kanamycin but not on ampicillin and chloramphenicol plates are a first confirmation of a successful gene doctoring process. The kanamycin resistance is introduced to the genome together with the *tat* operon. The recombineering plasmid confers resistance to chloramphenicol and the donor plasmid confers resistance ampicillin. Furthermore, positive clones were confirmed by creating a PCR product containing the genomic region of interest and showing on an agarose gel that there was an insertion (Figure 18). Clones with the right size of insert were sent for sequencing by GATC. Two

colonies were confirmed to contain the *A. tumefaciens tat* operon on the genome *E. coli* MC4100 genome.

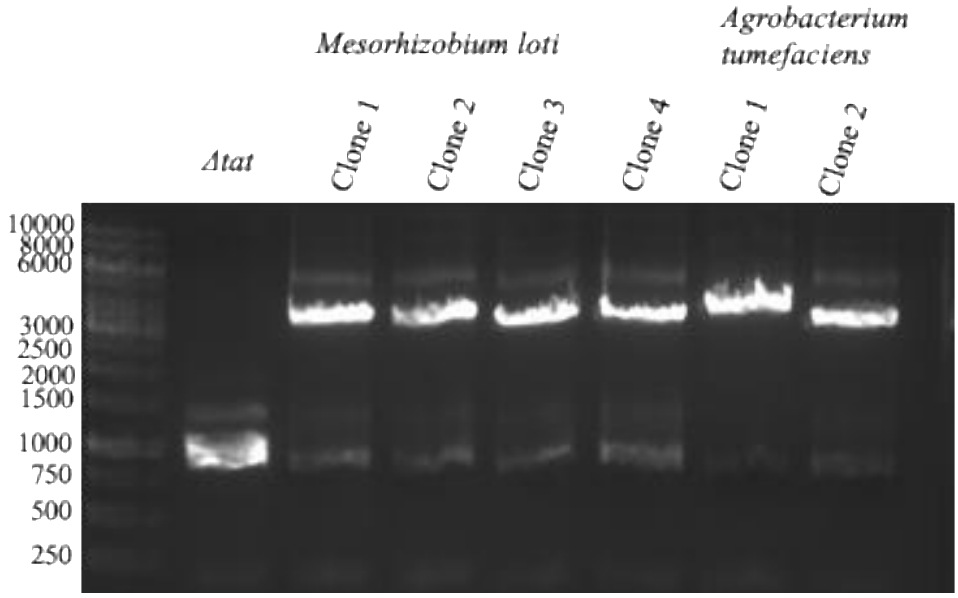


Figure 18: **DNA insert confirmation of the MC4100 colonies.** A colony PCR was performed on the colonies that possibly contain the new operon on the chromosome. The expected size of the Δ*tat* sequence as a negative control is 1191bp and the expected size of the *tat* operons inserted into the *E. coli* genome are expected at a size of 3653bp for the *A. tumefaciens* TatABC. Furthermore, shown here is the work on *Mesorhizobium loti* which was not continued.

Furthermore, the PCR products have been sequenced with GATC sequencing and the insertion of *A. tumefaciens tatABC* was be confirmed. The engineered *E. coli* MC4100 mutants contain next to the *A. tumefaciens* sequence a kanamycin cassette on the genome. The same process was undertaken using the *tatABC* sequence of *Mesorhizobium loti* however, a positive insertion of the operon could not be confirmed on the genome by sequencing. Therefore, the work on *Mesorhizobium loti* was not continued.

With this method, we were able to confirm two colonies of MC4100 containing *A. tumefaciens tatABC* operon on the genome (Clone 1 and Clone 2). Further work has been continued with both colonies. The clones have been sequenced starting in the homology region 1 until the homology region 2 and their sequences are shown in the appendix (page 217 and page 218). The in the appendix shown sequence corresponds to the TatABC region. The nucleotide sequence of Clone 1 shows one point mutation which is affecting the amino acid sequence, changing a lysine to an arginine in the TatB sequence in position 41. The sequence of Clone 2 shows no mutations.

The difference in the Clone 1 nucleotide sequence could have either occurred during the gene doctoring process or it could a be point mutation that occurred during the PCR use to generate the amplicon. In the next step, we investigated the functionality of the introduced *tat* in Clone 1 and Clone 2.

A strain with a non-functional Tat pathway lacks proteins important for the separation of the cell membrane and shows a ‘chaining’ or filamentous phenotype. This was used as first assessment of a functional insertion of the Tat genes onto the chromosome.

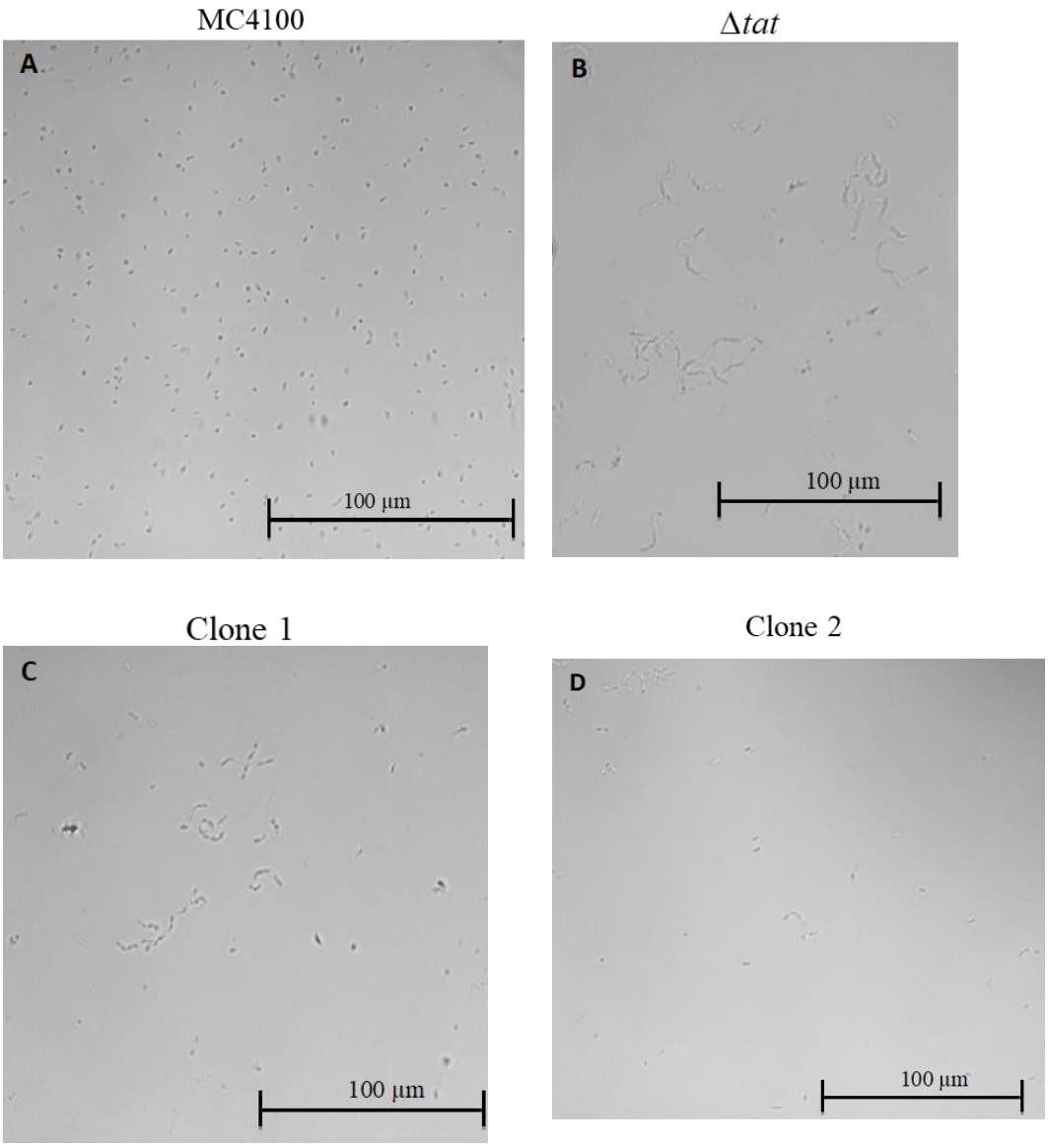


Figure 19: **Comparison of the different phenotypes of the new clones, wildtype cells, and Δ*tat*.** The figure shows microscope (Leica DMR microscope) images of the different phenotypes of **A** MC4100, **B** Δ*tat* which lacks proteins important for the separation of the cell membrane and shows a 'chaining' phenotype, **C** of MC4100 Clone 1, and **D** MC4100 Clone 2. Cells were grown in 10mL LB in shake flask and visualized on a Leica DMR microscope.

Both clones show a phenotype that is different to that of the Δ*tat* strain, Figure 19. Clone 1 cells seem to accumulate as clumps which suggest a defect in the Tat operon.

However, it shows no similarity to the filamentous chaining phenotype of the Δ*tat* strain. The morphology of Clone 2 is similar in appearance to that of MC4100.

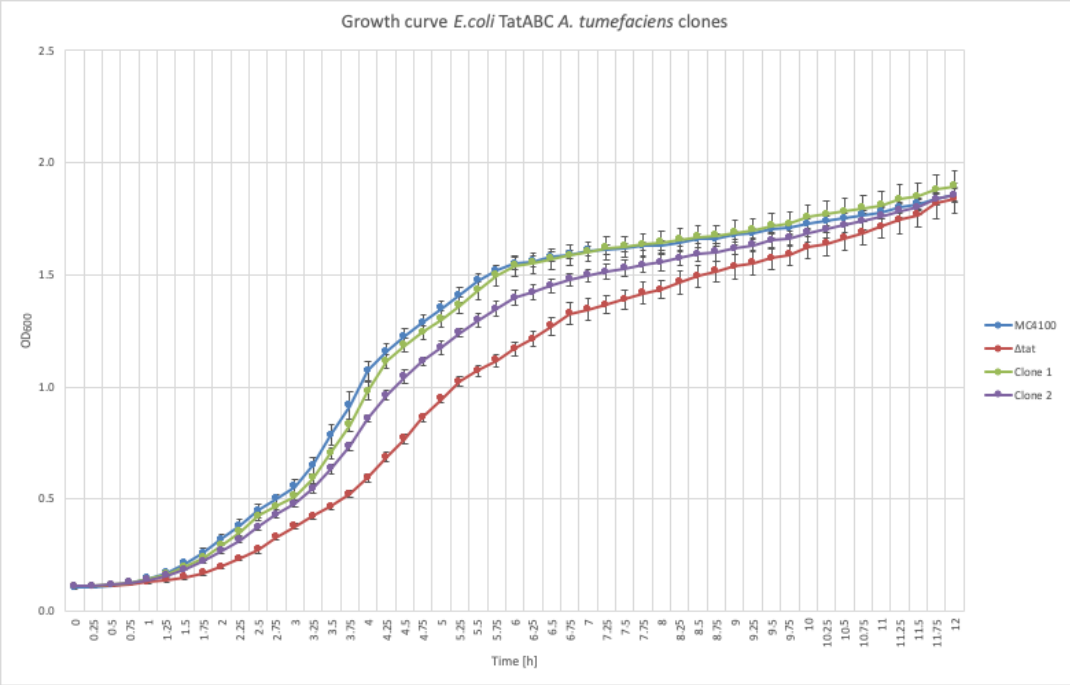


Figure 20: **Comparison of the growth of the two Tat clones to MC4100 and Δ*tat***. The growth of the MC4100 wildtype, Δ*tat* strain, and Clones 1 and 2 of the MC4100 *A. tumefaciens* TatABC strain was monitored in a plate reader in 1mL LB cultures. OD readings were taken every 10 minutes for 12 hours. The values presented here are averaged values from quadruplicates.

The insertion of *A. tumefaciens* TatABC onto chromosome of *E. coli* can improve the growth behaviour and shows improved growth compared to Δ*tat*, as shown in Figure 20.

3.2.3. MC4100 expressing the *Agrobacterium tumefaciens tatABC* operon is able to export the native Tat substrate TorA into the periplasm

In *E. coli* TatA, TatB, and TatC proteins have to be functional to translocate their cargo proteins into the periplasm (Wexler *et al.*, 2000). If trimethylamine N-*oxide* (TMAO) is provided in the medium TMAO reductase (TorA) activity will be found in the cytoplasmic, insoluble, and periplasmic fractions (Oates *et al.*, 2003). However, cells with insufficient Tat components are not able to export TorA into the periplasm and TorA will only occur in the cytoplasmic and insoluble fractions. We chose the TMAO reductase (TorA) assay, to confirm that *A. tumefaciens* TatABC is able to export an *E. coli* native Tat substrate. Thus, being able to fulfil *E. coli* Tat roles.

After the culture was grown for several hours in a medium containing TMAO (LB-GT medium) the cells were separated into cytoplasmic (C), insoluble (I), and periplasmic (P) fractions as described in 2.4.3 *E. coli* cell fractionation and were loaded onto a native CN-PAGE gel. The gel can be stained for TorA activity and the TorA appears as a white band. As Figure 21 shows, the two clones containing *A. tumefaciens* TatABC on the chromosome are capable of exporting TorA into the periplasm when TMAO is provided in the medium. Δ*tat* was chosen as a negative control because it lacks the necessary Tat system to export TorA. TorA is visible in the cytoplasmic and insoluble fraction in Δ*tat* but not in the periplasmic fraction, as expected. As shown in Figure 21A, the periplasmic export of Clone 1 is similar to that of wildtype cells. It is not clear why TorA migrates as two bands in some of the samples.

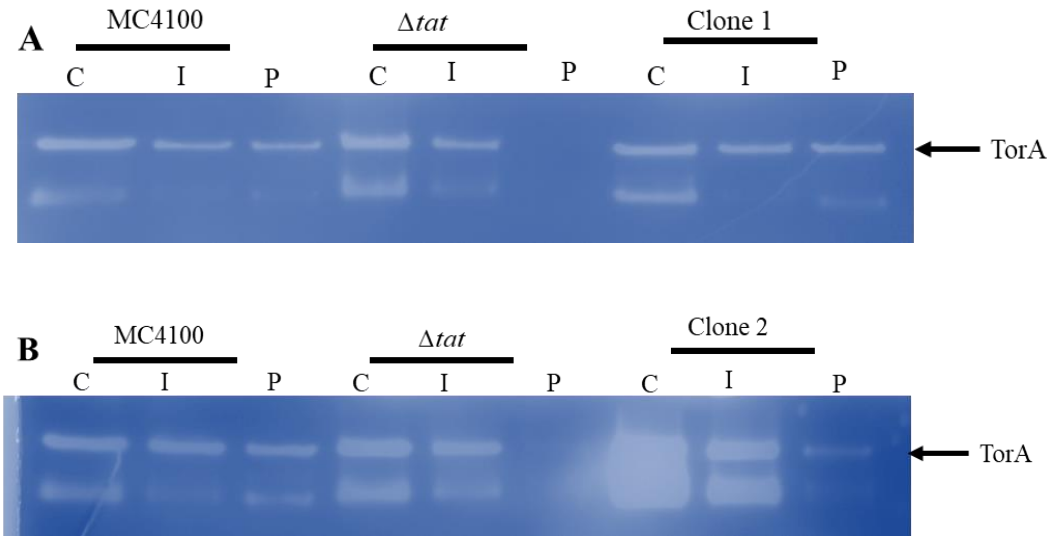


Figure 21: **Clones expressing the *A. tumefaciens tatABC* operon are capable of exporting a native Tat substrate.** Cells were grown in LB medium containing TMAO (LB-GT medium) at 37˚C in 50mL shake flask cultures. Cells were separated in cytoplasmic (C), insoluble (I), and periplasmic (P) fraction. The TorA (80 kDA, arrowed) is a white band that will appear in the periplasm in a functional Tat system. Compared to Δ*tat* there is a clear band of TMAO in the periplasm showing that TatABC is able to export a native protein. **A** shows the CN-PAGE of the first clone (Clone1) and **B** shows the CN-PAGE of the second clone (Clone 2). The gel is a representative of two gels and 10 µL sample have been loaded in each well.

Additionally, we aimed to confirm *A. tumefaciens* TatA integration into the membrane. Keeping in mind, that *E. coli* TatA shares 42% similarity with *A. tumefaciens* TatA, we have tried to detect TatA in the membrane of the new clones with a *E. coli* TatA specific antibody previously used in the lab. *E. coli* TatA has a size of roughly 10 kDa but in previous studies *E. coli* TatA was appearing at around 18 kDa size in SDS-PAGE gels (Gouffi *et al.*, 2004). The cells were grown in shake

flasks cultures in terrific broth (TB) medium and, as previously, separated into the three fractions (cytoplasm (C), insoluble (I), and periplasm (P)).

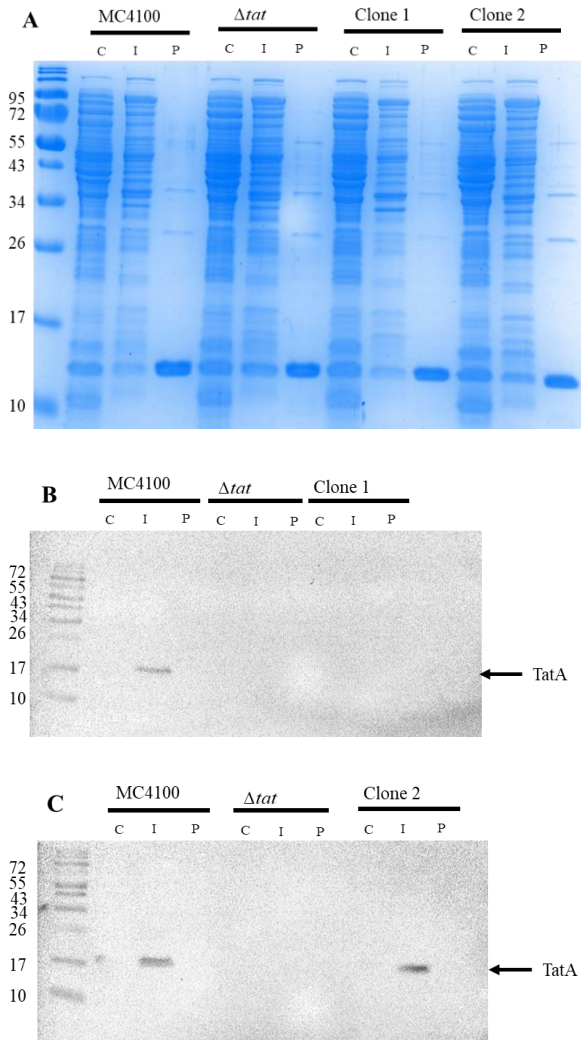


Figure 22: **TatA signal in MC4100, Δ*tat*, and the new *A. tumefaciens* TatABC.** The cells were grown in shake flask at 30°C in Terrific Broth (TB) medium and then were fractionated into cytoplasmic, insoluble, and periplasmic fractions. TatA (arrowed) signal is expected in the insoluble fraction. **A** shows the Coomassie stained SDS-PAGE gel, **B** shows the western blot containing Clone 1, and **C** shows the western blot containing Clone 2. The expected size is 9.8 kDa for *E. coli* TatA and 7.5 kDa for *A. tumefaciens* TatA. Samples were blotted using an anti-TatA antibody. The gel is a representative of two gels

and 10 µL sample have been loaded in each well. The exposure time of the western blots is 19 seconds.

In Figure 22B it was not possible to detect a TatA signal in the insoluble fraction of Clone 1 although as expected, it was possible to detect *E. coli* TatA in the insoluble fraction of the wildtype cells. Figure 22C shows that TatA from Clone 2 is detectable with an *E. coli* specific TatA antibody. This shows that there is a difference in the activity of the integrated Tat system in clones 1 and 2. The insertions have been sequenced and one point mutation in the *A. tumefaciens* TatABC region occurred in Clone 1 (alignment shown in the appendix page 219 and page 220). However, as the upstream promoter region has not been sequenced, it is possible that TatA is expressed at different levels in the two clones.

3.2.4. MC4100 clones expressing the *A. tumefaciens* TatABC is not capable of exporting human growth hormone into the periplasm

We are aiming to investigate the export ability and limitations of the new strain designed in this study. Are there any proteins that the *A. tumefaciens* TatABC clones are able to export that *E. coli* rejects? To answer the question, we started investigating if proteins that are exported by *E. coli* are also exported in the new strain. One protein that previously was exported successfully by the *E. coli* Tat system is human growth hormone (hGH) (Alanen *et al.*, 2015; Browning *et al.*, 2017).

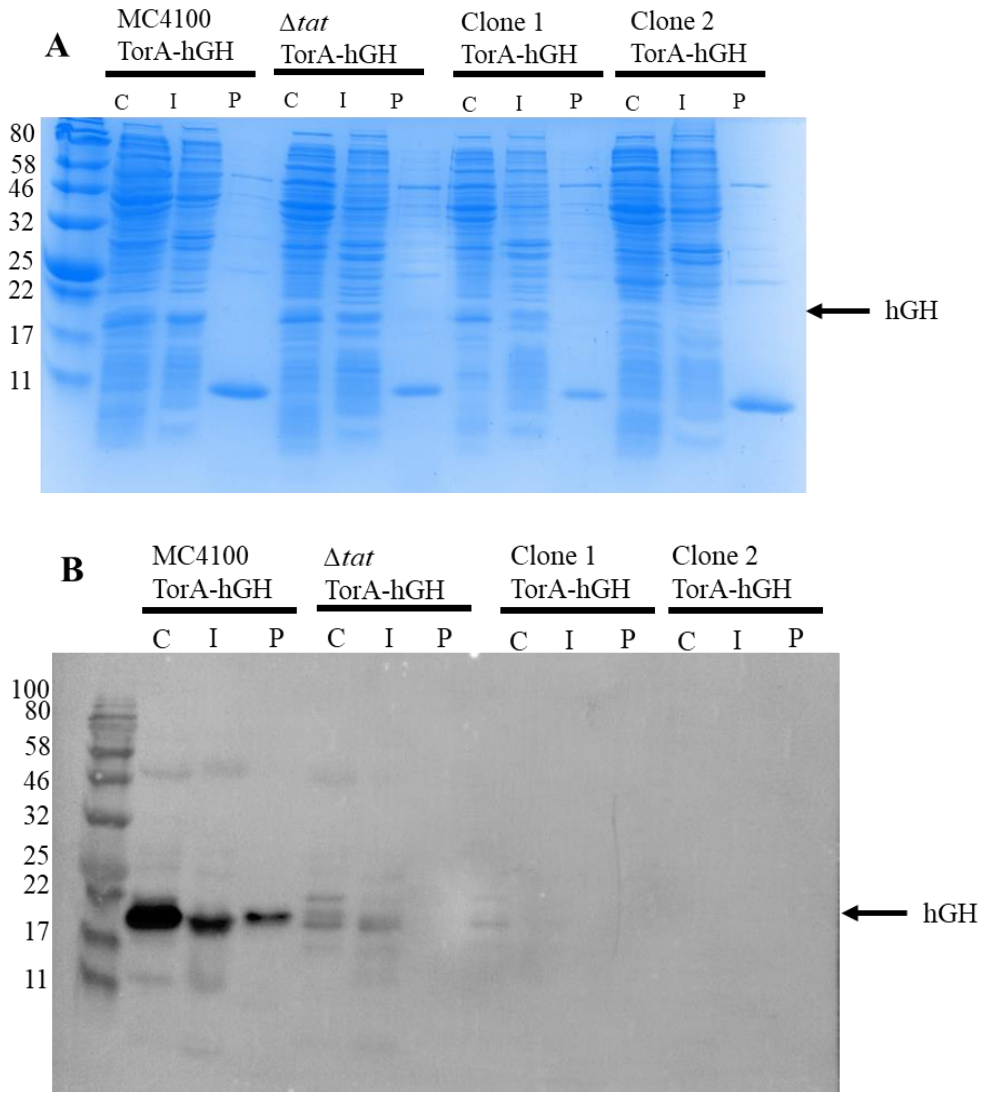


Figure 23: **The detection of hGH in cytoplasmic, insoluble, and periplasmic fractions following expression with the Tat specific signal peptide TorA in strains encoding the *E. coli* and *A. tumefaciens tat* operon.** The plasmid containing hGH (21 kDa, arrowed) was induced with 1 mM IPTG and was expressed in shake flask at 30°C in TMAO containing Luria Broth (LB) medium. The expression was 3 hours post-induction. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot of the fractions. The western blot was immunoblotted using an hGH-specific antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 5.8 seconds.

The experiment shows that MC4100 is capable of exporting hGH into the periplasm as expected and in the periplasmic fraction of Δ*tat* there is no hGH signal. Clones 1 and 2 do not show any traces of hGH on the western blot and it appears that hGH was degraded as only a faint hGH signal is visible in the cytoplasmic fraction of Clone 1 (Figure 23B). The two *A. tumefaciens* TatABC clones are not able to export hGH into the periplasm when expressed with a TorA signal peptide (Figure 23B). The native TorA is still present in the cell. As the wildtype strain is expressing the protein, it is likely that the protein in Clone 1 and Clone 2 is expressed but degraded. hGH has been in previous work shown to be susceptible to degradation (Browning *et al.*, 2017). Comparing Clone 1 and Clone 2, a faint band can be seen in the cytoplasm of Clone 1 that does not appear in Clone 2.

The same cells expressing hGH were used to perform the TMAO assay, as previously described, to evaluate whether native Tat substrates are expressed during the protein expression.

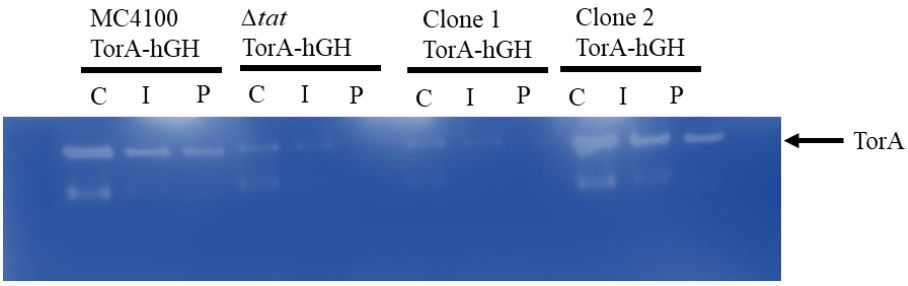


Figure 24: **TorA activity during TorA-SP-hGH expression.** TorA-SP-hGH was induced with 1 mM IPTG and was expressed at 30°C in TMAO containing Luria Broth (LB) medium. To show whether native substrate activity is still occuring when expressing a non-native protein via Tat, we have also taken samples for a TMAO assay throughout protein expression. Here shown is a CN-PAGE gel. The gel is a representative of two gels and 10 µL sample have been loaded in each well.

Clone 2, which was unable to export hGH to the periplasm (Figure 23B), was still able export TorA to the periplasm (Figure 24). In contrast Clone 1, in which a small amount of hGH was detected in the cytoplasm (Figure 23B) showed very little periplasmic TorA. It is possible that in this strain high level expression of TorA-SP-hGH interfered with the export of native TorA.

3.2.5. MC4100 *A. tumefaciens* TatABC is not capable of exporting single chain variable fragments into the periplasm

To further test the abilities of MC4100 clones containing *A. tumefaciens* TatABC on the chromosome we tested different single chain variable fragments (scFv's). We used an expression plasmid containing the TorA signal peptide linked to a single chain antibody variable fragment (TorA-SP-scFv) and expressed the plasmid in the two clones. The export was compared with MC4100 wildtype and Δ*tat*. These constructs have been shown to be exported by Tat in previous work (Matos *et al.*, 2014).

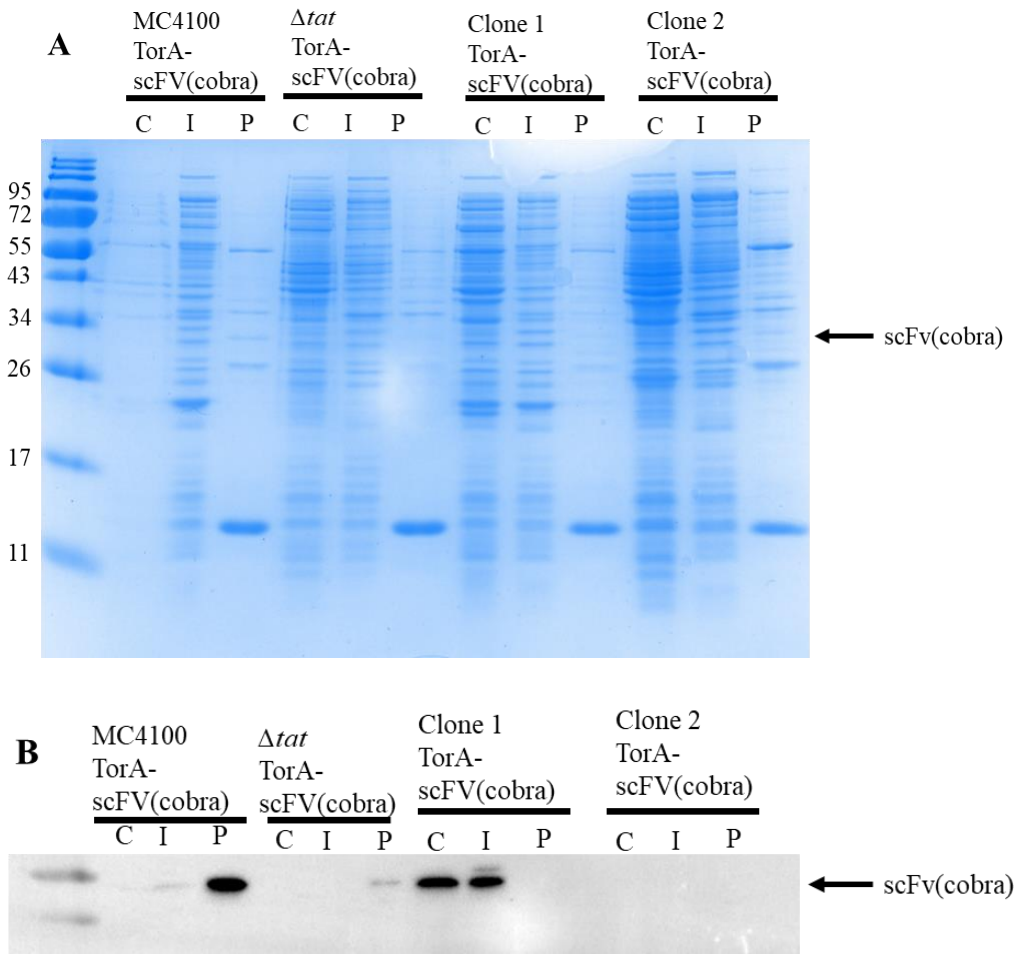


Figure 25: **Expression of scFv(cobra) with the Tat specific signal peptide TorA in the new strains.** Cells were grown in 50mL LB shake flask cultures at 30˚C. scFv(cobra) (29.5 kDa, arrowed) was induced with 1 mM IPTG after reaching an OD_{600} of 0.4 and cells were expressed for 3-hours post induction. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot of the fractions. The western blot was immunoblotted using a C-terminal anti-his antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 5.8 seconds.

Comparing the fractions of Clone 1 and Clone 2 on the Coomassie stained SDS-PAGE gel shows that the periplasm of Clone 2 has a higher-level cross-contamination with cytoplasmic proteins than Clone 1 (Figure 25A). Cytoplasmic contamination of

periplasmic samples is indicated by an appearance in the periplasmic fraction of multiple high molecular bands, (>55kDa). Comparing the periplasmic fractions in the Coomassie stained SDS gel, there are slight difference seen between the different strains. MC4100 exports scFv(cobra) into the periplasm as it can be seen due to the vast majority of scFv being found in the periplasmic fraction in the western blot. The two clones are not able to export TorA-SP-scFv(cobra) into the periplasm as shown in Figure 25B. Interestingly, Clone 1 expresses scFv(cobra) but is not exporting it into the periplasm while Clone 2 does not show any expression. The protein might also be degraded as scFvs are likely to be degraded when they are not exported (Alanen *et al.*, 2015).

We furthermore investigated a different scFv called scFv(M) which has been previously reported to be exported by the Tat pathway (Jones *et al.*, 2016).

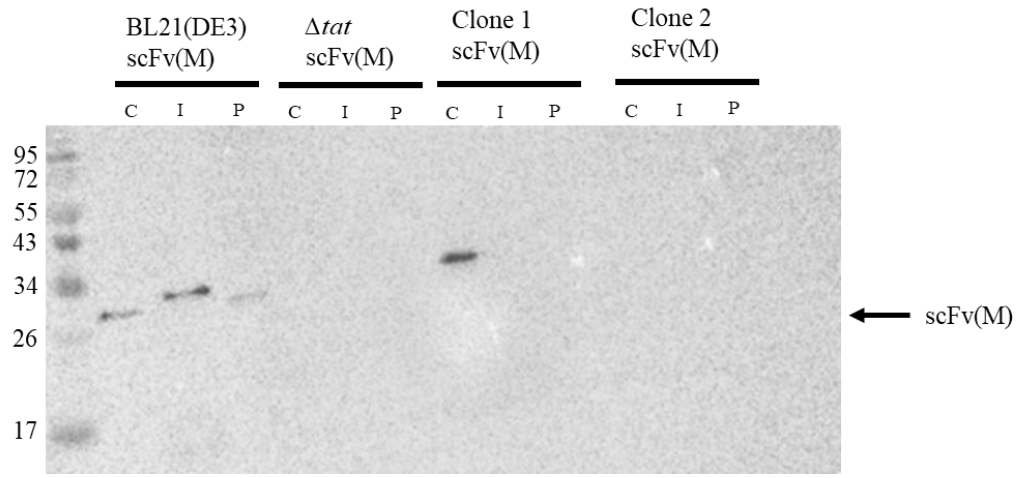


Figure 26: **Expression of single chain variable fragment (scFv) variant with the Tat specific signal peptide TorA in the new strains.** Cells were grown in 50mL LB shake flask cultures at 30˚C. After reaching an OD_{600} of 0.4 the plasmid containing scFv(M) (28 kDa, arrowed) was induced with 1 mM IPTG and was expressed for 3 hours. Cell samples were fractionated and immunoblotted. The western blot was immunoblotted using a C-terminal anti-his antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 5.8 seconds.

The two clones are also not able to export TorA-SP-scFv(M) into the periplasm (Figure 26). Clone 1 appears to express TorA-SP-scFv(M) but the signal is only visible in the cytoplasmic fraction of the cells. Clone 2, however, does not show any signs of TorA-SP-scFv(M) expression.

3.2.6. MC4100 *A. tumefaciens* TatABC is not capable of exporting GFP into the periplasm

Furthermore, we investigated whether TorA-SP-GFP can be exported in the new strains. GFP has been successfully exported by the Tat pathway in *E. coli* wildtype cells. (Barrett *et al.*, 2003; Bernhardt and De Boer, 2003).

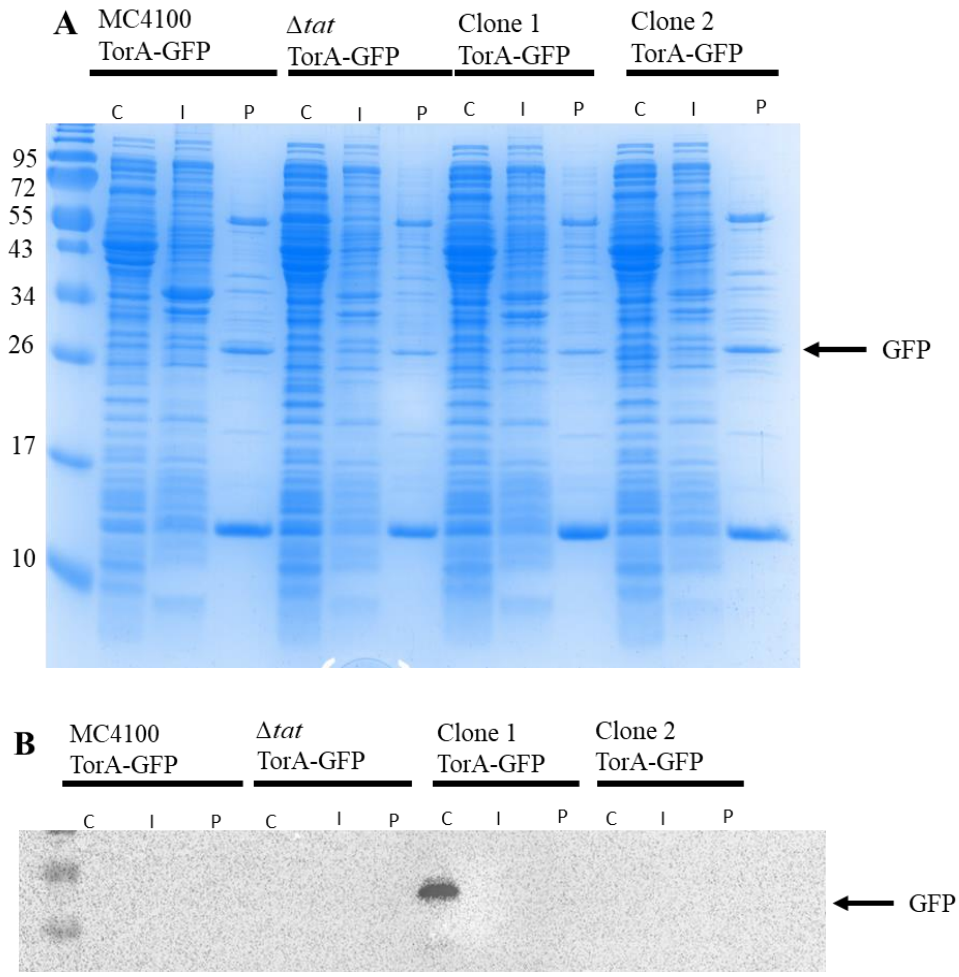


Figure 27: **Expression of TorA-SP-GFP in the new strains compared to MC4100 and Δ*tat*.** Cells were grown in 50mL LB shake flask cultures at 30˚C. The plasmid containing GFP (27 kDa, arrowed) was induced with 1 mM IPTG after reaching an OD_{600} of 0.4 and was expressed for 3-hours. Samples were fractionated. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot of the fractions. The western blot was immunoblotted using a C-terminal anti-his antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 5 seconds.

In Clone 1 TorA-SP-GFP was expressed but was not exported into the periplasm while TorA-SP-GFP was not expressed in Clone 2 (Figure 27). However, MC4100 did not

express TorA-SP-GFP in this experiment. Therefore, we repeated the expression in BL21 DE3 cells to show that the expression of this plasmid is possible. In Figure 28A and Figure 28B we can show that TorA-SP-GFP is expressed in BL21 DE3 cells.

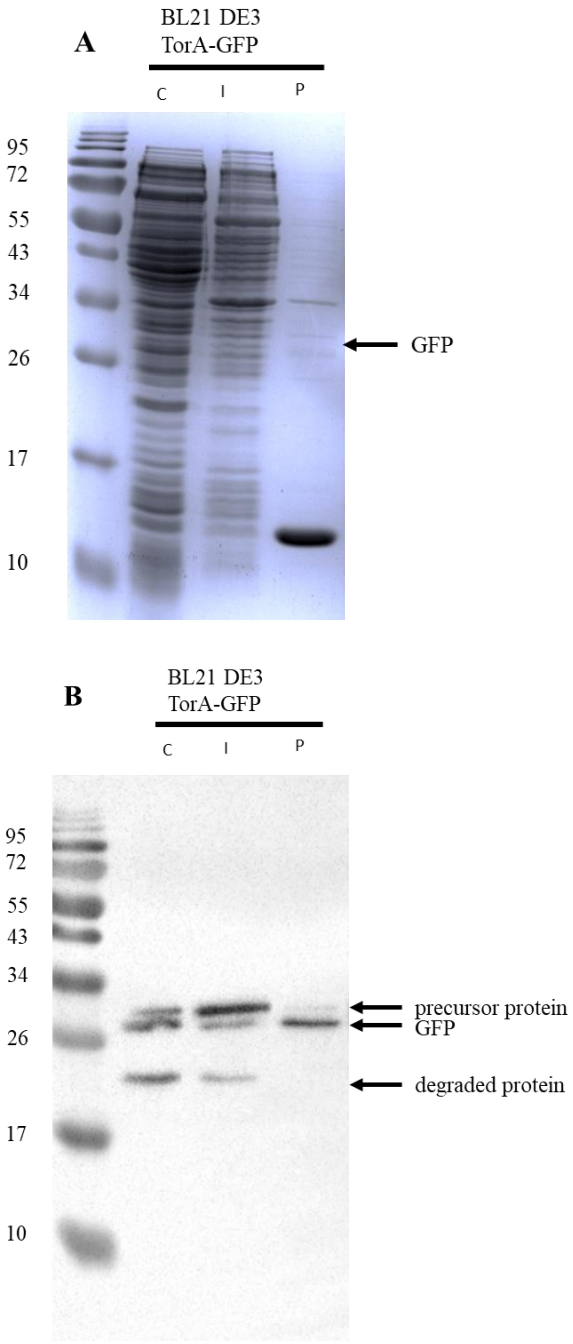


Figure 28: **GFP is expressed in BL21 DE3.** BL21 DE3 cells containing TorA-SP-GFP were expressed in a 50mL LB shake flask culture at 30˚C. The plasmid containing GFP (27 kDa, arrowed) was induced with 1 mM IPTG and was expressed. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. The western blot was immunoblotted using a C-terminal anti-his antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 5.8 seconds.

3.2.7. MC4100 *A. tumefaciens* TatABC is not capable of exporting co-translationally folded proteins

scFv's, hGH, and GFP are typically translocated by the *E. coli* Tat system and are well established for testing Tat translocation abilities. None of those proteins was translocated by Clone 1 or Clone 2. Therefore, we investigated further proteins, that were previously exported by *E. coli* wildtype Tat. The proteins investigated are co-translationally folded and their expected sizes are smaller than previously investigated proteins. Co-translational folding means that the substrate is folded during protein synthesis. Proteins that are folded rapidly are translocated by the Tat system. It is not known at what stage of synthesis Tat substrates interact with the Tat complex and therefore substrates that are folded co-translationally are an interesting target group. The here tested substrates are spectrin R16 (13.4 kDa) and Protein G B1 domain (Protein G) (7 kDa). Spectrin R16 is a three-helix bundle protein which is crucial for the integrity of plasma membranes(Glyakina *et al.*, 2018).

Neither Clone 1 nor Clone 2 expressed R16 (Figure 29A and Figure 29B). Furthermore, no signal could be achieved in the MC4100 wildtype as seen in Figure 29B. The original expression of this protein was carried out in BL21 (DE3) cells and expression trials need to be repeated in BL21 (DE3) as a positive control.

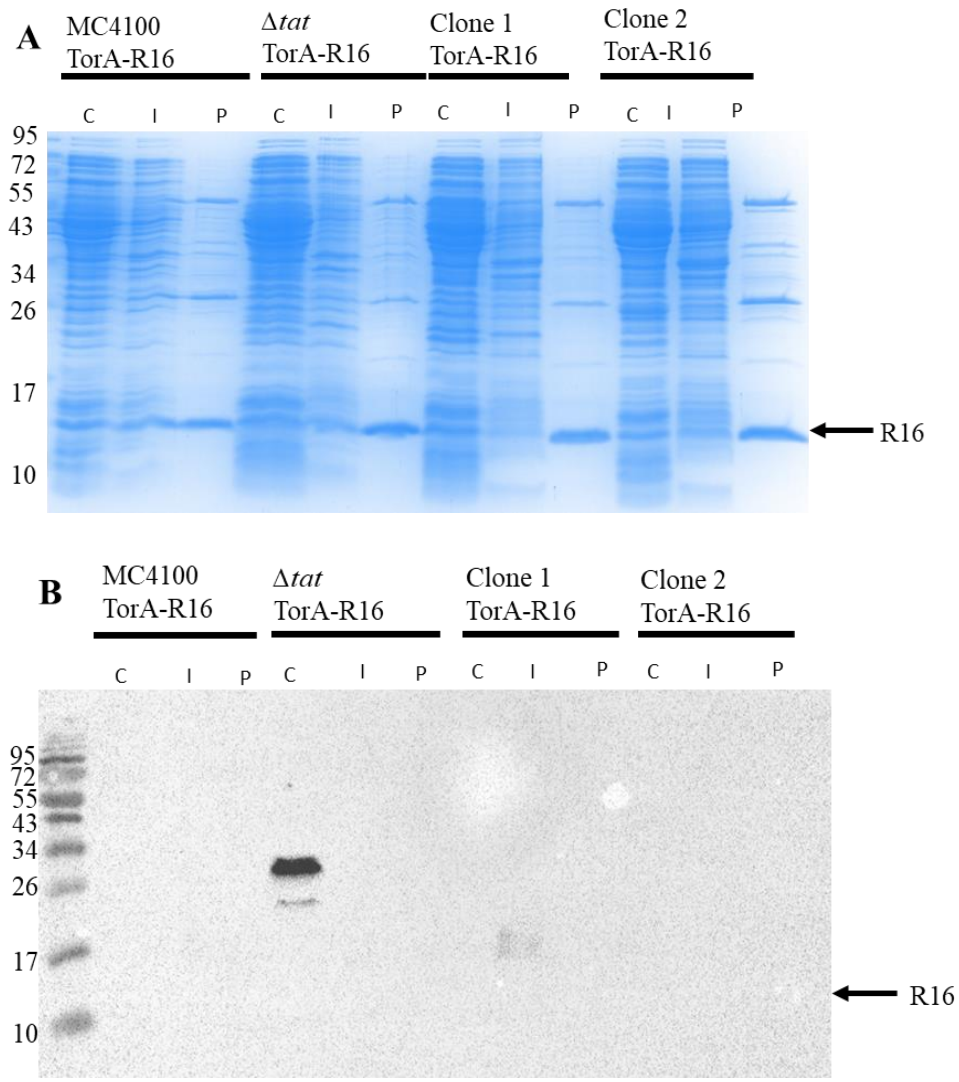


Figure 29: **Expression of TorA-SP-R16 in the new strains compared to MC4100 and Δ*tat*.** Cells were grown in 50mL LB shake flask cultures at 30˚C. The plasmid containing R16 (13.4 kDa, arrowed) was induced with 1 mM IPTG after reaching an OD_{600} of 0.4 and was expressed for 3-hours. Samples were fractionated. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot of the fractions. The western blot was immunoblotted using a C-terminal anti-his antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 5 seconds.

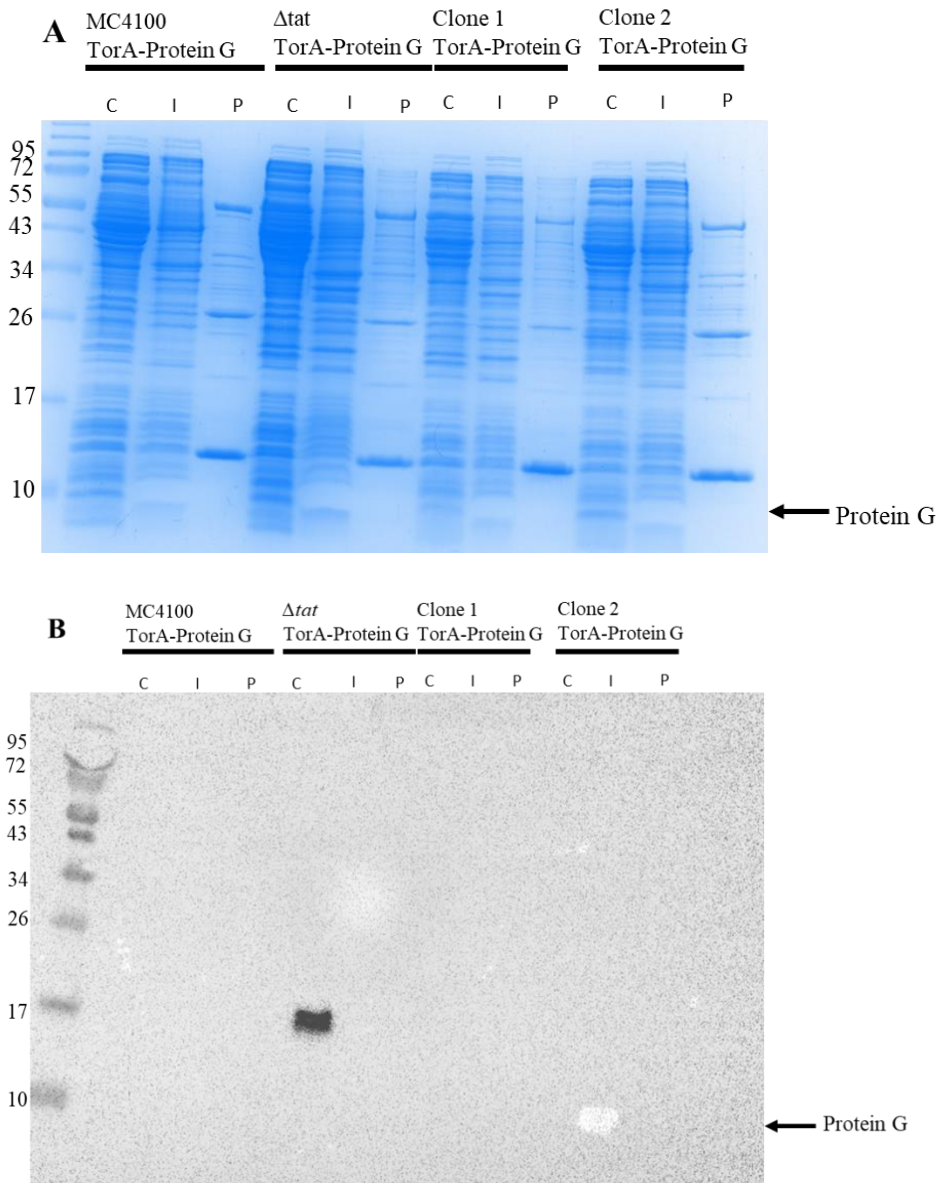


Figure 30: **Expression of TorA-SP-Protein G in the new strains compared to MC4100 and Δ*tat*.** Cells were grown in 50mL LB shake flask cultures at 30˚C. The plasmid containing Protein G (7 kDa, arrowed) was induced with 1 mM IPTG after reaching an OD_{600} of 0.4 and was expressed for 3-hours. Samples were fractionated. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot of the fractions. The western blot was immunoblotted using a C-terminal anti-his antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 5 seconds.

We furthermore investigated the Protein G domain, which also could not be expressed in both clones and in MC4100 (Figure 30A and Figure 30B). The original expression Protein G was carried out in BL21 (DE3) cells and expression trials would have needed to be repeated in BL21 (DE3) as a positive control which was not possible due to the research laboratory closure.

3.3. Discussion

In this study, we investigated the ability of an alternative *tat* operon to perform export in *E. coli* cells. This potentially could enable the production of biopharmaceutical targets in *E. coli* that currently require production in mammalian cells due to their complex folding structures. We chose the *tatABC* operon of *A. tumefaciens* and we show that inserting the operon onto the genome can improve the growth pattern of Δ*tat* cells. Cells with a non-functional Tat system are lacking the ability to separate the cell membrane which is referred to as 'chaining'. We can show that the non-native operon is capable of integrating into the genome and restores cell division. This suggests that native Tat proteins can be exported by the foreign Tat operon. This result is exciting because so far expression studies on other systems have involved expression of plasmid borne *tat* operons in the *E. coli* Δ*tat* strain. We furthermore proved by exporting the *E. coli* native Tat protein TorA into the periplasm that the new *E. coli* MC4100 clones are capable of exporting a Tat native substrate. Throughout the course of this study two clones have been created containing the *A. tumefaciens* TatABC sequence on the genome. Both sequences have been confirmed by sequencing and Clone 1 has a point mutation which is affecting the 41st amino acid mutating a lysine to an arginine. This could be due to point mutations arising during the PCR for the GATC sequence or throughout the gene doctoring process. Despite the mutation in TatB, Clone 1 is capable of exporting a *E. coli* native Tat protein TorA.

Furthermore, TatA of Clone 2 can be immunoblotted using an *E. coli* TatA specific antibody, but Clone 1 cannot. The sequencing shows that both clones have minor point mutations in their sequence which could be the reason for the differences in phenotype and export. Further research into the differences between Clone 1 and Clone 2, such

as transcription levels or membrane integration, could have allowed further insight into the expression differences.

Additionally, we conducted expression trials on both clones. Clone 1 is capable of expressing scFv(cobra) but neither Clone 1 nor Clone 2 are able to export scFv(cobra) into the periplasm. Further proteins assessed in this study are scFv(M), GFP, hGH, R16, and Protein G. None of those proteins was exported into the periplasm. These proteins were chosen as they have previously been shown to be exported using the *E. coli* Tat system and are good target proteins to test Tat export (Delisa, Tullman and Georgiou, 2003; Browning *et al.*, 2017). However, these proteins are degraded quickly and the reason for this is not clear. It could be that some sort of proofreading process is involved which degrades the proteins when their export is unsuccessful (Alanen *et al.*, 2015; Sutherland *et al.*, 2018).

The two MC4100 clones containing the *tatABC A. tumefaciens* operon on the genome are capable of exporting *E. coli* native proteins. However, their proofreading mechanisms are more selective than the *E. coli* native one. None of the substrates tested in this study could be exported into the periplasm. *A. tumefaciens* is mostly known for its infectious behaviour towards plants and the translocation of proteins by the Tat system of this bacterium has not been researched extensively. It is known that the TatA, TatB, and TatC from *A. tumefaciens* are similarly integrated into the cytoplasmic membrane compared to TatABC. Based on the current information, it is unclear why the *tat* operon from *A. tumefaciens* is more restrictive than the *E. coli tat* operon.

In summary, Clone 1 shows a point mutation in the TatB region, shows a phenotype change (clumping of cells), exports *E. coli* native Tat protein TorA very well, and

TatA is not picked up by an *E. coli* TatA specific antibody. Clone 2 has the correct codon optimised *A. tumefaciens* TatABC sequence, the filamentous phenotype is complemented, the *E. coli* native Tat protein TorA is exported but at reduced rate compared to Clone 1, and TatA is detected by an *E. coli* TatA specific antibody. It is thus not clear why Clone 2 shows no complementation of the phenotype – if anything, the correct Tat sequences and clear expression of TatA protein suggest that it should be superior to Clone 1. It could be that the gene doctoring process has introduced other mutations in the genome, but we have no information about any such additional changes.

In future experiments, the new clones could be further tested about their expression ability. Further plasmids for the gene doctoring process were designed and in future experiments I would like to insert TatABC from *Mesorhizobium loti* and TatAAC from *Synechocystis sp.* onto the *E. coli* genome and investigate their export abilities.

Furthermore, the proteins investigated in this study showed better expression in BL21 (DE3) wildtype cells than MC4100. Therefore, in further experiments I would like to change the BL21 (DE3) genome to see whether the expression can be improved.

4. Optimising protein expression using a nitrate inducible promoter

4.1. Introduction

The production of proteins is essential for every organism. In organisms, expression systems are naturally controlled by promoters (Haugen, Ross and Gourse, 2008). Those promoters have been used to express recombinant proteins, the *tac* promoter for example is industrially well established for expression in *E. coli* because it is a strong promoter and is easily regulated (Rosano and Ceccarelli, 2014; Öztürk, Ergün and Çalık, 2017). The *tac* promoter is a hybrid promoter derived from the *trp* and *lac* promoter (de Boer, Comstock and Vasser, 1983). In the past, gene expression of up to 50% of total cellular protein in *E. coli* has been achieved (Miroux and Walker, 1996). This shows promoters are a powerful tool for biopharmaceutical production. Improving promoter strength as well as improving inducibility and tightness of regulation can have different benefits. It can improve the quality of the expressed protein and it can improve cost efficiency. Optimising the promoter used for expression of the target protein allows production of high quantities of recombinant protein at low costs. Induction of the *tac* promoter using IPTG can be costly in large processes and hence cheaper alternatives are required. During an induction experiment, cells are grown until they reach mid-log phase. Then the cells are induced to initiate protein expression. After a process-dependent expression time the cells are harvested for protein purification.

Furthermore, promoters have to be tightly controlled because access levels of target protein can induce secretion stress and can cause the protein to go into inclusion bodies (Kawe, Horn and Plückthun, 2009). Costs are unfortunately often increased by the inducer needed for the promoter. The previously mentioned *tac* promoter is induced by Isopropyl-β-D-thiogalactopyranosid (IPTG), as are others such as the *grac* promoter (Phan, Nguyen and Schumann, 2012).

In this study, the aim was to investigate if an alternative promoter, p*narG-CC*, could provide a cheaper and more effective tool for the biotechnological industry. The *narG-CC* promoter has been designed by the Steve Busby laboratory from the University of Birmingham and a fusion of the the *narG* promoter and the *CC* (-40.5) promoter. A schematic presentation of the promoter is shown in Figure 31. Binding sites of NarL which is a transacting regulator, the integration host factor (IHF) and a cyclic AMP receptor protein (CRP) (Andrew J Darwin and Stewart, 1996; Browning, Butala and Busby, 2019) are used in this promoter. IHF is only found in gram-negative bacteria and introduces DNA modification that supports DNA interaction in nucleoprotein arrays (Goosen and van de Putte, 1995). CRP controls transcription by binding to the effector cAMP (Yang *et al.*, 2016). CRP is a strong activator when it is at position -41.5 upstream of the transcriptional start site. When it is at position -40.5, however, CRP is less active and the promoter requires another transcription factor (Rossiter *et al.*, 2011, 2015). This support is provided by the regulatory element NarL.

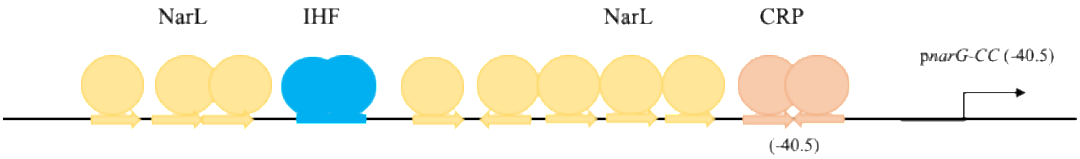


Figure 31: **The figure shows a schematic presentation of the *narG-CC* (-40.5) promoter.** The location of the binding sites for NarL, IHF and CRP are indicated. The start site of transcription is shown by a bent arrow. (Adapted from a presentation given by Douglas Browning at the Recombinant Protein Production 10 conference).

Sodium nitrate is easily available even in countries that may have difficulties to receive biopharmaceutical supplies. Furthermore, sodium nitrate is a cheaper resource than IPTG. As nitrate can be used by *E. coli* cells during anaerobic conditions, the

metabolism of the cells needs to be adapted so that sodium nitrate is not metabolised throughout the expression process (Richardson and Watmough, 1999). Further challenges a new promoter has to overcome are efficiency and the transcriptional activity needs to be tightly controlled (Makrides, 1996; Terpe, 2006). Therefore, we aimed to determine whether p*narG-CC* can express similar amounts of protein compared to more commonly used promoters and how tightly controlled the promoter is. Additionally, we investigated whether the promoter allows export of protein into the periplasm via the Tat pathway, as well as whether nitrate is continuously available for fermentation and therefore a single induction is supported.

4.2. Results

4.2.1. p*narG*-CC can be used for single induction

To be an efficient tool for the biotechnology industry the *narG*-CC promoter has to be able to be expressed with a single induction, like the *tac* promoter. Therefore, it is important to ascertain whether the nitrate used for induction is metabolised by the host strain. Wildtype *E. coli* has three pathways involved in nitrate metabolism under anaerobic conditions. The strains used in this work have a deletion of the major nitrate pathway, the *narGHJI* operon, and are called N11 and Tn3.1 (Douglas Browning, University of Birmingham, unpublished work). N11 is an *E. coli* K-12 based strain which has *narG* (the major nitrate reductase) knocked out. The strain Tn3.1 has the same deletion but is based on *E. coli* K-12 W3110 TatExpress. Both strains have been engineered in the Steve Busby lab using gene doctoring (Lee *et al.*, 2009).

To determine whether the sodium nitrate would be metabolised by the cells we ran a fed-batch fermentation containing *E. coli* N11. This was to prevent the strain from using nitrate to respire under anaerobic conditions. We transformed a plasmid expressing GFP driven by the *narG*-CC promoter into the N11 strain. After induction with 20 mM sodium nitrate, time points were taken during the fermentation run and we measured the nitrate concentration in the medium over the time course of 38 hours. If the nitrate concentration is stable throughout the run, the promoter in combination with the strain is feasible for comparison to the *tac* promoter.

Sodium nitrate appears to be continuously present in the medium at approximately 20 mM concentration and continuously induces the *narG*-CC promoter. The nitrate assay proves that a 20 mM nitrate induction is a sufficient amount of inducer for a 38-hour fermentation process (Figure 32). Additionally, in the western blots of this fermentation run (Figure 35 on page 119) the protein concentration of GFP appears to

increase which supports the fact that nitrate is continuously available throughout fermentation.

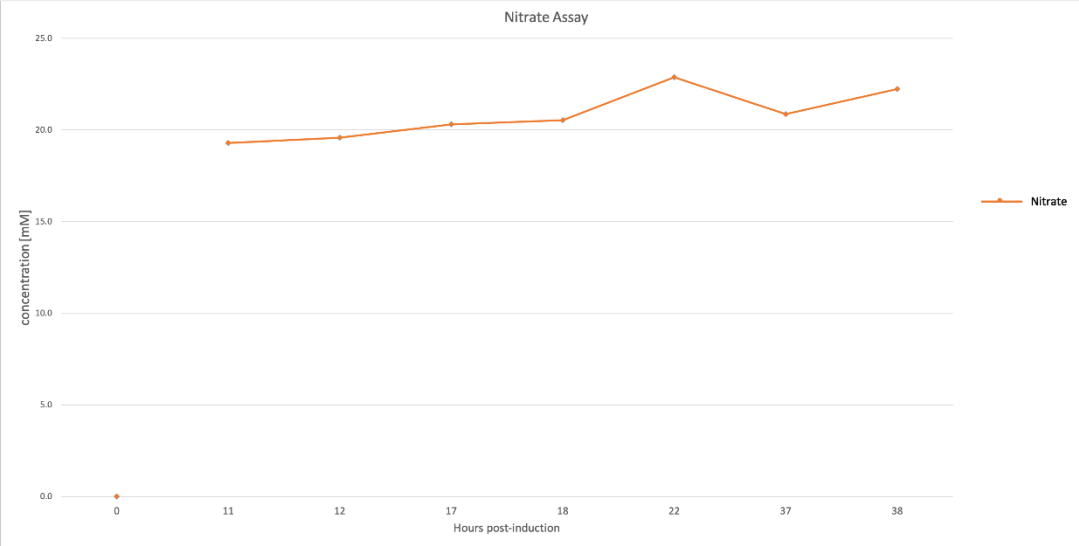


Figure 32: **Nitrate and nitrite concentration in the medium of a fed-batch fermentation.** GFP was expressed using a sodium nitrate inducible promoter (p*narG*-CC) in a fed-batch fermentation in SM6 medium. 20 mM sodium nitrate was added as inducer after 11 hours growth in the fermenter and the nitrate concentration has been measured post-induction with a commercial nitrate colorimetric assay as described in 2.8.1 Nitrite/ Nitrate assay kit (Sigma Aldrich). Induction is marked by the time point 0 hours. The N11 strain does not contain *narG*, which encodes nitrate reductase, on the genome.

4.2.2. p*narG*-CC shows tightly controlled expression in MS and SM6 media

Ideally the new promoter should be tightly controlled and not express the protein of interest before induction. We investigated different media to determine which ones allow tight regulation of the promoter. We compared a fermentation medium, SM6, to the MS medium used in Birmingham, both described in the Materials and Methods. Both media are minimal salt media that contain salts and nitrogen, and glycerol as

energy source. A huge benefit of minimal salt media is the reproducibility due to each of its components being defined (Kolter, 1951). The main difference between SM6 and MS media is the amount of glycerol in the media. SM6 medium contains 95 g/L glycerol and MS medium contains 4 g/L glycerol. Therefore, the availability of glycerol as a carbon source varies strongly. The media also have minor differences in the supplemented salts.

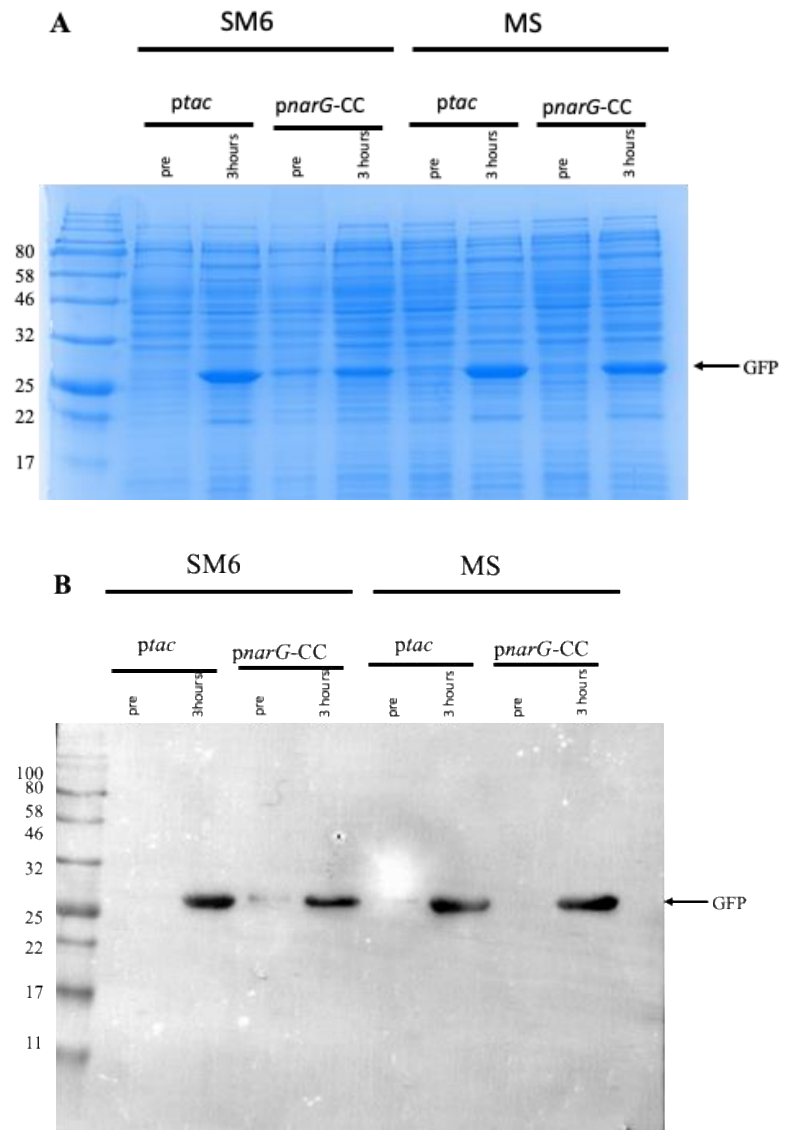


Figure 33: **Comparison of expression of GFP in SM6 medium to MS medium.** GFP (27 kDa, arrowed) was expressed in 50mL shake flasks cultures at 37˚C. SM6 and MS media was used. Samples were taken before induction and 3 hours post-induction and the whole extract for analysis was prepared. We investigated both the p*tac* and the p*narG*-CC promoters in regard to the tightness of the regulation of expression of GFP in N11 cells. Samples were analysed by **A** Coomassie stained SDS-PAGE gel and **B** western blot. The samples were immunoblotted using an anti-GFP antibody. The gel is a representative of three gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 2.8 seconds. The whole cell extract has been used.

The MS medium shows tight control of expression compared to SM6 medium. The recombinantly expressed GFP induced using nitrate (p*narG*-CC) or IPTG (p*tac*) can be clearly seen in the Coomassie stained SDS-PAGE gel as both promoters show high levels of expression (Figure 33A). The comparison to p*tac* expressing GFP is the first evidence that the *narG*-CC promoter expresses high levels of protein in shake flask performed. Previous research by the Busby lab in Birmingham focused on comparing different *narG*-CC promoter designs in performance. In SM6 medium the *narG*-CC promoter appears to be slightly leaky, as in Figure 33B a low-level signal in the pre-induction sample can be seen. The expression in MS medium is tightly controlled. In further experiments, we changed the base used to increase the pH of the SM6 medium to sodium hydroxide instead of ammonia to investigate whether changes in the media lead to tighter expression (Figure 34A and Figure 34B).

Using SM6 as a medium is preferred, as it is the medium used in fermentation in our system and the fermentation protocol is optimised for SM6.

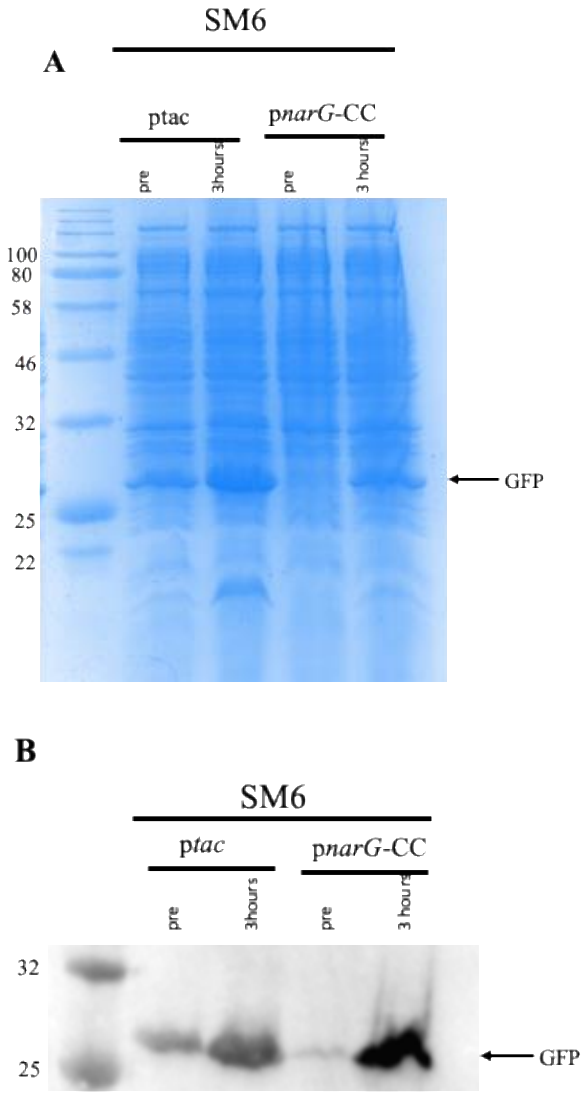


Figure 34: **Comparison of control of expression of GFP from p*tac* and the *narG*-CC promoters.** The protein was expressed in 50mL shake flask cultures using SM6 medium containing 10 % sodium hydroxide as a base. The cell line used here is N11 and the cells were grown at 37˚C in SM6 and MS media. Samples were taken before induction and 3 hours post induction and whole cell extract was prepared. Samples were analysed by **A** Coomassie stained SDS-PAGE gel and **B** western blot of GFP (27 kDa, arrowed) containing samples. The samples were immunoblotted using an anti- GFP antibody. The gel is a representative of three gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 2.8 seconds.

The *narG*-CC promoter was compared to the IPTG induced p*tac*, which shows slightly leaky expression of GFP in the SM6 medium containing 10 % sodium hydroxide (Figure 34). Furthermore, the *narG*-CC promoter seems to be reaching higher levels of expression of GFP than p*tac*. The *narG*-CC promoter is still slightly leaky (Figure 34B, signal in the pre-induction sample) even in the SM6 medium variant. The medium can influence the transcriptional regulation as seen on the *tac* promoter.

p*narG*-CC is capable of high-level expression in shake flask in different minimal media. The level of expression in shake flask experiments is comparably high as the expression of the *tac* promoter.

4.2.3. Expression of GFP in fed-batch fermentation

In the next step, we investigated the expression of GFP under control of p*narG-CC* under fermentation conditions. Fed-batch fermentation provides ideal conditions to produce high cell mass which can support expression of metabolites. Figure 35 shows that expression of GFP with the *narG*-CC promoter can be scaled-up from a 50 mL shake flask to a 500 mL fermenter.

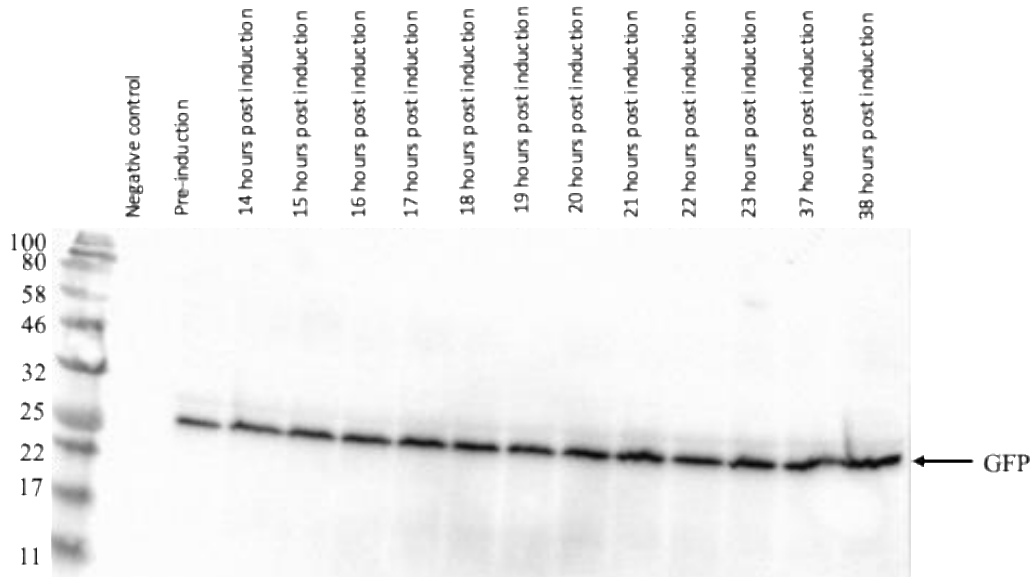


Figure 35: **GFP increases throughout the fermentation under the control of the *narG*-CC promoter.** *E. coli* N11 cells expressing GFP (27 kDa, arrowed) from the *narG*-CC promoter were grown in fed-batch fermentation as described in section 2.4.2 and samples were taken prior to the induction. The cells were grown under fed-batch conditions in SM6 medium with glycerol added at a rate of 0.06 uL/min. Throughout the fermentation and the whole cell extract was prepared for immunoblotting. Here shown is the western blot of the expression of p*narG*-CC GFP in N11 cells in SM6 medium. At an OD_{600} of 36 the pre-induction sample was taken, and the cells were induced with 20mM sodium nitrate. As negative control N11 cells expressing an empty plasmid were used. The fractions were immunoblotted using an anti-GFP antibody. The gel is a representative of one gel and 10 µL sample have been loaded in each well. The exposure time of the western blot is 3 seconds.

The experiment shown in Figure 35 was repeated, as the cells were not growing over an OD_{600} of 36. In the next fermentation process, we increased the glycerol feed from 0.06 µL/min of maximum pump feed to 0.15 µL/min to improve the growth conditions and achieved an OD_{600} of 136 (fermentation in Figure 35 and Figure 37). The GFP is

clearly visible in the culture (Figure 36) and the GFP levels could have been quantified by fluorescence. However, we were not able to do this due to a lack of time. Cells expressing GFP from the *tac* promoter can show similar levels of GFP expression as shown in the following section (4.2.4 pnarG-CC is able to express comparable amount of GFP to ptac). Furthermore, induced expression could have been compared to an experiment without induction because the nitrate in the media could have meant equally strong expression.

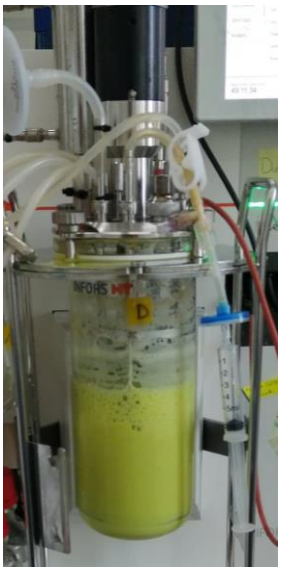


Figure 36: **Fermentation expressing p*narG*-CC GFP.** The culture shows a yellow-green colour due to the expression of GFP.

The GFP is clearly recognisable in the Coomassie-stained gel as a band at about 27 kDa Figure 37. In both fermentation processes GFP was expressed pre-induction, indicating that expression is not tightly controlled.

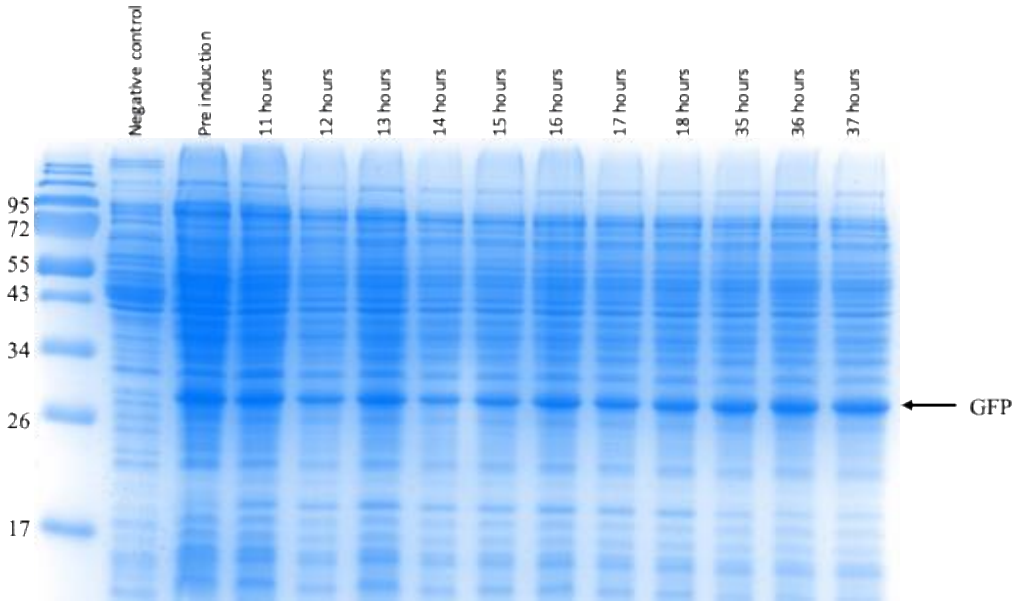


Figure 37: **Coomassie stained SDS-PAGE gel showing the production of GFP from *E. coli* N11 expressing the GFP (27 kDa, arrowed) gene from the nitrate-inducible *narG*-CC promoter.** The cells were grown under fed-batch conditions in SM6 medium with glycerol added at a rate of 0.15 uL/min. The culture was induced after 11h with 20 mM sodium nitrate. For this experiment the glycerol feed has been increased from 0.06 µL/min of maximum pump feed to 0.15 µL/min. As negative control N11 cells expressing an empty plasmid were used. Here shown is the whole cell extract. Samples were analysed by Coomassie stained SDS-PAGE gel.

The results presented in Figure 37 were achieved by running the fermenter with SM6 medium and increasing the feed rate of glycerol from 0.06 µL/min to 0.15 µL/min. The increased feed rate shows a higher OD during the GFP production as shown in Figure 38. Due to the length of the fermentation process, it was not possible to take data points every hour.

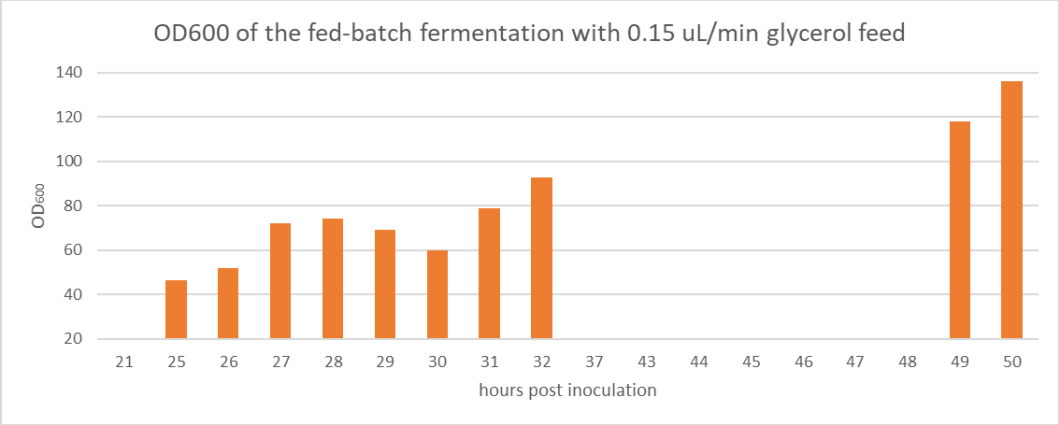


Figure 38: **Growth curve of the fermentation of *E. coli* N11 with 0.15 uL/min glycerol feed rate.** The cells were grown under fed-batch conditions in SM6 medium. The cultures were induced with 20 mM sodium nitrate.

In the next fed-batch fermentation experiment we used SM6 pH adjusted with 10% sodium hydroxide instead of ammonium because of a potential interference of the ammonium with the *narG*-CC promoter.

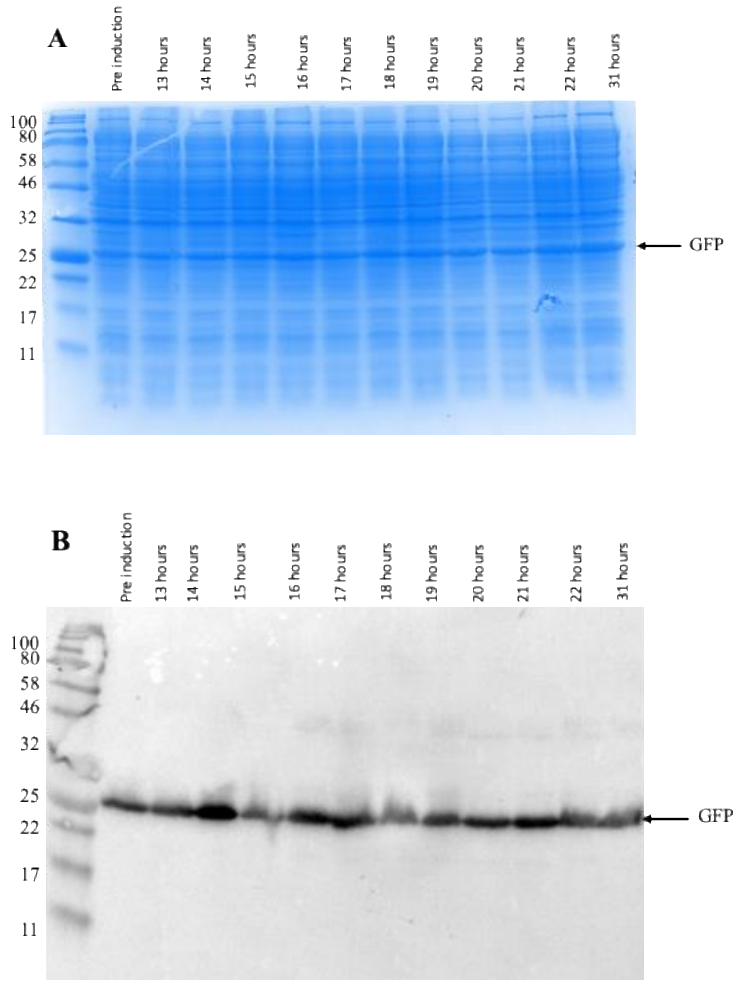


Figure 39: **Coomassie stained SDS-PAGE gel and western blot showing the production of GFP from *E. coli* N11 expressing the GFP (27 kDa, arrowed) gene from the nitrate-inducible *narG*-CC promoter.** The cells were grown under fed-batch conditions in modified SM6 medium (NaOH instead of ammonium) with glycerol added at a rate of 0.15 uL/min. The culture was induced after 11 h with 20 mM sodium nitrate. pH was monitored using ammonium and 25% sulfuric acid. Here shown is the whole cell extract of the cells. As negative control N11 cells expressing an empty plasmid were used. Samples were analysed by **A** Coomassie stained SDS-PAGE gel and **B** western blot. The fractions were immunoblotted using an anti-GFP antibody. The gel is a representative of one gel and 10 µL sample have been loaded in each well. The exposure time of the western blot is 2.2 seconds.

The expression of GFP in the fed-batch fermentation with modified growth conditions (glycerol feed rate from 0.06 µL/min to 0.15 µL/min) was high based on that GFP is the most abundant protein in the cells. However, tightly controlled expression could not be achieved in the fermenter. In this figure the pre-induction sample shows a high level of GFP signal indicating expression prior to induction. Adjusting the pH of the SM6 medium with ammonium could have caused the pre-induction. However, as ammonium was still used as a feed during the fermentation, it is not possible to determine if the ammonium has an influence on the *narG*-CC promoter Based on the tightly controlled expression in shake flask (Figure 33), the experiment should be carried out using a different method to adjust the pH.

We also conducted fermentation experiments with MS medium. During these fermentations, the cells did not show any significant growth over the period of 48 hours.

4.2.4. p*narG*-CC is able to express comparable amount of GFP to p*tac*

To determine whether the *narG-CC* promoter can produce similar amounts of GFP to the *tac* promoter, we have compared the production of GFP in the fermentation runs in SM6 medium pH adjusted with NH_4OH at 0.06 µL/min glycerol feed. When the glycerol feed was increased to 0.15 µL/min (Figure 39) the cells grew better, and they were producing more GFP than p*tac*.

In the first presented fermentation (Figure 35), the p*narG*-CC GFP expressing cells grew poorly and a final OD_{600} of 32 was reached. Under the same conditions the cells expressing p*tac* GFP reached an OD_{600} of 94. The *narG*-CC GFP expressing cells

produced less GFP than p*tac* expressing cells. This is clearly visible on the Coomassie stained SDS-PAGE gel (Figure 40A) and on the western blot (Figure 40B).

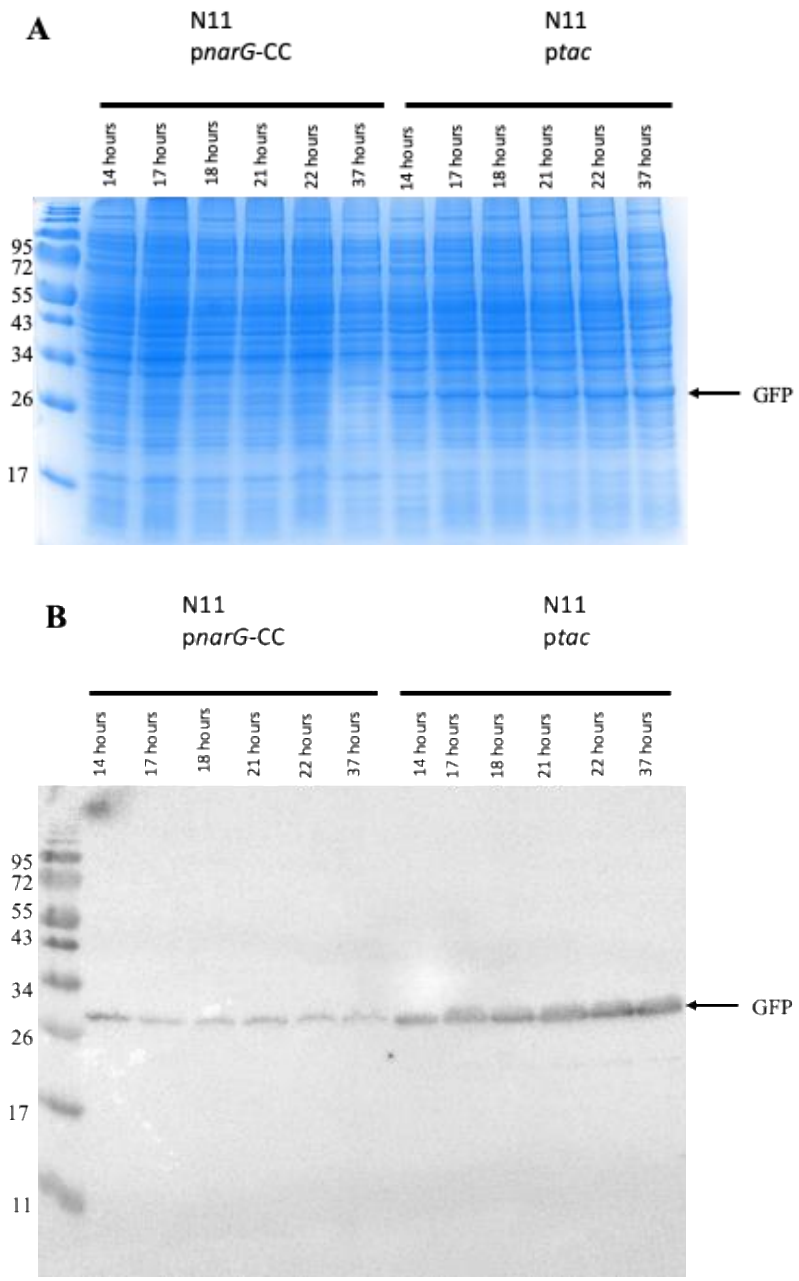

A
N11
pnarG-CC
N11
ptac
14 hours
17 hours
18 hours
21 hours
22 hours
37 hours
14 hours
17 hours
18 hours
21 hours
22 hours
37 hours
95
72
55
43
34
26
17
GFP
B
N11
pnarG-CC
N11
ptac
14 hours
17 hours
18 hours
21 hours
22 hours
37 hours
14 hours
17 hours
18 hours
21 hours
22 hours
37 hours
95
72
55
43
34
26
17
11
GFP

Figure 40: **Comparison of GFP production from *E. coli* N11using the *narG*-CC promoter and the *tac* promoter.** Samples were taken throughout a fed-batch fermentation run in SM6 medium and a feed rate of glycerol at 0.06 µL/min. The N11 strain was used for fermentation. Cells expressing the *tac* promoter were induced with 1mM IPTG and cells expressing the *narG*-CC promoter were induced using 20 mM sodium nitrate. The culture containing p*narG-CC* reached a final OD_{600} of 32 and the culture containing p*tac* reached an OD_{600} of 94. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Here, the *tac* promoter produced more GFP and GFP is visible even on the Coomassie stained SDS-PAGE gel, see arrow. The fractions were immunoblotted using an anti-GFP antibody. The gel is a representative of one gel and 10 µL sample have been loaded in each well. The exposure time of the western blot is 2.2 seconds.

In a fed-batch fermentation with a glycerol feed rate of 0.15 µL/min p*tac* is not exporting as much GFP (shown in Figure 41) as it has in the previous experiment (Figure 40) were a glycerol feed rate of 0.06 µL/min was used. The reasons for this are unknown. However, we did not have time to repeat this experiment.

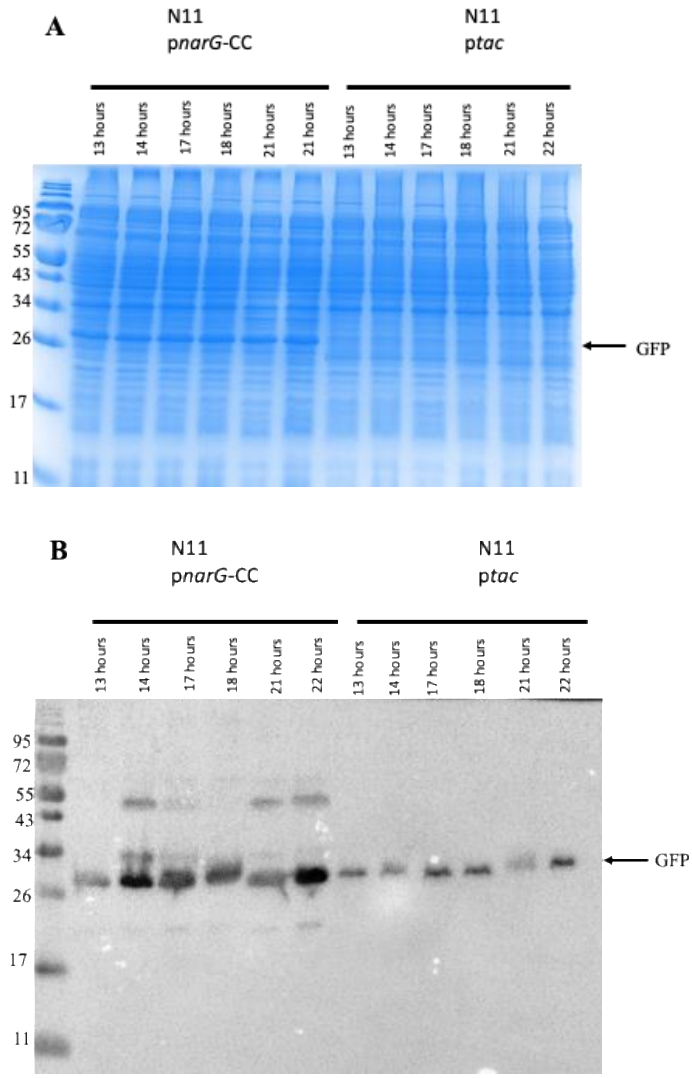


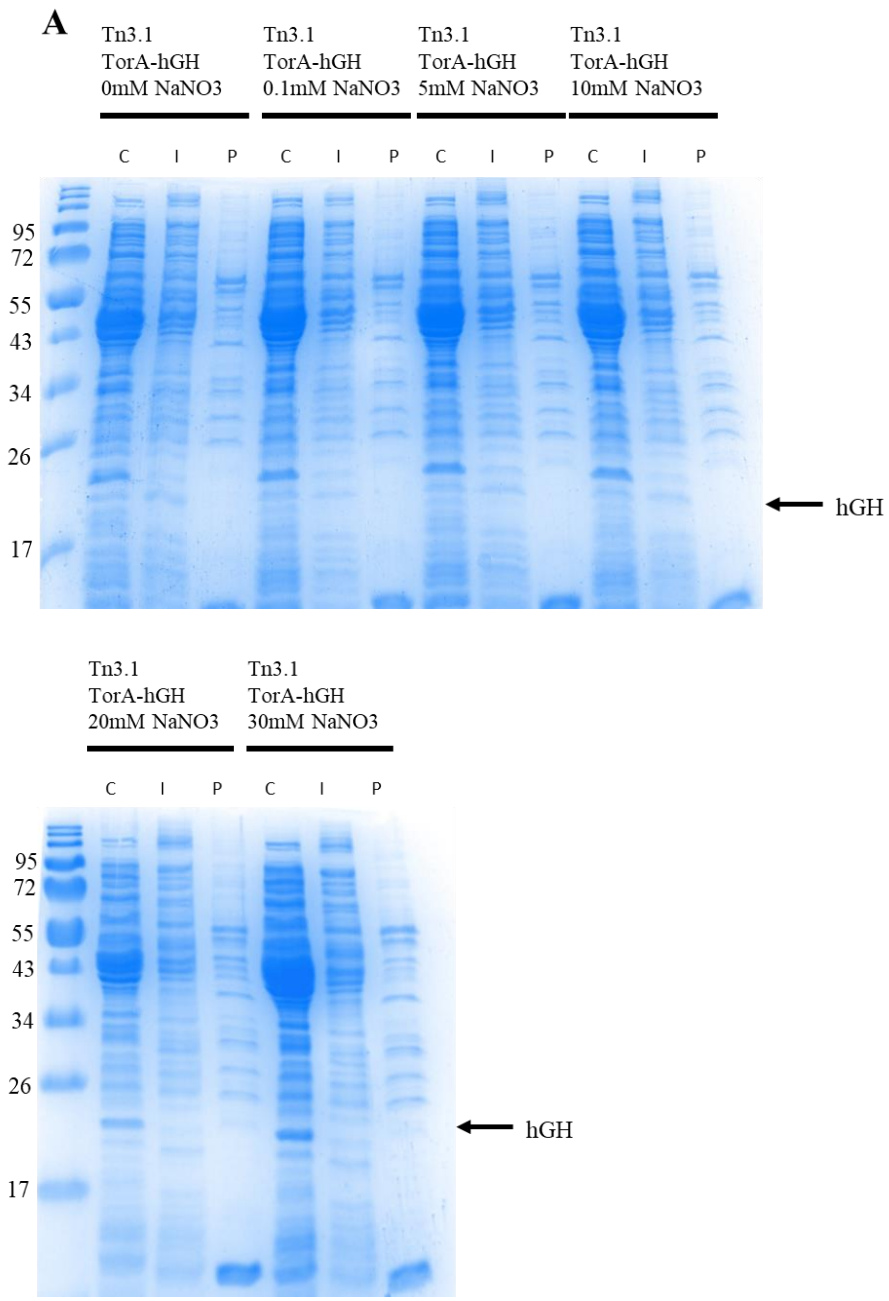
Figure 41: **Comparison of GFP concentration expressed using the *narG*-CC promoter compared to the *tac* promoter.** This experiment used SM6 medium with a glycerol feed rate of 0.15 µL/min. Samples were induced with 20mM sodium nitrate (p*narG*-CC) and 1mM IPTG (p*tac*). Under these conditions the *E. coli* strain N11 reached an OD_{600} of 100. Here shown is the whole cell extract of the cells. Samples were analysed by **A** Coomassie stained SDS-PAGE gel and **B** western blot. In this run the *narG*-CC promoter produced more GFP and GFP is visible even on the Coomassie stained SDS-PAGE gel. The fractions were immunoblotted using an anti-GFP antibody. The gel is a representative of one gel and 10 µL sample have been loaded in each well. The exposure time of the western blot is 2.2 seconds.

In summary, the *narG*-CC promoter was capable of achieving similarly high levels of recombinant protein production as the *tac* promoter, providing the batch culture condition are adjusted for the higher consumption of glycerol. However, the high-level expression in the pre-induction sample shows that expression is not tightly controlled, and further experiments are required to investigate the reason for the leaky expression in fed-batch fermentation.

4.2.5. p*narG*-CC can express TorA-SP-hGH in shake flask under specific conditions

In the next step, we expressed human growth hormone (hGH), TorA-SP-hGH, via the Tat pathway into the periplasm in the Tn3.1 strain which over-expressed Tat components and has *narG* deleted (unpublished work, Douglas Browning, University of Birmingham). Periplasmic expression is favourable as it allows easier and cheaper downstream processing (Matos *et al.*, 2012) and hGH is a protein of pharmaceutical interest that was previously successfully expressed and exported into the periplasm (Alanen *et al.*, 2015; Browning *et al.*, 2017). In the experiment shown in Figure 42, the cells were grown in LB medium in order to induce the *narG*-CC promoter. Samples were fractionated into the cytoplasmic (C), insoluble(I), and periplasmic (P) fractions and immunoblotted. western blots of fractionated samples show that TorA-SP-hGH is exported to the periplasm (periplasmic signal in Figure 42B). Various concentrations of sodium nitrate were tested to establish whether there are differences in the expression levels. We can show that hGH does not show any expression before

induction and that at levels of 5 mM to 30 mM sodium nitrate TorA-SP-hGH is being expressed and efficiently exported into the periplasm. hGH cannot be seen on the Coomassie stained SDS-PAGE gel (Figure 42A). However, in the western blot a double band can be seen. This is likely caused by gel artefacts as clipping of hGH in the periplasm is not usually seen. Alternatively, it could be caused by F-methylation.



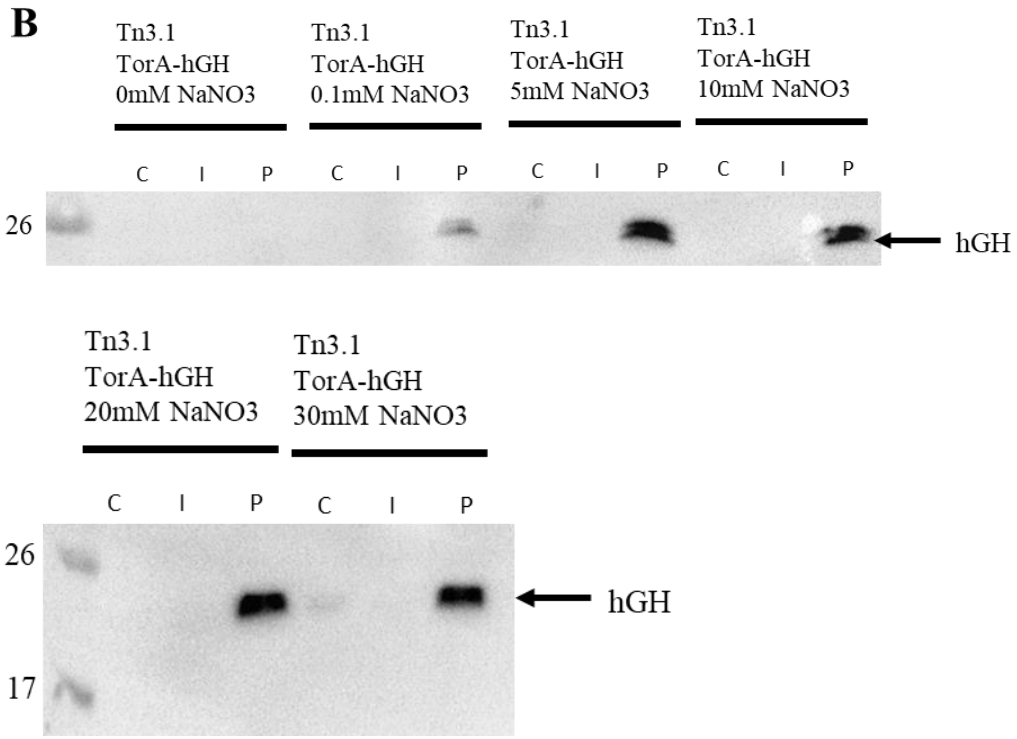


Figure 42: **Inducer concentration has minor impact on TorA-SP-hGH expression using the *narG*-CC promoter**. Here expressed is TorA-SP-hGH (21 kDa, arrowed) in the strain Tn3.1 in 50mL LB shake flasks cultures at 37˚C. Tn3.1 is a 'TatExpress' strain with an additional deletion of the major nitrate reductase. Samples were fractionated 3 hours post induction. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. The fractions were immunoblotted using an anti-hGH antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blots is 2 seconds.

We further conducted an experiment in MS medium with p*narG*-CC expressing TorA-SP-hGH in Tn3.1 in a shake flask. Minimal media are the preferred media for fermentation because their components are well defined which allows for a high reproducibility. After seeing successful periplasmic expression of TorA-SP-hGH in LB medium as shown in Figure 42 we expected to see expression in one of the minimal salt media. We chose MS medium for this experiment, as it showed similar expression

levels of GFP in SM6 when compared to previous experiments (Figure 33) and is (for us unknown reasons) the preferred medium for our collaborators at the University of Birmingham.

A 3-hour induction in MS medium at 30 ºC did not produce enough hGH to be detected on a western blot using an anti-hGH antibody (Figure 43A).

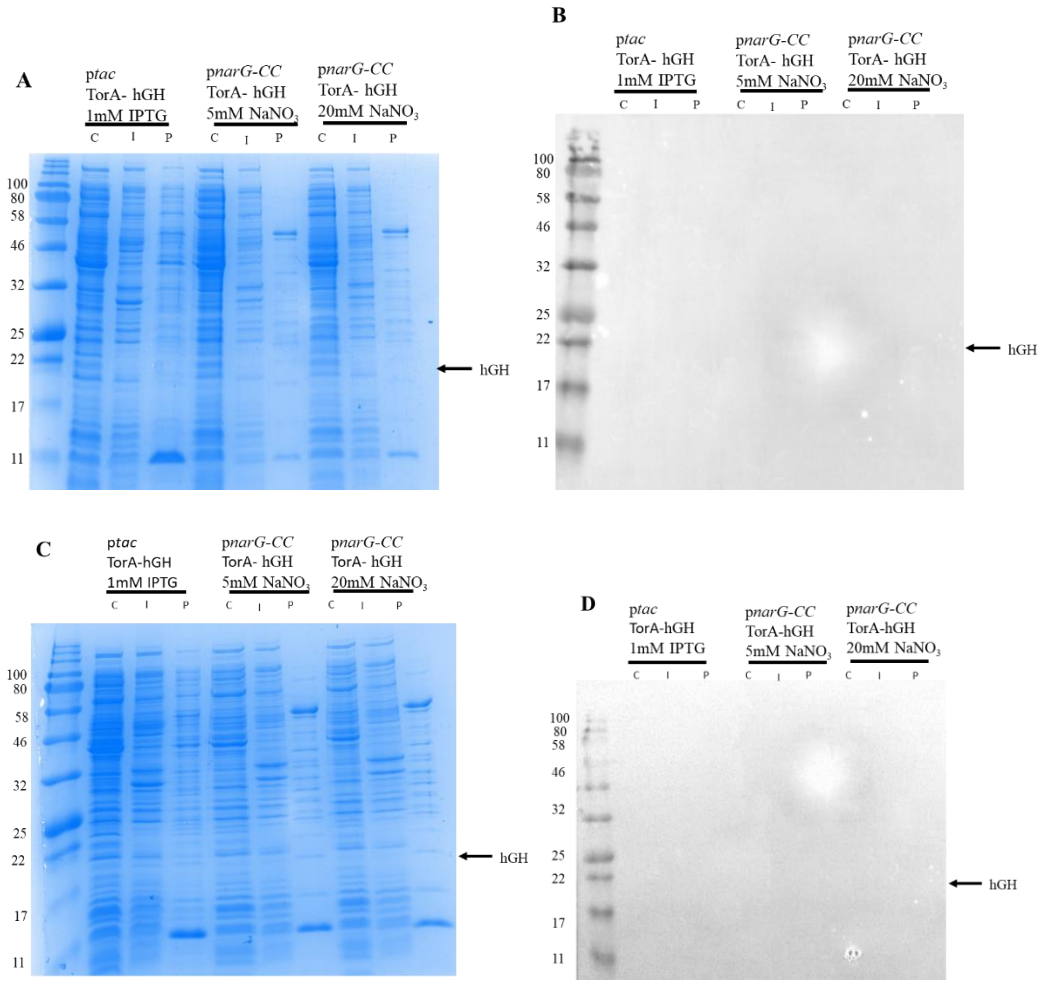


Figure 43: **Expression of TorA-SP-hGH in the Tn3.1 strain did not result in export using the p*narG*-CC promoter.** In **A** and **B** Tn3.1 TorA-SP-hGH was expressed using the *narG*-CC and *tac* promoter. The cells were grown at 30˚C and induced with 1mM IPTG (p*tac*), 5mM and 20mM sodium nitrate (p*narG*-CC). There was no detectable hGH in the fractions. In **C** and **D** Tn3.1 TorA-SP-hGH was expressed using the *tac* and *narG*-CC promoter. The cells were grown at 25˚C and were induced with 1mM IPTG (p*tac*), 5mM and 20mM sodium nitrate (p*narG*-CC). **A** and **C** show the Coomassie stained SDS-PAGE gel. **B** and **D** show the western blot. There was no detectable hGH in the fractions (cytoplasmic, C; insoluble, I; periplasmic, P). The fractions were immunoblotted using an hGH antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blots is 2 seconds.

In Figure 43 we expressed TorA-SP-hGH from the *narG-CC* promoter at 30˚C and 25˚C. The data in Figure 42 demonstrates that the p*narG*-CC promoter is capable of expressing TorA-SP-hGH in LB medium, and that it was exported to the periplasm. However, the concentration of hGH expressed in MS medium, with a 3-hour induction period, is too low to be detected by the antibody and no export could be detected. Lower temperatures can improve protein folding and therefore protein production (Weickert *et al.*, 1996). Hence, we conducted the same experiment under lower temperature to improve folding of the protein. The same experiment at 25 ºC also failed to show any protein (Figure 43B).

4.2.6. TorA-hGH can be expressed in a fermenter

In the next stage, we attempted to express TorA-SP-hGH from the *narG*-CC promoter in a fed-batch fermentation anticipating some expression of hGH related to the longer expression time compared to shake flask the shake flask experiment shown in Figure 42.

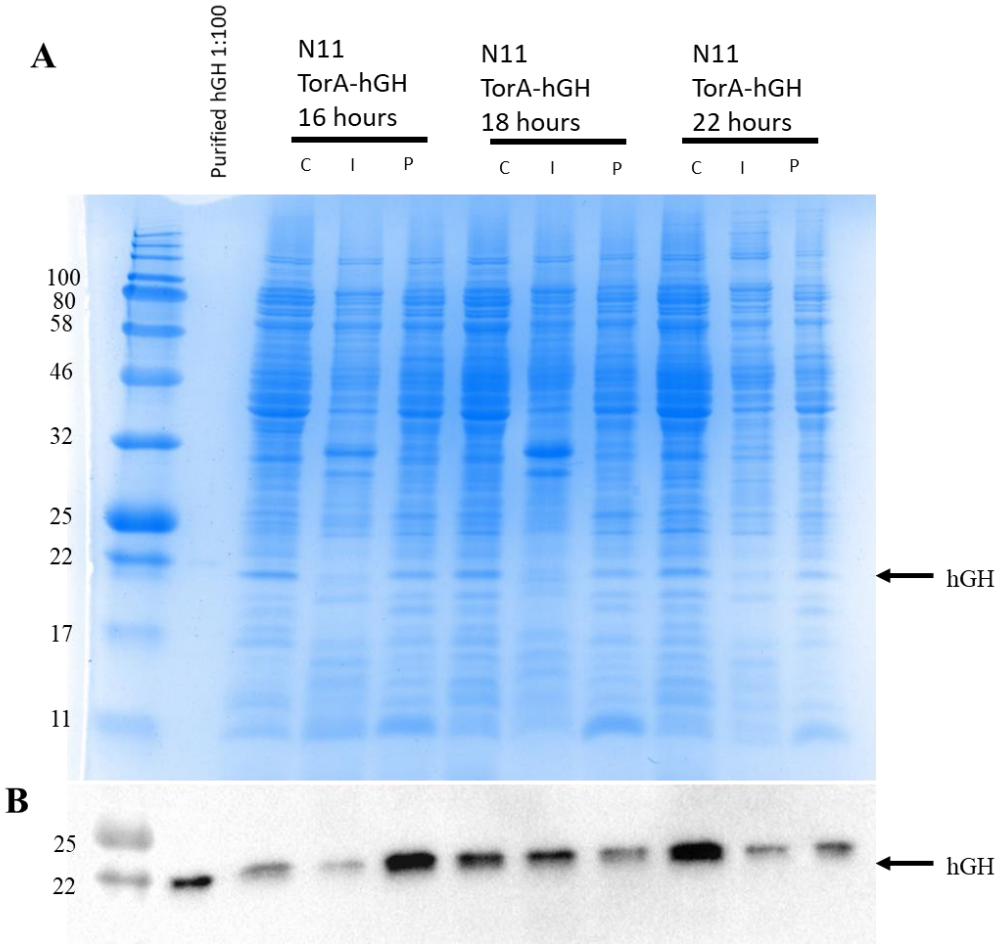


Figure 44: **TorA-SP-hGH expressed in N11 throughout fermentation in SM6 medium using p*narG*-CC.** The cell line N11 containing the p*narG*-CC TorA-SP-hGH (21 kDa, arrowed) plasmid were grown in a fed-batch fermenter containing SM6. Cell samples were taken after the induction with 20 mM sodium nitrate. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. The fractions were immunoblotted using an anti-hGH antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blots is 2.2 seconds.

In Figure 44 different fractionated samples of a fed-batch fermentation with N11 cells are shown. In the Coomassie stained SDS-PAGE gel numerous bands above 55kDa are visible in the periplasmic fraction. This is an indicator for cytoplasmic contamination of the periplasm during the fractionation protocol. Because of the

cytoplasmic contamination in the periplasmic samples in Figure 44 it cannot be shown that the hGH was exported to the periplasm by the Tat pathway. It is likely that hGH is exported into the periplasm in the sample taken 16 hours post-induction because the strongest signal occurs in the periplasmic fraction. However, we cannot say this for certain as the cells appear more fragile and a more gentle fractionation procedure is required.

E. coli strain Tn3.1 expresses more Tat proteins. We therefore decided to use strain Tn3.1 to express TorA-SP-hGH from the *narG*-CC promoter to determine whether the number of Tat translocases in strain N11 was the factor limiting hGH production. However, the data (Figure 45) indicate that this was not the case as no was hGH detected in any of the cytoplasmic, insoluble, and periplasmic fractions.

Tn3.1 cannot express p*narG*-CC TorA-SP-hGH into the periplasm in a fed-batch fermentation process and no hGH could be shown in the cytoplasm and insoluble fraction (Figure 45A and Figure 45B).

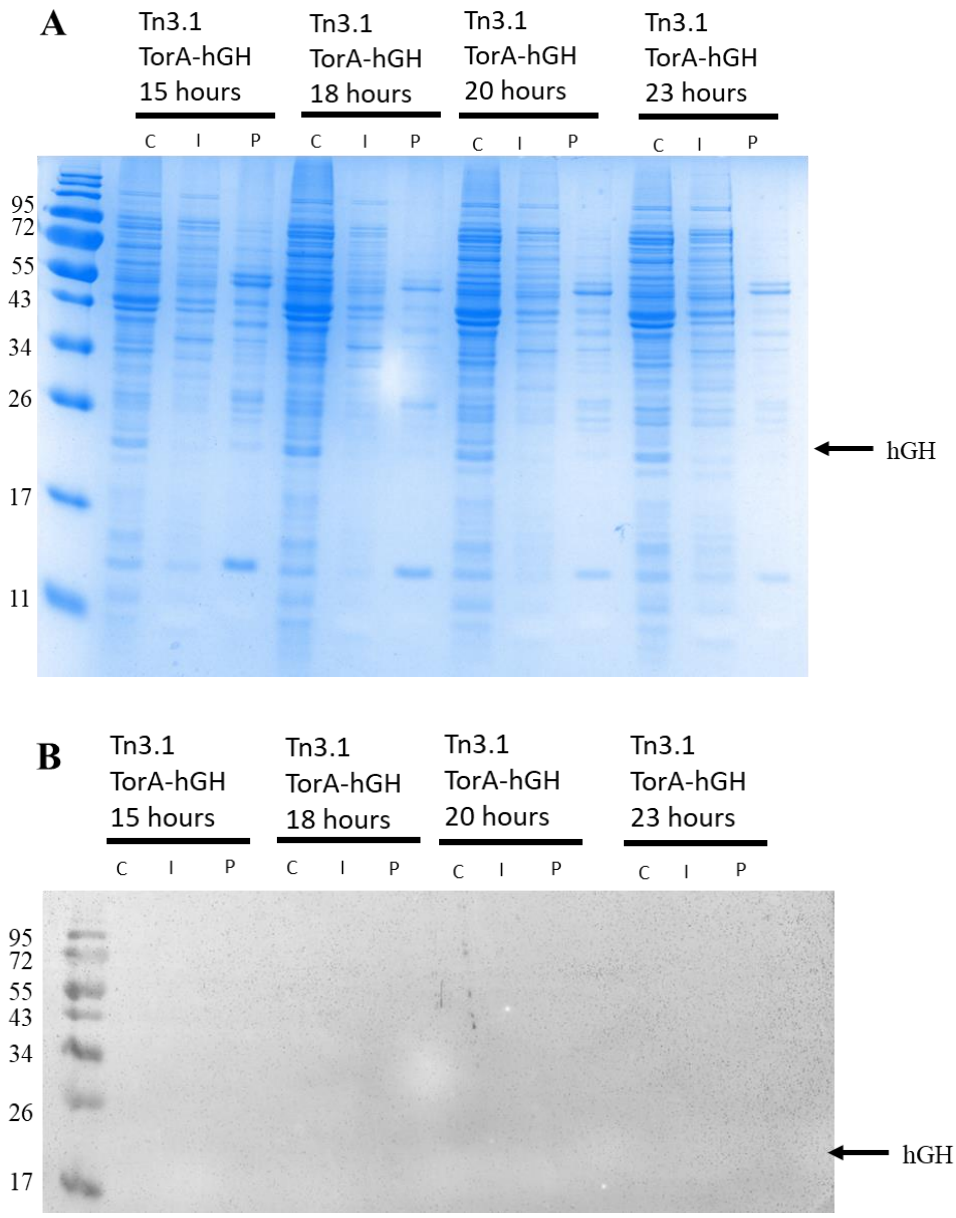


Figure 45: **TorA-SP-hGH expressed in Tn3.1 throughout fermentation in SM6 medium using the *narG*-CC promoter.** The cell line Tn3.1 containing the p*narG*-CC TorA-SP-hGH plasmid (21 kDa, arrowed) were grown in a fed-batch fermenter containing SM6. Cell samples were taken after the induction with 20 mM sodium nitrate and were fractionated. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. The fractions were immunoblotted using an anti-hGH antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blots is 2.2 seconds.

4.2.7. TorA-SP-scFv cannot be exported to the periplasm using p*narG*-CC

To test another protein that is relevant to pharmacology, we used p*narG*-CC to express TorA-SP-scFv with p*narG-CC* at shake flask scale. Whilst TorA-SP-scFv was expressed in SM6, it is not exported into the periplasm (Figure 46B) and the protein is not visible as an additional protein band on the Coomassie stained SDS-PAGE gel (Figure 46A). However, the signal of the sample using the *tac* promoter is far stronger than the signal of any of the p*narG*-CC samples. Furthermore, in the western blot in Figure 46 there is a double band in the insoluble fraction. This is most likely caused by proteolytic clipping as it is described for scFv by Alanen *et al*. (Alanen *et al.*, 2015).

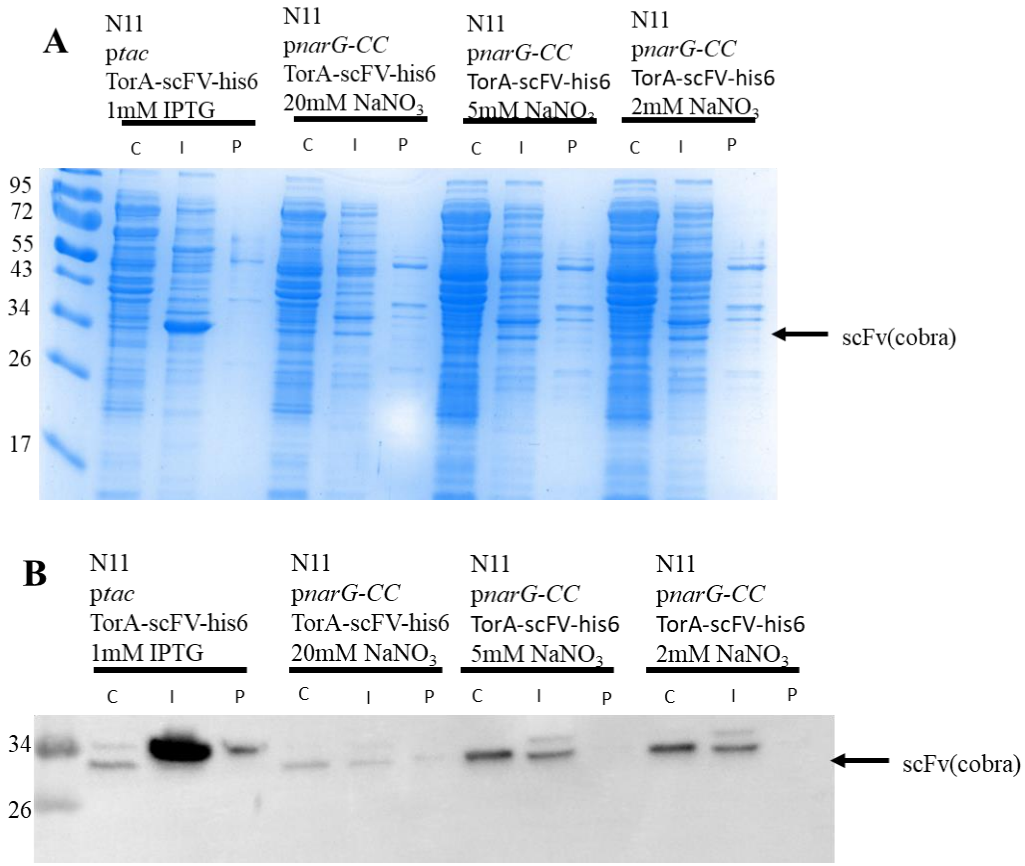


Figure 46: **Expression of p*narG*-CC TorA-SP-scFv with varying inducer concentrations.** Expression of TorA-SP-scFv(cobra) (29.5 kDa, arrowed) was conducted in N11 in SM6 medium at 30˚C in 50mL shake flask. Both plasmids were expressed for 3 hours after reaching an OD_{600} of 0.5. **A** shows the Coomassie stained SDS-PAGE gel and **B** shows the western blot. The fractions were immunoblotted using a C-terminal anti-his antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blots is 2.2 seconds.

4.3. Discussion

The p*narG*-CC promoter can be induced with a single dose of the sodium nitrate because the concentration of the inducer remains constant throughout growth (Figure 32). As shown previously, the p*narG*-CC promoter leads to similar levels of GFP during fed-batch fermentation as when expressed from the p*tac* promoter when the glycerol feed rate is at 0.15 µL/min. However, expression of GFP in fed-batch fermentation is not tightly controlled and further experiments would have been required to test whether tightly controlled expression can be achieved. In future experiments, we could have adjusted the pH with NaOH during the fermentation process and additionally check the nitrate levels in the SM6 medium.

Expression from p*narG*-CC can be tightly controlled at shake flask scale in MS medium and SM6 variant, but not during fermentation with the same medium. This suggest that there is some nitrate in the medium due to the presence of ammonium used to adjust the pH of the medium. Furthermore, N11 and Tn3.1 are both strains that are lacking the *narG* genes and the consequences for cell growth and metabolism of these deletions are not known.

Additionally, p*narG*-CC expresses similar amounts of GFP as p*tac* does in minimal salt media and is able to express higher amounts of GFP in a variant of SM6 medium. We show here that the glycerol feed impacts the ability on the yield and that the growth correlates strongly with the quantity of GFP expressed. Depending on the cell density reached during a fed-batch fermentation process p*narG*-CC could outperform the *tac* promoter.

p*narG*-CC can also be used to express TorA-SP-hGH, which is successfully exported to the periplasm in LB in small scale expression experiments. Regulated expression has not been achieved in MS medium in shake flask. Additionally, in shake flask

experiments in SM6 medium no hGH could be detected in any of the fractions. This may be related to the different growth conditions of the cells in different media. Furthermore, various amounts of sodium nitrate were tested in shake flask experiments to establish whether there are differences in the protein levels (Figure 42). We can show that TorA-SP-hGH is not present before induction and appears to be tightly controlled compared to GFP and that at levels of 5mM to 30mM sodium nitrate TorA-SP-hGH is being expressed into the periplasm (Figure 42). In fed-batch fermentation conditions, expression of TorA-SP-hGH was achieved. However, the Coomassie stained SDS-PAGE gel showed cytoplasmic contamination in the periplasm fraction. Both N11, which is *narG* deletion strain, and Tn3.1, which is a *narG* deletion strain overexpressing the Tat system, show cytoplasmic contamination in the periplasm after the fractionation (Figure 44).

To determine whether the data was protein specific we analysed the production of another industrially relevant protein, namely TorA-SP-scFV. Here we found that TorA-SP-scFV was expressed but was not exported to the periplasm (Figure 46).

Whilst the *narG*-CC promoter shows good expression for GFP it does not lead to the export of TorA-SP-hGH into the periplasm under fermentation conditions without cytoplasmic contamination in the periplasmic fraction (Figure 44). However, TorA-SP-hGH can be exported successfully into the periplasm in LB medium in shake flask (Figure 42).

The *narG*-CC promoter is good for a strong expression of proteins such as GFP and can reduce expression costs due to the low price of its inducer. In further experiments, other targets such as superoxidase dismutase (SOD1) and R16, could be tested in regard to their periplasmic export. Spectrin R16 is a three-helix bundle protein which is crucial for the integrity of plasma membranes (Glyakina *et al.*, 2018) and SOD1 is

an antioxidant which catalyses the dismutation of superoxide radicals (Younus, 2018). Furthermore, it could be investigated whether the promoter together with other signal peptides from preproteins such as prePhoD, could improve periplasmic export.

5. Production and export of PETase by the Tat pathway

5.1. Introduction

Poly (ethylene) terephthalate (PET) is the most commonly used plastic worldwide and is typically found in many household products, such as drinking bottles, salad packaging, shampoo, and window cleaner. PET is a thermoplastic which consists of single carbon or heteroatomic chains and is fully recyclable. However, in Europe 52% of PET is recycled and, in the U.S., only 31% (PET Resin Association, 2015). Considering that in 2017 worldwide 30.3 million tons of PET were produced, this shows a need for better recycling procedures (Plastics Insight, 2019). Less than 10% of the plastic used worldwide is currently recycled and traces of plastic were found in deep sea and other remote areas. Novel ways of recycling this plastic have to be established. Enzymatic degradation of plastic with PET-hydrolases, such as FsC from *Fusarium solani pisi*, was studied but the enzymatic activity was low and in incubation studies only 5% of PET was degraded after 96 hours (Ronkvist *et al.*, 2009; Groß *et al.*, 2017).

Throughout a 5-year monitoring of bottle recycling yards, a research group found a bacterium that is capable of degrading PET to utilise it as a carbon source. In 2016, Yoshida *et al*. published their research showing that the bacterium *Ideonella sakaiensis* is capable of producing an enzyme called PETase (Yoshida *et al.*, 2016). The enzyme breaks down PET into intermediate mono (2-hydroxyethyl) terephthalic acid (MHET) and in a secondary step MHETase breaks down MHET to two monomers (ethylene glycol and terephthalic acid) (Bornscheuer, 2016). PETase allows *Ideonella sakaiensis* to use PET as its major carbon and energy source (Austin *et al.*, 2018). MHET is the first intermediate to produce PET and currently has to be

synthesized. This is an exciting discovery as the enzymatic activity of PETase is 88-fold higher than the FsC-hydrolase (Yoshida *et al.*, 2016).

Although, PETase has only been recently discovered, its structure and function are well studied. Structural studies have shown that PETase has a disulphide bond that is predicted to be essential for the enzymatic activity and the protein is predicted to be a Tat substrate in *I. sakaiensis* (Fecker *et al.*, 2018; Huang *et al.*, 2018; Liu *et al.*, 2018). Therefore, exporting the protein via the Tat pathway would be potential advantageous. Previously, PETase was expressed in *Bacillus subtilis* via the Sec and Tat pathway. The most promising expression was shown with the *I. sakaiensis* native signal peptide SP_{PETase}. (Huang *et al.*, 2018) Furthermore, they achieved best expression in a Δ*tat* deletion strain of *Bacillus subtilis* and the research group reasoned that the Tat translocation components have an inhibitory effect on the secretion of PETase when fused to the native signal peptide. The PETase signal peptide is predicted to be a Tat signal peptide, however it does not have the Tat motif which may lead to a failure to transport through the Tat pathway.

In this study, we aim to investigate whether the plastic-degrading enzyme, PETase, can be exported by the Tat pathway in *E. coli*. PETase is a protein with significant industrial interest, which has previously been shown to be difficult to export in an active state in high concentrations (Papadopoulou, Hecht and Buller, 2019). Tat-mediated export could allow export of correctly folded substrate which makes it a useful pathway to use for PETase. To achieve Tat export, we have investigated the activity of four different signal peptides targeting the Tat pathway. The signal peptides used are AmiC, HyaA, TorA, and PhoD*N. Additionally, we have compared protein expression of PETase in CyDisCo strains that can generate disulphide bonds in the cytoplasm. a general expression plasmid. We aimed to purify the exported protein to

do further studies and evaluate the enzymatic activity. Furthermore, we aimed to upscale the protein production using fed-batch fermentation.

5.2. Results

5.2.1. PETase is not expressed at high levels using the TorA signal peptide

The TorA signal peptide has shown good activity for exporting non-native proteins to the periplasm of *E. coli*, with proteins such as GFP and hGH (Bernhardt and De Boer, 2003; Alanen *et al.*, 2015). Here, we used the TorA signal peptide to target PETase to the Tat pathway using the *tac* promoter expressing PETase in the pEXTII plasmid. Expression studies using the TorA signal peptide were done at 30˚C, 25 ˚C and at 20˚C in shake flask using LB media. In the 25˚C expression experiment samples were taken three hours post induction, and after overnight expression (approximately 16 hours post-induction) as shown in Figure 47. Samples were fractionated into the cytoplasmic (C), insoluble (I), and periplasmic (P) fractions. The TorA signal peptide does not appear to support PETase export and no protein could be visualised on the western blot as shown in Figure 47B or on the Coomassie stained SDS-PAGE gel (Figure 47A).

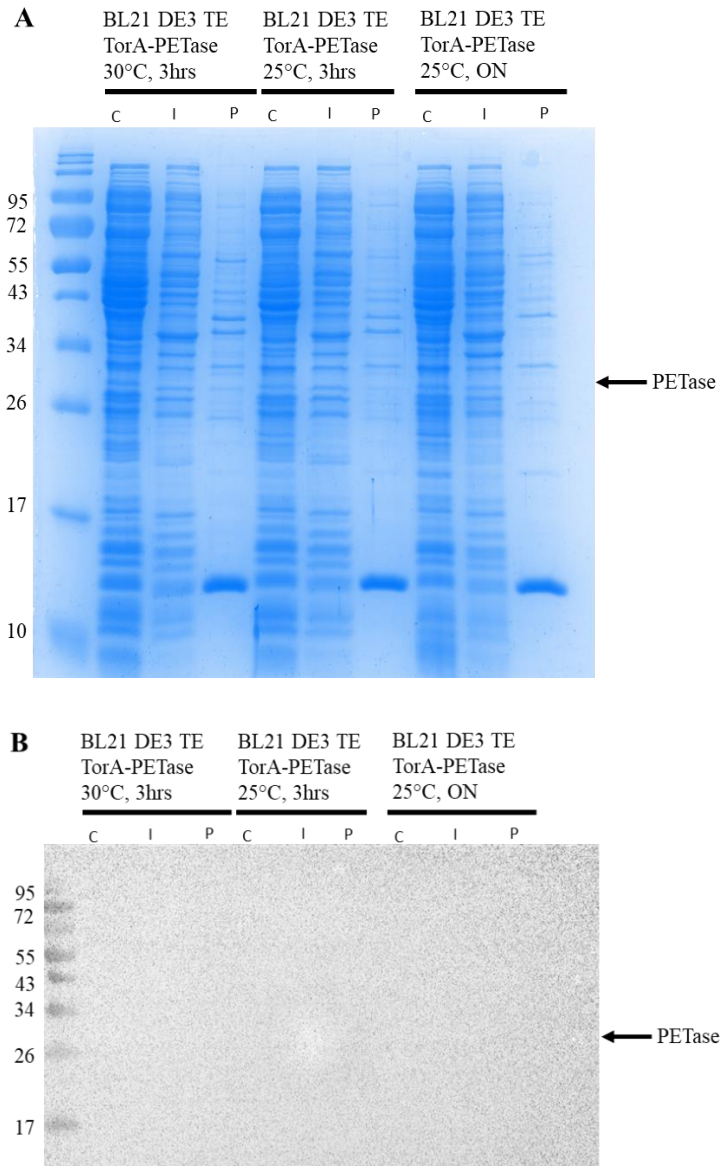


Figure 47: **Comparison of the expression of TorA-SP-PETase in *E. coli* strain BL21 DE3 TE at different temperatures and different incubation times.** PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strain BL21 DE3 TE in 50 mL shake flasks cultures at different temperatures in LB. Samples were taken 3 hours and 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Samples were blotted using a C-terminal anti-his antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 3 seconds.

Furthermore, expression at 20˚C was tested. Lower temperatures can improve folding and therefore can allow a protein to be expressed if it is degraded (Weickert *et al.*, 1996).

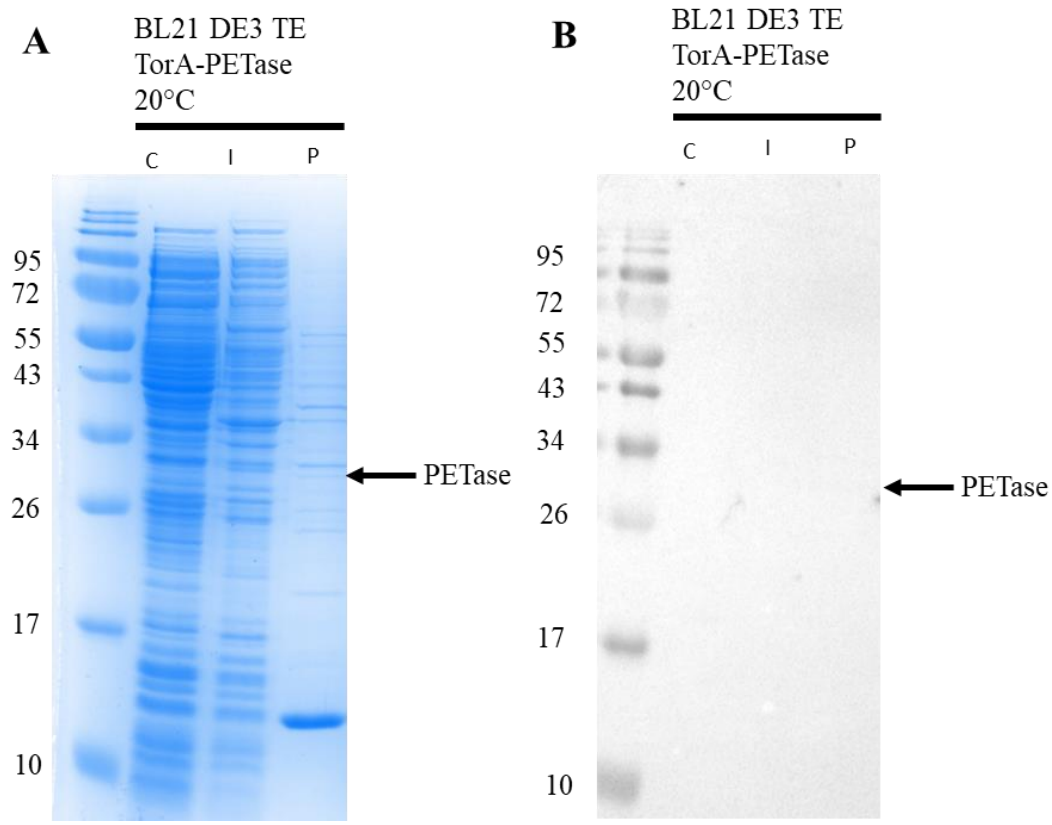


Figure 48: **TorA-SP-PETase does not express at 20˚C.** PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strain BL21 DE3 TE in 50 mL shake flasks cultures at 20˚C in LB. Samples were taken 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. All samples were blotted using a C-terminal anti-his antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 3 seconds.

TorA-SP-PETase is also not expressed at 20˚C as the Coomassie stained SDS-PAGE gel in Figure 48A and the western blot in Figure 48B show. Due to the lack of expression in any of the experiments no further expression experiments were conducted using the TorA signal peptide. It appears that the TorA signal peptide is not suitable for the substrate and the protein of interest might be degraded immediately.

5.2.2. PETase is successfully expressed using the HyaA signal peptide

The next signal peptide that was tested for export is HyaA. HyaA is also a Tat specific signal peptide. In the first experiment, we compared expression at two different temperatures, 25˚C and 30˚C.

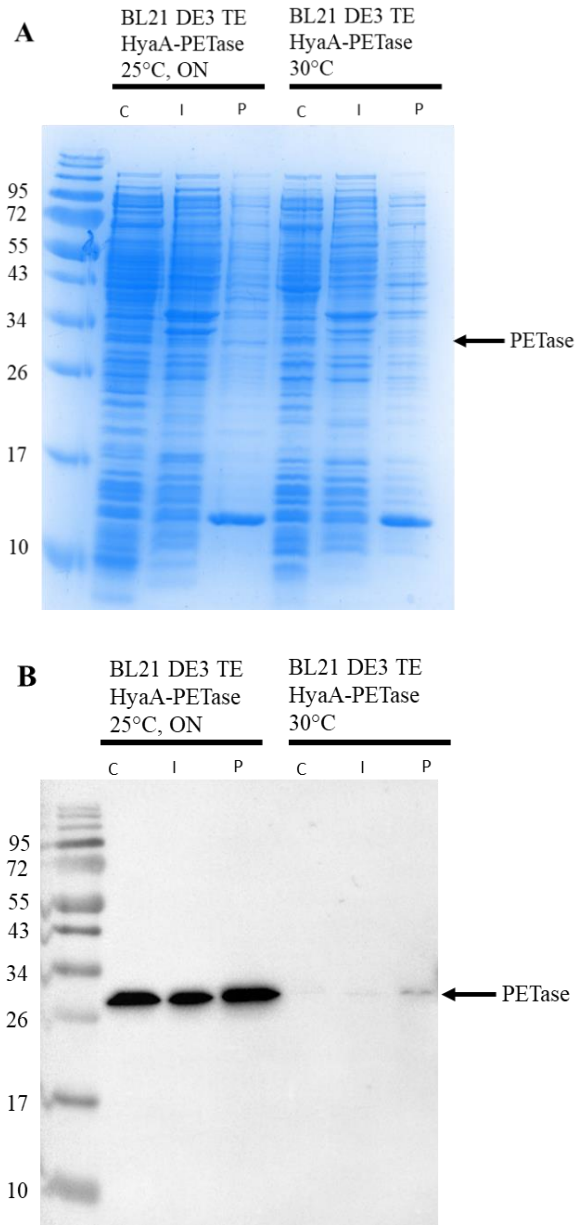


Figure 49: **The influence of temperature on the production of HyaA-SP-PETase by *E. coli* strain BL21 DE3 TE.** PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strain BL21 DE3 TE in 50 mL shake flasks cultures at different temperatures in LB. Samples were taken 3 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Samples were blotted using a C-terminal anti-his antibody The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 3 seconds.

As shown in Figure 49B, there is new evidence that PETase is successfully exported into the periplasm using the signal peptide HyaA. The export shows a higher efficiency at 25˚C than at 30˚C and at 25˚C PETase is found in all cell fractions. PETase is probably one of the most abundant bands in the periplasm. However, to fully confirm this further controls are needed. Additionally, PETase appears the same size in all fractions. This could be related to the clipping of the signal peptide or that the protein is mature size because it is degraded in the cytoplasm. This has been previously observed by Alanen *et al.* in 2015 (Alanen *et al.*, 2015).

At 30˚C the his-tagged substrate only shows a weak signal in the periplasmic fraction. To show that the export is efficient a Coomassie stained SDS-PAGE gel (Figure 49A) has been done. Through a Coomassie stained SDS-PAGE gel it can be shown whether the periplasmic fraction is contaminated with cytoplasmic proteins. Numerous bands above 55 kDa are a sign of cytoplasmic contamination in the periplasm. Despite the low temperature expression, the cells show faint signs of contamination in the periplasm on the Coomassie stained SDS-PAGE gel which could be related to lysis during the fractionation process or during the expression (Figure 49A).

In the next step we repeated this experiment at 20˚C to see whether lowering the temperature further can improve protein expression and prevent stress on the cells.

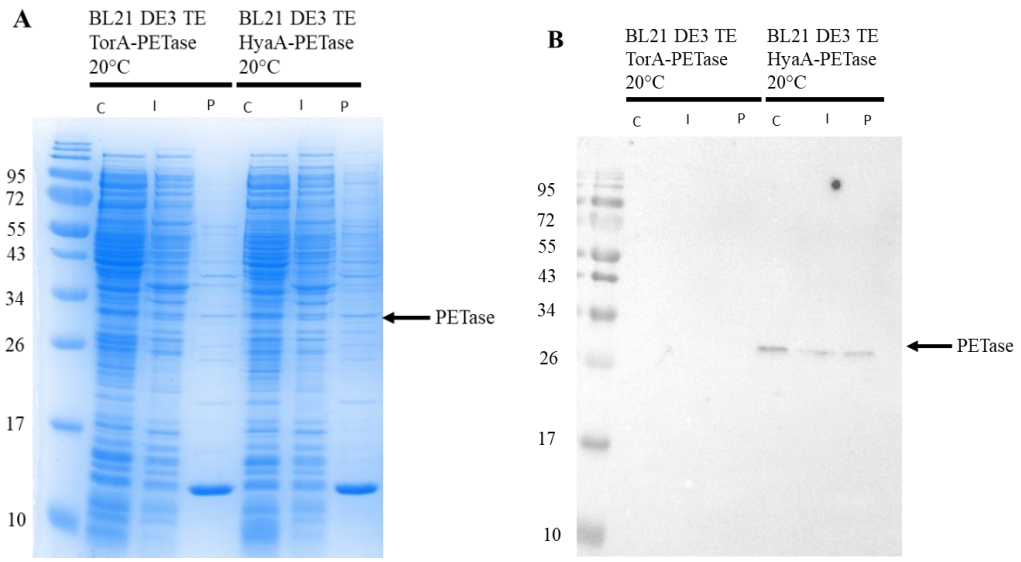


Figure 50: **PETase is synthesised at 20˚C in low concentrations using the HyaA signal peptide.** PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strain BL21 DE3 TE in 50 mL shake flasks cultures at 20˚C in LB. Samples were taken 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Also shown here is the TorA-SP-PETase expression which was shown previously. Also shown here is the repeat of the TorA-SP-PETase expression shown previously. Samples were blotted using a C-terminal anti-his antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 3 seconds.

Expression at 25˚C appears to be optimal. Decreasing the temperature to 20˚C resulted in a faint signal in all fractions (Figure 50B) probably due to the lower temperature. A longer incubation time could have led to an accumulation of more protein. At 30˚C the protein appears to be degraded. Also shown here is the TorA-SP-PETase expression which was shown previously in Figure 48. The SDS gel could not be repeated as separate gels due to Covid-19 lockdown. Further experiments on this signal peptide will be conducted at 25˚C as it results in the highest protein

concentration. Additionally, a construct was tested containing PDI and ErV1p (CyDisCo components) on the plasmid. Both components when co-expressed with PETase support cytoplasmic disulphide bond formation (Matos *et al.*, 2014). PETase has a disulphide bond that is predicted to be essential for the enzymatic activity of PETase (Fecker *et al.*, 2018; Liu *et al.*, 2018). The CyDisCo components could support the cytoplasmic disulphide bond formation and might improve the protein yields.

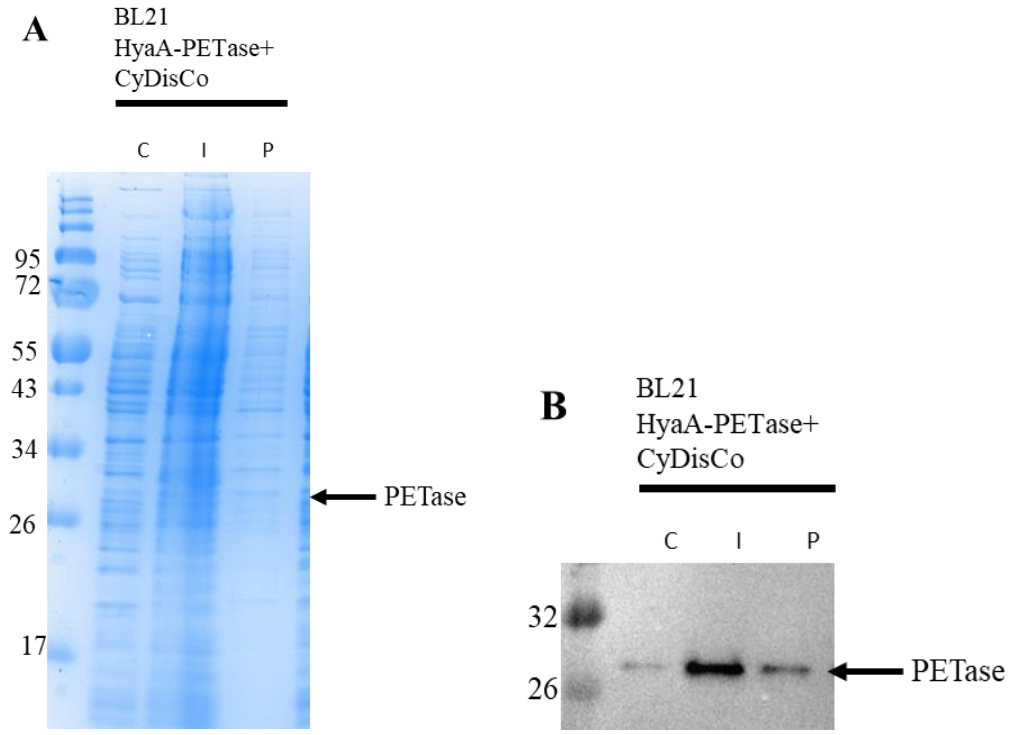


Figure 51: **Expression of HyaA-SP-PETase with CyDisCo components.** PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strain BL21 in 50 mL shake flasks cultures at 25˚C in LB. Samples were taken 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Samples were blotted using a C-terminal anti-his antibody. The gel is a representative of two gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 3 seconds.

HyaA-SP-PETase with CyDisCo components is successfully expressing PETase as shown in Figure 51B. However, the majority of protein appears to be in the insoluble fraction. The construct was expressed in BL21 and in a further experiment, expression in BL21 DE3 'TatExpress' (TE) is planned. 'TatExpress' *E. coli* strain contains an additional promoter to overexpress the *tatABCD* operon and yields could be increased using this strain (Browning *et al.*, 2017). The experiment has previously been conducted in BL21 DE3 TE cells, but no export could be confirmed due to high levels of cytoplasmic proteins in the periplasmic fraction. The result is not shown here because the experiment was not repeated. Therefore, it needs to be repeated to investigate if export can be improved or the overexpression of Tat membrane proteins and PETase leads to the lysis of the cells.

As the cells in the previous experiments lysed, we wanted to verify that PETase was exported by the Tat pathway. Therefore, we expressed the construct in MC4100 Δ*tat* cells.

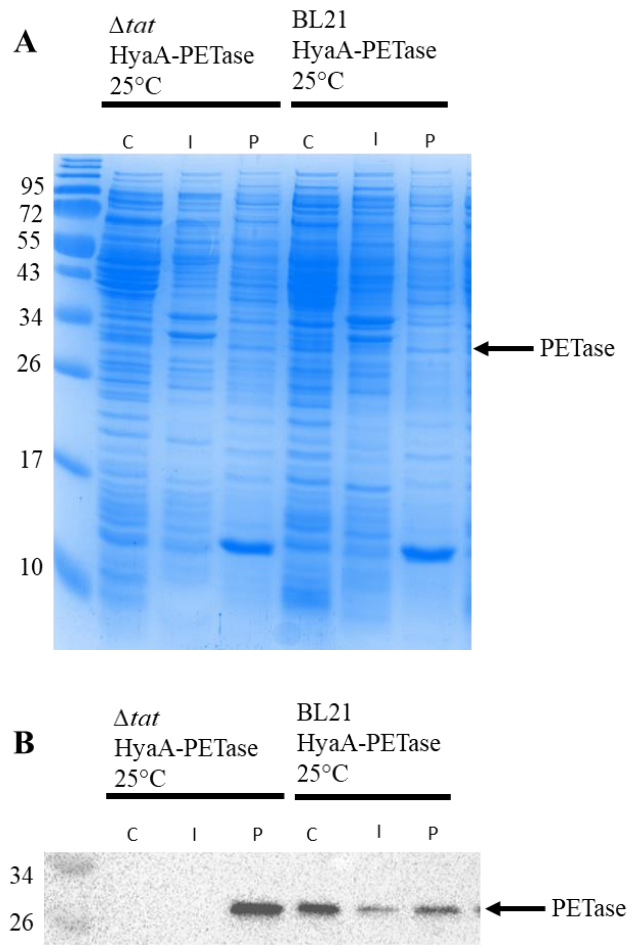


Figure 52: **Synthesis of HyaA-SP-PETase in a MC4100 Δ*tat* (*tatABCDE*) and BL21.** PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strain BL21 in 50 mL shake flasks cultures at 25˚C in LB. Samples were taken 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Samples were blotted using a C-terminal anti-his antibody. The gel is a representative of three gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 3 seconds.

In Figure 52A, the periplasmic fraction shows a high cytoplasmic contamination. Synthesis of HyaA-SP-PETase causes the cells to fractionate poorly. Interestingly, the

export is much more efficient in the Δ*tat* strain (Figure 52B) suggesting that PETase is exported by Sec.

Furthermore, we can show that the choice of media has an influence on the expression. In Figure 53A and B the Coomassie stained SDS-PAGE gel and western blot of a shake flask expression of HyaA-SP-PETase in LB and TB media shown. TB medium is a phosphate buffered medium which supports growth due to a higher glycerol content as carbon source and the phosphate buffer prevents radical pH changes.

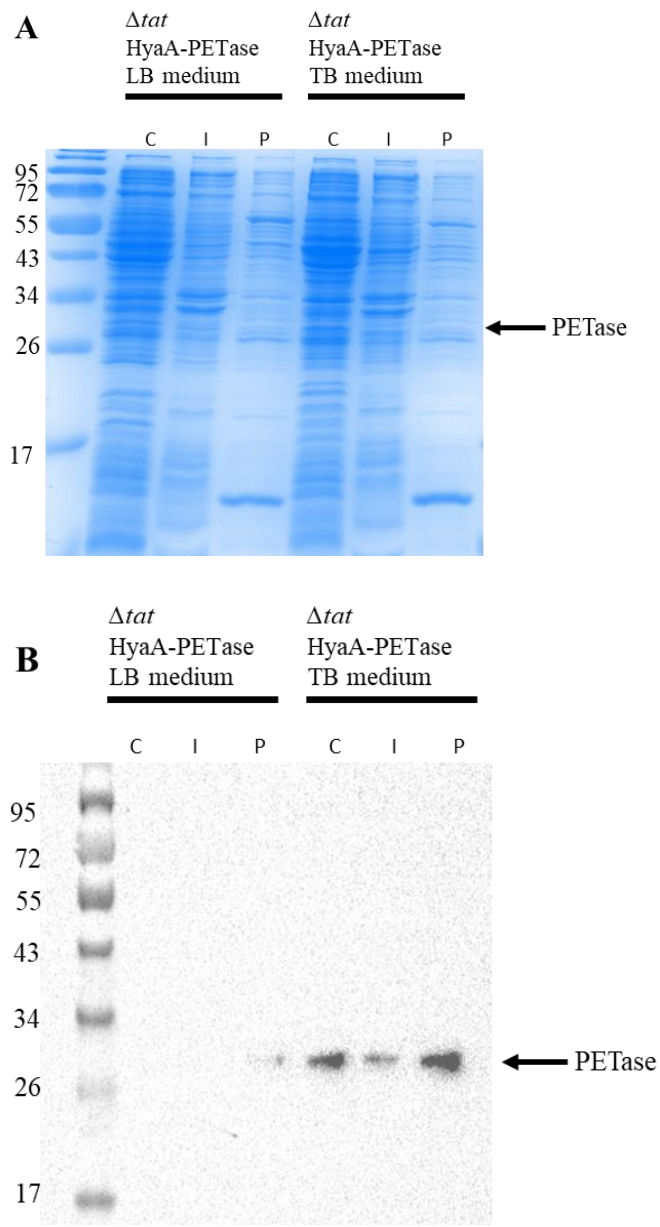
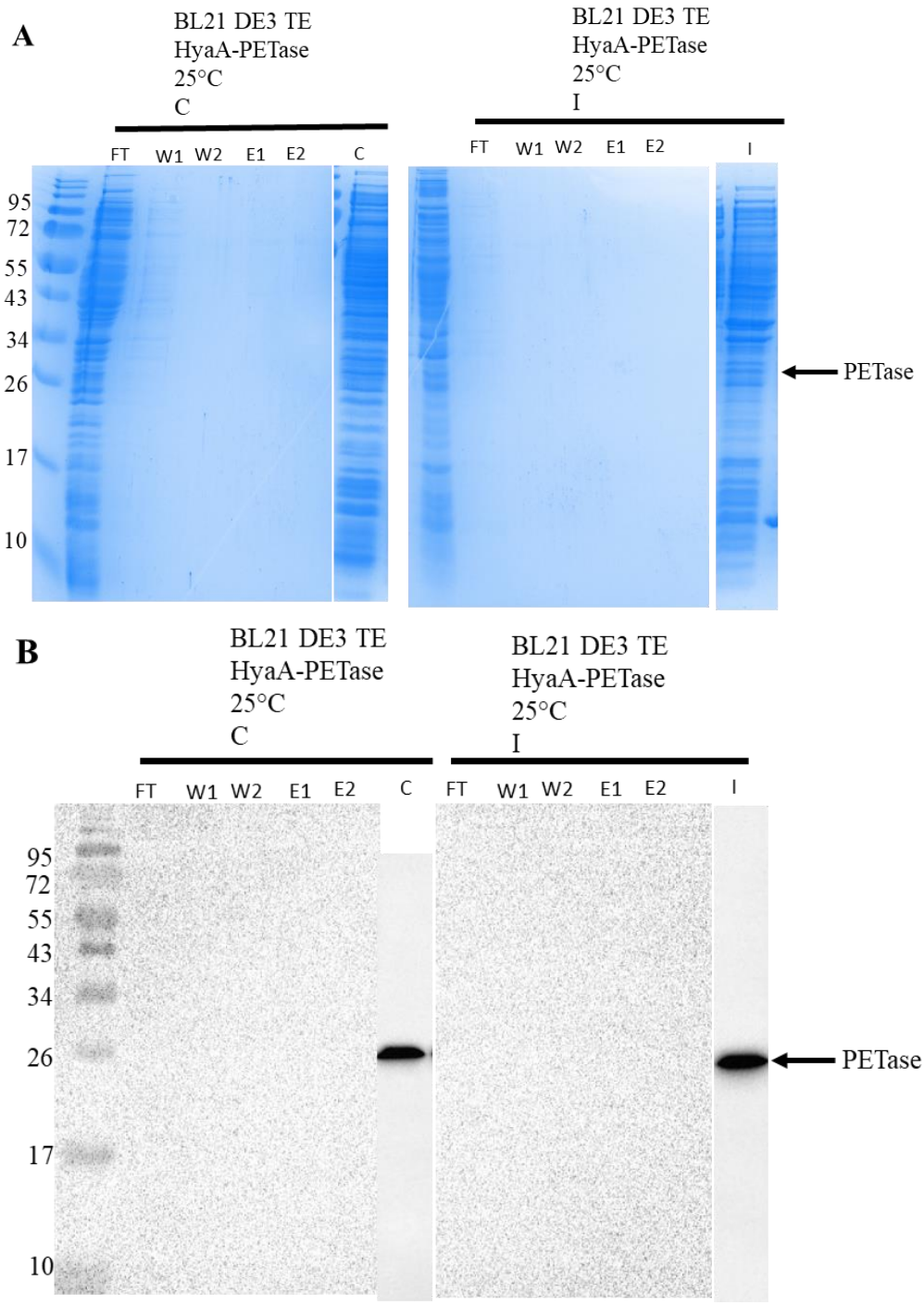


Figure 53: **Comparison of expression of HyaA- SP-PETase in Δ*tat* in different media**. PETase (32 kDa, arrowed) was expressed in the strain Δ*tat* in 400 mL shake flasks cultures in LB and TB of a pEXTII plasmid using the *tac* promoter. Samples were taken 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Samples were blotted using a C-terminal anti-his antibody. The gel is a representative of three gels and 10 µL sample have been loaded in each well. The exposure time of the western blot is 20.6 seconds.

The western blot in Figure 53B shows that TB medium can facilitate protein expression and improves expression compared to LB medium.

5.2.3. PETase could not be purified using a bench Ni-NTA spin column

In the next step, we aimed to purify the his-tagged protein from Δ*tat* HyaA-SP-PETase periplasmic sample and from all fractions of the BL21 DE3 TE HyaA-SP-PETase samples grown at 25˚C, as shown in Figure 49. Here a Ni-NTA spin column kit was chosen, which is described in the ‘Materials and Methods’ section.


A
BL21 DE3 TE
HyaA-PETase
25°C
C
BL21 DE3 TE
HyaA-PETase
25°C
I
FT W1 W2 E1 E2 C
FT W1 W2 E1 E2 I
95
72
55
43
34
26
17
10
PETase
B
BL21 DE3 TE
HyaA-PETase
25°C
C
BL21 DE3 TE
HyaA-PETase
25°C
I
FT W1 W2 E1 E2 C
FT W1 W2 E1 E2 I
95
72
55
43
34
26
17
10
PETase

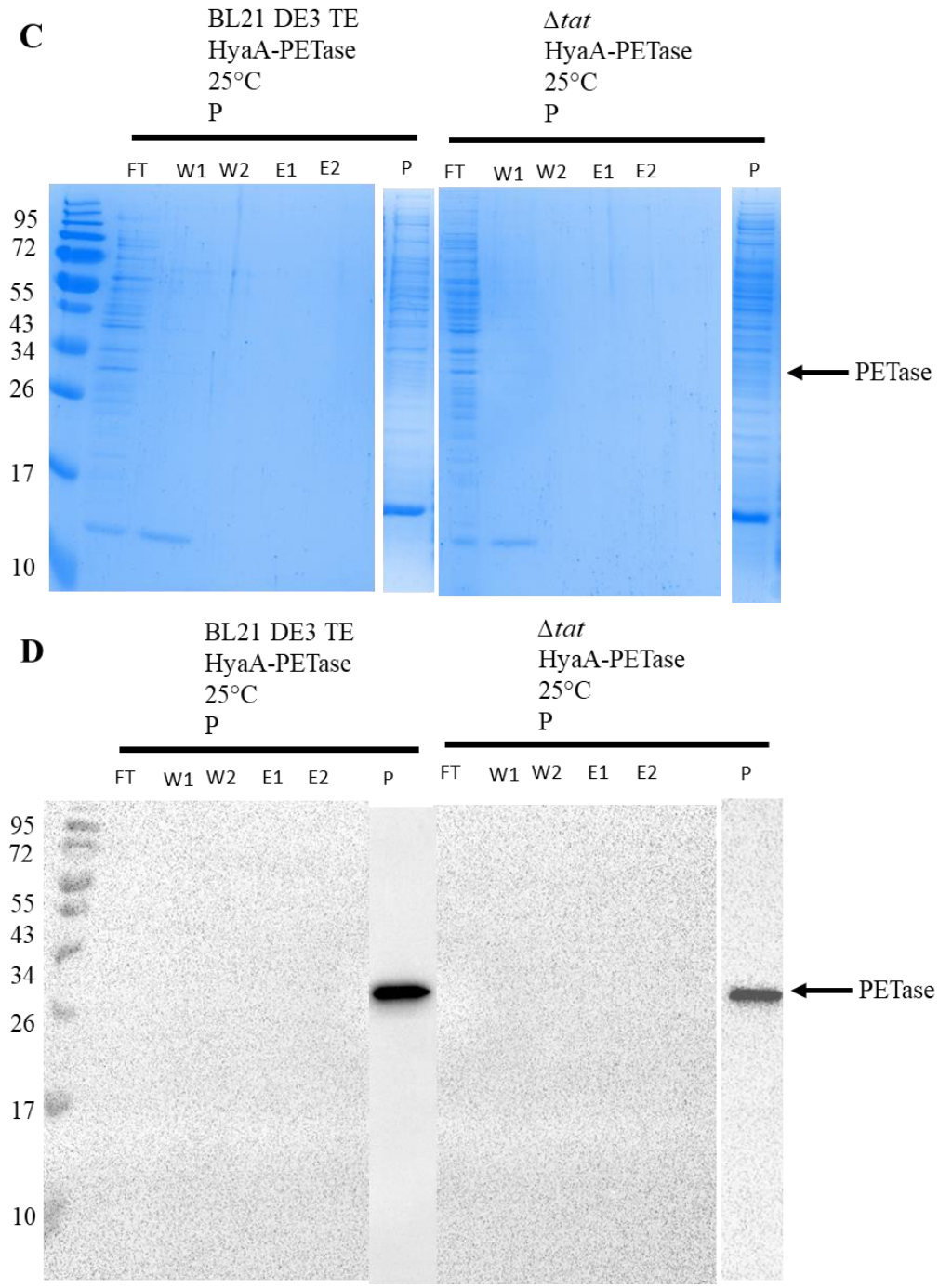

C
BL21 DE3 TE
HyaA-PETase
25°C
P
Δtat
HyaA-PETase
25°C
P
FT W1 W2 E1 E2 P
FT W1 W2 E1 E2 P
95
72
55
43
34
26
17
10
PETase
D
BL21 DE3 TE
HyaA-PETase
25°C
P
Δtat
HyaA-PETase
25°C
P
FT W1 W2 E1 E2 P
FT W1 W2 E1 E2 P
95
72
55
43
34
26
17
10
PETase

Figure 54: **Bench top purification of PETase containing samples. A** and **C** show the Coomassie stained SDS-PAGE gel **B** and **D** show the western blot. Samples shown in previous images contained PETase in fractions. These samples are shown next to the purified sections, however, have not been loaded onto the same SDS-PAGE gel. To purify these PETase containing samples, we loaded the samples on a Ni-spin purification column and then blotted the different fractions (cytoplasmic, C; insoluble, I; periplasmic, P) with a C-terminal anti-his antibody. Samples are separated in the flow through (FT) section, the first wash (W1), the second wash (W2), the first elution (E1), and the second elution (E2). The size of PETase is 32 kDA. Samples were blotted using a C-terminal anti-his antibody. The gels are a representative of one gel and 10 µL sample have been loaded in each well. The exposure time of the western blots are 3 seconds.

Throughout the purification process samples are collected from the flow through (FT), the two wash steps (W1 and W2), and the elution steps (E1 and E2). Typically, the purified protein is expected to be in the elution fractions if it is expressed. Therefore, samples have been loaded onto an SDS-PAGE gel separately to analyse whether PETase was expressed. However, the protein could not be eluted in any of the fractions (Figure 54) and the purification kit used is not suitable for the purification of our construct.

5.2.4. PETase is being expressed using the AmiC signal peptide both with and without disulphide bond supporting components

AmiC is a promiscuous signal peptide and can target substrates to the Tat pathway as well as the Sec pathway (Tullman-Ercek *et al.*, 2007). Expression studies were done using the AmiC signal peptide at 30˚C and at 25˚C. In the 25˚C expression experiment

samples were taken three hours post induction, and after overnight expression (approximately 16 hours).

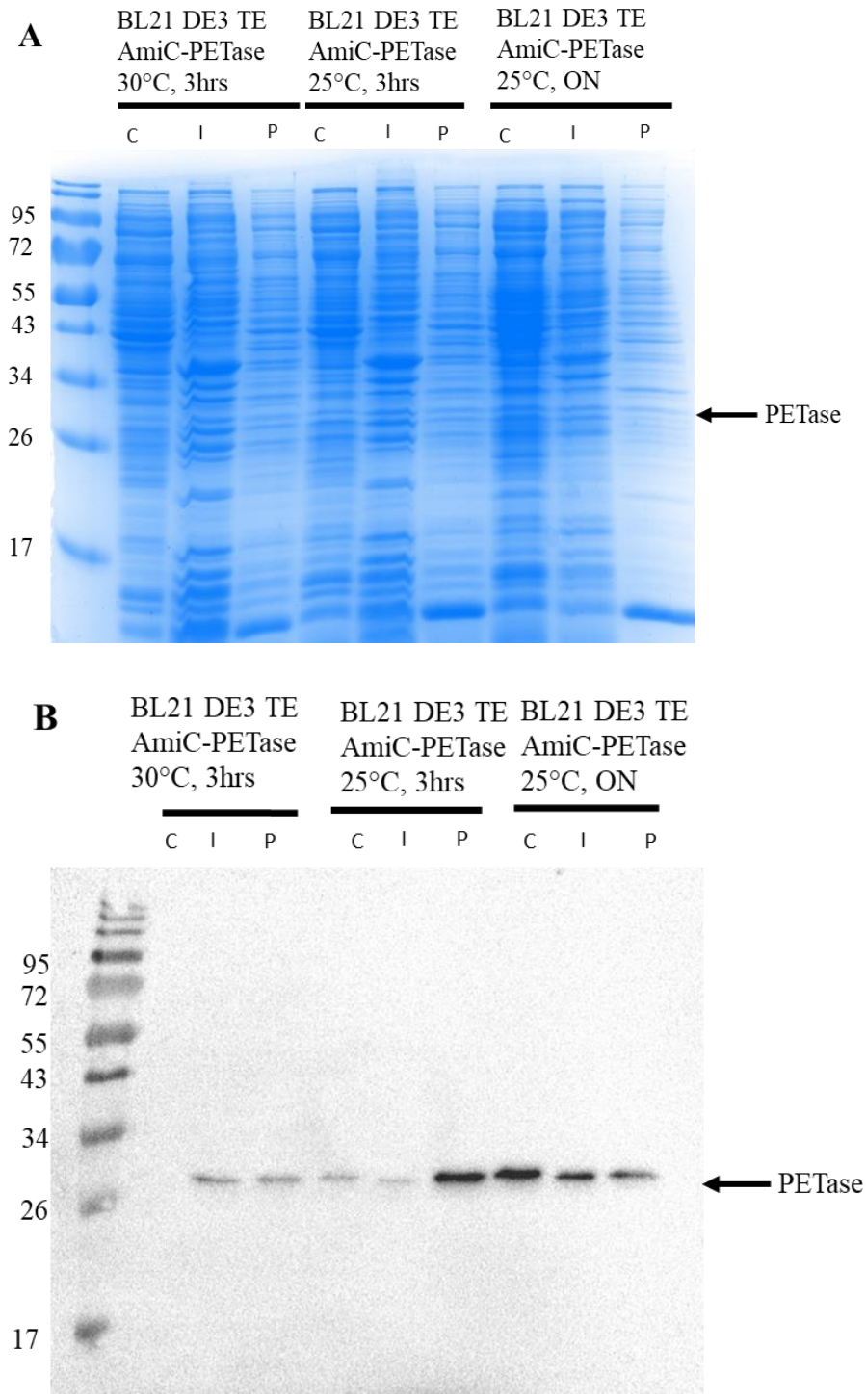

A
BL21 DE3 TE
AmiC-PETase
30°C, 3hrs
BL21 DE3 TE
AmiC-PETase
25°C, 3hrs
BL21 DE3 TE
AmiC-PETase
25°C, ON
C I P C I P C I P
95
72
55
43
34
26
17
PETase
B
BL21 DE3 TE
AmiC-PETase
30°C, 3hrs
BL21 DE3 TE
AmiC-PETase
25°C, 3hrs
BL21 DE3 TE
AmiC-PETase
25°C, ON
C I P C I P C I P
95
72
55
43
34
26
17
PETase

Figure 55: **Comparison of expression of AmiC-SP-PETase in BL21 DE3 TE at different temperatures**. PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strain BL21 in 50 mL shake flasks cultures at different temperatures in LB. Samples were taken 3 hours and 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Samples were blotted using a C-terminal anti-his antibody. The gels are a representative of two gel and 10 µL sample have been loaded in each well. The exposure time of the western blots are 3 seconds.

Figure 55B shows that PETase is being expressed using the AmiC signal peptide. However, the fractions could not be clearly separated, and the periplasmic fraction shows a high amount of proteins above 55 kDa which is a sign of cytoplasmic contamination in the periplasm (Figure 55A). The cells appear to be more prone to lysis and a gentler fractionation protocol could have been used.

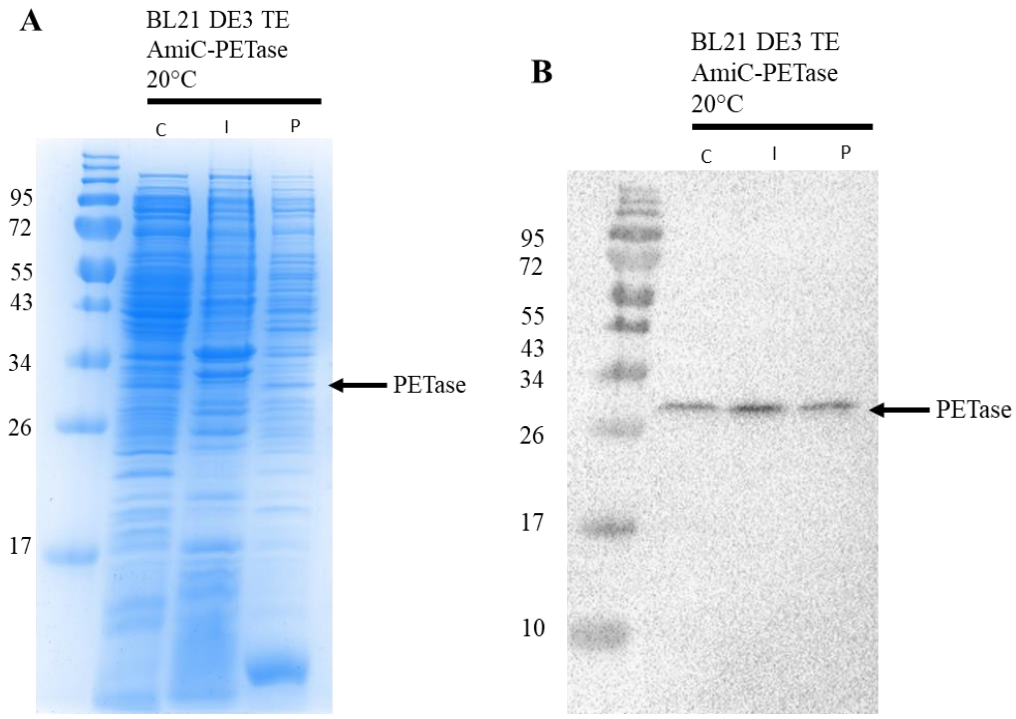


Figure 56: **AmiC-SP-PETase is expressed at lower temperatures.** PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strain BL21 DE3 TE in 50 mL shake flasks cultures at 20˚C in LB. Samples were taken 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Samples were blotted using a C-terminal anti-his antibody. The gels are a representative of two gel and 10 µL sample have been loaded in each well. The exposure time of the western blots are 3 seconds.

Further experiments on the AmiC signal peptide will be conducted at 25˚C.

A construct supporting cytoplasmic disulphide bond formation and containing the AmiC signal peptide has also been tested.

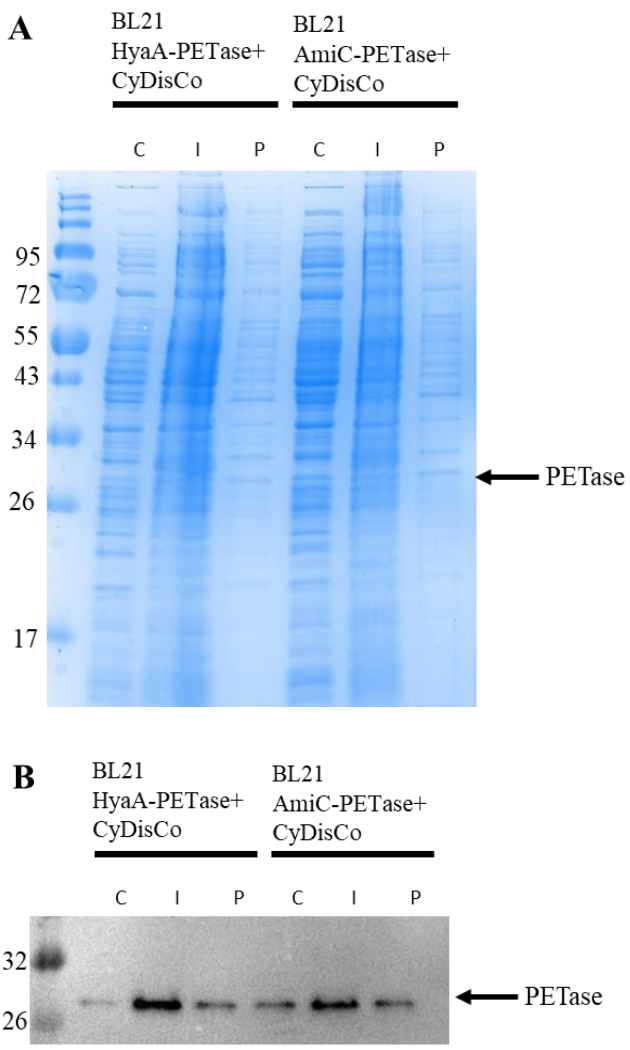


Figure 57: **Expression of HyaA-SP-PETase with CyDisCo components and AmiC-SP-PETase with CyDisCo components.** PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strain BL21 in 50 mL shake flasks cultures at 25˚C in LB. Samples were taken 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. The gels are a representative of two gel and 10 µL sample have been loaded in each well. The exposure time of the western blots are 3 seconds.

In Figure 57A and Figure 57B, the expression of AmiC-SP-PETase with CyDisCo components is presented. The fractions show that the periplasmic fraction is contaminated with cytoplasmic proteins (Figure 57A).

The same experiment was conducted in BL21 DE3 TE cells, but no export could be achieved due to extreme lysis (data not shown). Ideally this experiment would have been repeated to investigate whether the lysis is related to the synthesis of PETase or related to the fractionation process. To investigate, if AmiC-SP-PETase export is Tat mediated, we have expressed the AmiC-SP-PETase construct in MC4100 Δ*tat*.

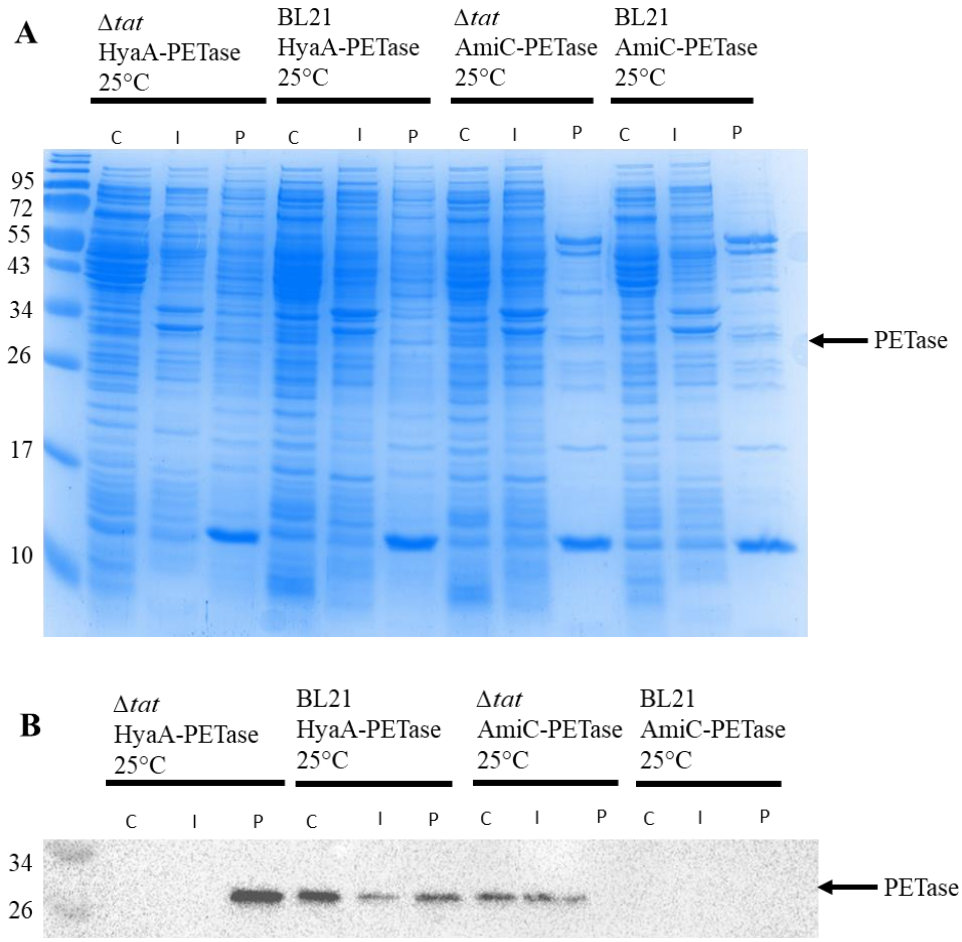


Figure 58: **Expression in Δ*tat* to determine whether expression is Tat mediated.** PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strain BL21 and Δ*tat* in 50 mL shake flasks cultures at 25˚C in LB. Samples were taken 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Focus in this image is the AmiC-SP-PETase expression and the blot would need repeating to show only AmiC expression. This was not possible to do due to the Covid-19 lockdown. The Δ*tat* strain does not contain *tatABCDE* and therefore if the protein found in the periplasm in Δ*tat* it would have been expressed via the Sec pathway. Samples were blotted using a C-terminal anti-his antibody. The gels are a representative of three gel and 10 µL sample have been loaded in each well. The exposure time of the western blots are 3 seconds.

Figure 58A shows that the cells expressing HyaA-SP-PETase are showing more proteins in the periplasm than the cells expressing AmiC-SP-PETase. Also, no export into the periplasm occurs in a Δ*tat* strain which proves that AmiC leads to a translocation of PETase by the Tat pathway (Figure 58B). Using the signal peptide HyaA does not lead to export by the Tat pathway and protein is found in the periplasmic fraction when expressed in a Δ*tat* strain (Figure 58B).

5.2.5. PETase cannot be expressed using PhoD*N as a signal peptide in initial experiments

PhoD is a Tat specific signal peptide which is found in *Bacillus subtilis*. The here used signal peptide has a mutation of alanine in position 51 to asparagine. Therefore, it is called PhoD*N.

The cloning of the construct containing the PhoD*N signal peptide was started before the Covid-19 lockdown and due to the research laboratory restrictions only a single expression trial could be conducted. The here presented data is preliminary and ideally repeats would have been conducted.

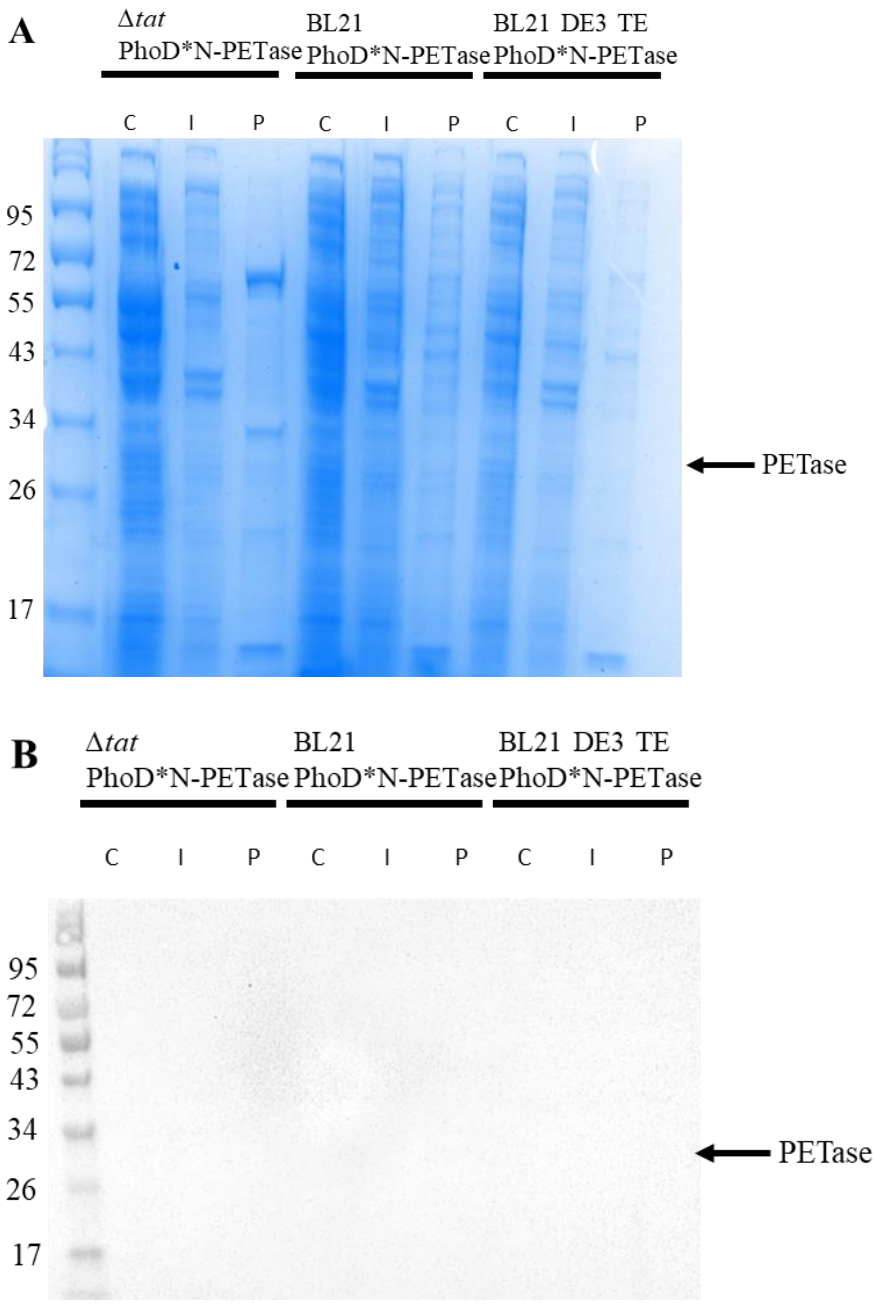


Figure 59: **Expression of PhoD*N-SP-PETase.** PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strains Δ*tat*, BL21, and BL21 DE3 TE in 50 mL shake flasks cultures at 25˚C in LB. Samples were taken 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Samples were blotted using a C-terminal anti-his antibody. The gel is a representative of one gel and 10 µL sample have been loaded in each well. The exposure time of the western blots are 3 seconds.

Figure 59A and Figure 59B show no signs of expression using the signal peptide PhoD*N in different cell lines.

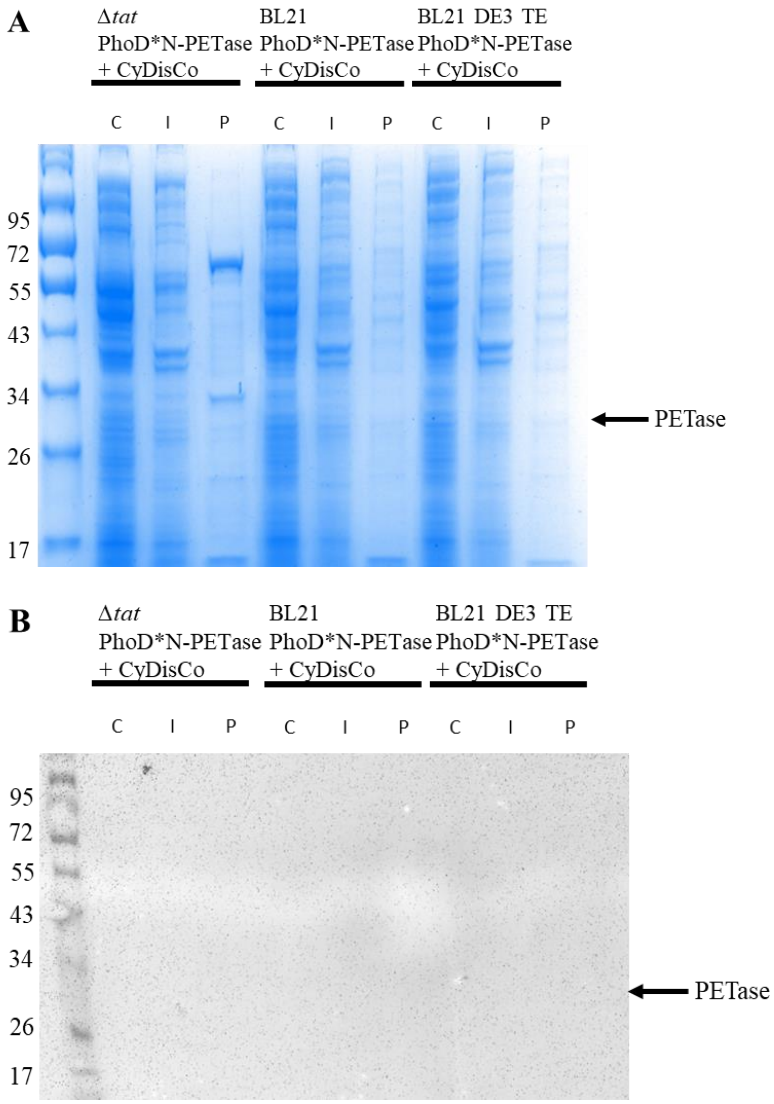


Figure 60: **Expression of PhoD*N-SP-PETase with CyDisCo components.** PETase (32 kDa, arrowed) was expressed of a pEXTII plasmid using the *tac* promoter in the strains Δ*tat*, BL21, and BL21 DE3 TE in 50 mL shake flasks cultures at 25˚C in LB medium. Samples were taken 16 hours post-induction and were separated into fractions. Cells were induced with 100 µM IPTG. **A** shows the Coomassie stained SDS-PAGE gel **B** shows the western blot. Samples were blotted using a C-terminal anti-his antibody. The gel is a representative of one gel and 10 µL sample have been loaded in each well. The exposure time of the western blots are 3 seconds.

PhoD*N-SP-PETase with CyDisCo does also not show any expression at 25˚C as the Coomassie stained SDS-PAGE gel in Figure 60A and the western blot in Figure 60B show.

5.2.6. Further experiments that had been planned

Based on the results achieved before Covid-19 forced the University to close the research laboratories, further experiments had been planned. The signal peptides HyaA and AmiC were proven to successfully express PETase in the constructs containing the CyDisCo components and the constructs without CyDisCo components. Therefore, fed-batch experiments are planned to increase the sample volume for purification. Fed-batch fermentation was planned on BL21 HyaA-SP-PETase with CyDisCo, BL21 AmiC-SP-PETase with CyDisCo, BL21 DE3 TE HyaA-SP-PETase, and BL21 DE3 TE AmiC-SP-PETase. The periplasmic fraction of those samples was planned to be purified on a HisTrap column. A goal of this research was to investigate whether *E. coli* can express active PETase and which yields that can be achieved. Due to the lockdown, we could not progress to the activity assays which would have included incubating PET with the purified protein and analyse the product with HPLC.

5.3. Discussion

The four signal peptides that were used in this study have previously shown good Tat export. In this study, the signal peptides TorA and PhoD*N could not achieve export of PETase. However, the signal peptides HyaA and AmiC could achieve promising export into the periplasm. When using the AmiC signal peptide the cells appear to lyse. We can show that PETase is expressed by the Sec pathway when the signal peptide HyaA is used (Figure 52). PETase is not exported in the periplasm when the signal peptide AmiC is used (Figure 58) when using a strain that is lacking the Tat pathway.

We could not achieve sufficient expression at 30˚C with HyaA and AmiC. PETase was shown to be heat labile at 40˚C and expression at 30˚C might be too close to the heat labile point (Huang *et al.*, 2018). If a high concentration of the protein is toxic to the cells, expression at lower temperature might beneficiary to reduce cell stress and to improve protein expression. Expression at 25˚C achieved the most promising export.

The majority of expression experiments throughout this study, showed some signs of cytoplasmic contamination in the periplasmic fraction of the cells. This was determined by comparing the amount of proteins above 55 kDa visualised in the periplasmic fraction on a Coomassie stained SDS-PAGE gel. *I. sakaiensis*, which is a gram-negative, aerobic bacteria found in samples from a bottle recycling yard in Japan, secretes PETase into the extracellular milieu (Huang *et al.*, 2018). However, it is interesting that the cells are more affected by stress when the signal peptide HyaA was used compared to the signal peptide AmiC. This might be related to the export machinery used during the export with the different signal peptides. It appears that

cells are prone to cell stress when the Sec system is used when the signal peptide HyaA is used for targeting PETase into the periplasm. However, we cannot determine whether the lysis occurred during the expression or during the fractionation and it may be that the cells expressing HyaA-SP-PETase are frailer and therefore require a more gentle fractionation method.

Comparing export in Figure 49 and in Figure 52 (ideally would be compared on the same blot) the 'TatExpress' strain appears to improve export of HyaA-SP-PETase. Due to the lockdown the samples could not be shown on the same western blot.

The periplasmic band in the HyaA-SP-PETase expression using Δ*tat* suggests that PETase is exported by the Sec pathway. This experiment would be interesting to repeat using the construct HyaA-SP-PETase with CyDisCo to see how the folding of the protein affects the export.

Furthermore, we tried Ni-NTA spin column purification on the PETase samples with the HyaA signal peptide and we could not elute the protein. Further purification experiments are required optimising the conditions for the elution of the protein.

The expression of the plasmids additionally containing the CyDisCo components shows that the majority of the protein is in the membrane fraction. There is periplasmic export and it would be exciting to investigate whether the purified protein of this samples is more active than the protein samples that were expressed without the CyDisCo components.

Different research groups have been expressing PETase using different Tat and Sec specific signal peptides. Successful expression has been reported in *E. coli* and *Bacillus subtilis* (Huang *et al.*, 2018; Seo *et al.*, 2019). However, the reported yields

are between 3 and 6.2 mg/L. In this study, we could show efficient export of PETase using the signal peptide HyaA and AmiC which has not been shown before. Furthermore, this is worth pursuing because of the low yields reported in the literature (Huang *et al.*, 2018; Seo *et al.*, 2019).

6. Final discussion

Many proteins are produced in mammalian cells such as CHO cells. While they are able to express complex proteins that require disulphide bond formation, CHO cultures are also costly to maintain (Ranade, 2010; Waegeman and Soetaert, 2011). Scientists are continuously searching for ways to decrease the costly factors involved in industrial protein production. *E coli* is still a preferred expression host and the expression of proteins in *E. coli* is a crucial part of the landscape of biopharmaceutical production systems (Waegeman and Soetaert, 2011; Mohamed N. Baeshen, 2014). While wildtype microorganisms are limited in their capabilities to export complex proteins containing cofactors or disulphide bonds, microorganisms can be engineered to export such proteins (Lee *et al.*, 2009; Thomason *et al.*, 2014). In Gram negative bacterial host systems, expression and secretion into the periplasm is preferred because the protein is more accessible during downstream processing and purification processes. This is a further cost benefit of microorganisms (Quax, 1997).

The Tat pathway could offer a solution to the expression problem in microorganisms as proteins exported by the Tat pathway are exported into the periplasm in a fully folded state (Barrett *et al.*, 2003; Panahandeh *et al.*, 2008). Native Tat substrates often contain cofactors and require complex folding (Palmer, Sargent and Berks, 2005). Furthermore, the Tat pathway has its own inbuilt proofreading mechanism which, despite being poorly understood, helps to maintain the quality of the secreted products (Robinson and Bolhuis, 2004; Gangl *et al.*, 2015). The proofreading mechanisms allows only correctly folded proteins to export into the periplasm. However, it often rejects proteins of interest. Overcoming the proofreading mechanism and being able to exploit the Tat pathway could enable *E. coli* to export proteins up to 150 kDa in size (Delisa, Tullman and Georgiou, 2003; Palmer and Berks, 2012).

In this study, we have investigated different approaches to facilitate cost effective protein expression in *E. coli*. The first investigated approach is genomic engineering an alternative *tat* operon (TatABC *A. tumefaciens*) into the *E. coli* genome (Chapter 3). Furthermore, we expressed proteins with an alternative promoter that is induced with sodium nitrate, which is comparatively cheap compared to established inducers such as IPTG (Chapter 4) to improve bacterial expression systems. Additionally, we tested if Tat signal peptides can enable a protein of major interest (PETase) to be expressed and exported in a functional state by the Tat pathway (Chapter 5).

TatABC *A tumefaciens* in the *E. coli* genome can export Tat native proteins and does not affect growth or phenotype.

Each known Tat system exports a variety of native proteins which differs between the different Tat systems and organisms the system occurs in. This suggests that the proofreading mechanism of the Tat systems may differ as well. Furthermore, it has been shown in the past that a non-functional Tat system in *E. coli* has a severe impact on the growth and cell membrane of the cells (Wexler *et al.*, 2000; Lee *et al.*, 2002). In this study, we showed that an alternative *tat* operon expressed on the *E. coli* genome can export *E. coli* native Tat proteins and does not affect the growth or phenotype compared to wildtype cells. Previously, alternative Tat systems were expressed from a second plasmid in *E. coli* and expression and growth has been impacted by the cell stress related to a two-plasmid expression (Frain *et al.*, 2017).

TatABC *A. tumefaciens* MC4100 mutants are not able to express proteins such as hGH, GFP, scFv(cobra), scFv(M), and co-translationally folded proteins (R16 and Protein G).

We have shown that *A. tumefaciens* Tat is capable of exporting TMAO reductase in a *E. coli* Δ*tat* strain. However, expression of non-native proteins has not been reported

(Oates *et al.*, 2003). We show here that *tat* operon is highly selective in its export of proteins and cannot express and export a variety of proteins such as hGH, GFP, single chain variable fragments, and co-translationally folded proteins such as R16 and Protein G. It does not provide an improvement for the biopharmaceutical production of complex proteins so far rejected by the Tat system.

To better understand the proofreading and rejection mechanisms of the Tat pathway further research would need to be done by inserting other *tat* operons onto the *E. coli* genome and further test the export capabilities. The differences in the *tat* operons could reveal important insight into the proofreading mechanism and provide an opportunity to find less selective Tat systems and be able to export complex and large proteins into the periplasm. Throughout the course of this study, further gene doctoring constructs have been designed and it would be exciting to insert them into the genome to investigate their abilities.

A sodium nitrate induced promoter (p*narG*-CC) can express high levels of GFP comparably to the commonly established p*tac* promoter (IPTG induced) in shake flask and fed-batch fermentation.

To further decrease costs of protein production in microorganisms we tested a sodium nitrate inducible promoter (p*narG*-CC). As previously noted, the commonly used inducer IPTG is costly and requires storage at -15 to -25˚C which may not be possible in countries entering the biopharmaceutical market. The promoter used here is a novel promoter system. Here we show that the promoter using sodium nitrate has high expression levels when expressing GFP in shake flask and fed-batch fermentation experiments. These are promising findings as regulating gene expression in native

promoters is challenging and designing new promoters requires complex screening of different promoter assemblies (Liu *et al.*, 2019).

The *narG*-CC promoter allows expression and secretion of hGH into the periplasm but not scFv.

We aim to express proteins into the periplasm because downstream processing is cheaper as the protein is more accessible for purification processes. Therefore, we investigated the ability of the promoter to mediate the export proteins into the periplasm by the Tat pathway. We showed that hGH is exported into the periplasm. However, this study also showed that scFv was not expressed into the periplasm. Overall, this promoter shows promising performance for protein expression.

Based on the findings in this study, future experiments also involve testing a larger spectrum of proteins and other signal peptides in combination with the *narG*-CC promoter and investigating its ability to encourage periplasmic export.

PETase expression is more efficient at 25˚C compared to 30˚C and 20˚C.

When expressing PETase in *E. coli*, we found that the temperature that the culture is expressed at has a huge impact on the expression efficiency. In previous studies by Huang *et al*. it has been shown that PETase is heat labile at 40˚C and expression at lower temperatures is preferred (Huang *et al.*, 2018). We can show here that growth at 25˚C shows the best expression compared to growth at 30˚C and 20˚C. Research published earlier this year shows that when PETase is expressed in *B. subtilis* the highest enzyme activity obtained when grown at 28˚C (Wang *et al.*, 2020). Further research into the activity of the here expressed protein is required to investigate the quality of the expressed protein.

PETase can be expressed and secreted into the periplasm using the signal peptide of HyaA (probably through the Sec pathway) but cells are lysed when using AmiC in a Tat system containing cell line.

Our research interest is directed to proteins from high industrial value such as the recently discovered esterase called PETase. This enzyme is capable of degrading polymers and has a high enzymatic activity compared to other proteins capable of degrading polymers (Ronkvist *et al.*, 2009; Groß *et al.*, 2017). Natively the protein is exported by the Tat pathway but despite all efforts expression in industrial preferred hosts is challenging (Huang *et al.*, 2018). It previously could not be exported by the Tat pathway in studies using *Bacillus subtilis*. However, it is expressed using a Tat signal peptide (Huang *et al.*, 2018).

We targeted PETase to the Tat pathway using the signal peptides TorA, HyaA, AmiC and PhoD*N and could show expression into the periplasm with HyaA. We show that signal peptides are a powerful for protein expression, but were not able to show Tat-specific expression with these signal peptides. However, due to the lockdown, no conclusion of the activity and the yield of the protein could be made.

The protein is exported into the extracellular milieu in *I. sakaiens*is and possibly could be toxic to the cells. Further experiments will need to be conducted to research whether the cells are quarantining the protein in lysosomes.

In this study we successfully express proteins in high yields for biopharmaceutical production in *E. coli* variants. We could not disable the Tat rejection mechanism through creating TatABC *A. tumefaciens* MC4100 mutant strains. However, we obtained high yields of protein using a novel promoter and we achieve promising

expression results using different Tat signal peptides for periplasmic export of a complex folded protein.

Research presented here regarding the enzyme PETase could be continued with fed-batch fermentation of the plasmids AmiC-SP-PETase with CyDisCo, AmiC-SP-PETase, HyaA-SP-PETase with CyDisCo, and HyaA-SP-PETase. Furthermore, in future experiments the protein needs to be purified and its activity needs to be determined. For this, PET could be incubated with the purified protein and the PET samples be analysed using scanning electron microscopy (SEM). If the protein is toxic to the cells an alternative promoter system could be used such as the *RHA* promoter which has been proposed to support expression of toxic proteins (Giacalone *et al.*, 2006).

7. References

Alami, M.; Lüke, I.; Deitermann, S.; Eisner, G; Koch, H.; Brunner, J.; Müller M. (2003) 'Differential interactions between a twin-arginine signal peptide and its translocase in *Escherichia coli*', *Molecular Cell*, 12(4), pp. 937–946. doi: 10.1016/S1097-2765(03)00398-8.

Alanen, H.; 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 a2b and antibody fragments to the periplasm by the *Escherichia coli* Tat pathway in the absence of prior disulfide bond formation', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1853(3), pp. 756–763. doi: 10.1016/j.bbamcr.2014.12.027.

Alcock, F.; Stansfeld, P.; Basit, H. Habersetzer, J.; Baker, M.; Palmer, T.; Wallace, M. I.; Berks, B. C. (2016) 'Assembling the Tat protein translocase', *eLife*, 5, pp. 1–28. doi: 10.7554/eLife.20718.

Alfano, J. R. and Collmer, A. (2001) 'Mechanisms of Bacterial Pathogenesis in Plants: Familiar Foes in a Foreign Kingdom', in *Principles of Bacterial Pathogenesis*, pp. 179–226. doi: 10.1016/b978-012304220-0/50006-6.

Austin, H. P.; Allen, M. D.; Donohoe, B. S.; Rorrer, N. A.; Kearns, F. L.; Silveira, R. L.; Pollard, B. C.; Dominick, G.; Duman, R.; Omari, K. E.; Mykhaylyk, V.; Wagner, A.; Michener, W. E.; Amore, A.; Skaf, M. S.; Crowley, M. F.; Thorne, A. W.; Johnson, C. W.; Lee Woodcock, H.; McGeehan, J. E.; Beckham, G. T. (2018) 'Characterization and engineering of a plastic-degrading aromatic polyesterase', *Proceedings of the National Academy of Sciences of the United States of America*, 115(19), pp. E4350–E4357. doi: 10.1073/pnas.1718804115.

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), pp. 1–9. doi: 10.1186/s12934-014-0141-0.

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', *Journal of Biological Chemistry*, 287(10), pp. 7335–7344. doi: 10.1074/jbc.M111.326355.

Balchin, D., Hayer-Hartl, M. and Hartl, F. U. (2016) 'In vivo aspects of protein folding and quality control', *Science*, 353(6294). doi: 10.1126/science.aac4354.

Bardwell, J. C. A., McGovern, K. and Beckwith, J. (1991) 'Identification of a protein required for disulfide bond formation in vivo', *Cell*, 67(3), pp. 581–589. doi: 10.1016/0092-8674(91)90532-4.

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', *FEBS Journal*, 276(1), pp. 232–243. doi: 10.1111/j.1742-4658.2008.06776.x.

Barrett, C. M. L.; Ray, N.; Thomas, J. D.; Robinson, C.; Bolhuis, A. (2003) 'Quantitative export of a reporter protein, GFP, by the twin-arginine translocation pathway in *Escherichia coli*', *Biochemical and Biophysical Research Communications*, 304(2), pp. 279–284. doi: 10.1016/S0006-291X(03)00583-7.

van den Berg, B..; W. M. Clemons Jr; Collinson, I.; Modis, Y.; Hartmann, E.; Harrison, S. C.; Rapoport, T. A. (2004) 'X-ray structure of a protein-conducting channel', *Nature* 427 pp. 36 - 44. doi: 10.1038/nature02218

Berks, B. C. (1996) 'A common export pathway for proteins binding complex redox

cofactors?', *Molecular Microbiology*, 22(3), pp. 393–404. doi: 10.1046/j.1365-2958.1996.00114.x.

Berks, B. C. (2015) 'The twin-arginine protein translocation pathway', *Annual Review of Biochemistry*, 84, pp. 843–864. doi: 10.1146/annurev-biochem-060614-034251.

Berks, B. C., Lea, S. M. and Stansfeld, P. J. (2014) 'Structural biology of Tat protein transport', *Current Opinion in Structural Biology*, 27(1), pp. 32–37. doi: 10.1016/j.sbi.2014.03.003.

Berks, B. C., Palmer, T. and Sargent, F. (2003) 'The Tat protein translocation pathway and its role in microbial physiology', *Advances in Microbial Physiology*, 47, pp. 187–254. doi: 10.1016/S0065-2911(03)47004-5.

Bernhardt, T. G. and De Boer, P. A. J. (2003) 'The *Escherichia coli* amidase AmiC is a periplasmic septal ring component exported via the twin-arginine transport pathway', *Molecular Microbiology*, 48(5), pp. 1171–1182. doi: 10.1046/j.1365-2958.2003.03511.x.

Bielser, J. M.; Wolf, M.; Souquet, J.; Broly, H.; Morbidelli, M. (2018) 'Perfusion mammalian cell culture for recombinant protein manufacturing – A critical review', *Biotechnology Advances*, 36(4), pp. 1328–1340. doi: 10.1016/j.biotechadv.2018.04.011.

Bittner, L. M., Arends, J. and Narberhaus, F. (2017) 'When, how and why? Regulated proteolysis by the essential FtsH protease in *Escherichia coli*', *Biological Chemistry*, 398(5–6), pp. 625–635. doi: 10.1515/hsz-2016-0302.

Blaudeck, N.; Kreutzenbeck, R.; 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), pp. 2811–

2819. doi: 10.1128/JB.185.9.2811-2819.2003.

de Boer, H. A., Comstock, L. J. and Vasser, M. (1983) 'The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters.', *Proceedings of the National Academy of Sciences of the United States of America*, 80(1), pp. 21–25. doi: 10.1073/pnas.80.1.21.

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), pp. 20213–20219. doi: 10.1074/jbc.M100682200.

Bornscheuer, U. T. (2016) 'Feeding on plastic A bacterium completely degrades poly(ethylene terephthalate)', *Science*, 351(6278), pp. 1154–1155. doi: 10.1016/0390-6035(79)90080-4.

Brosius, J., Erfle, M. and Storella, J. (1985) 'Spacing of the -10 and -35 regions in the *tac* promoter. Effects on its in vivo activity', *Journal of Biological Chemistry*, 260, pp. 3539–3541.

Browning, D. F., Richards, K. L., Peswani, A. R., Roobol, J., Busby, Stephen 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*, pp. 1–24. doi: 10.1002/bit.26434.

Browning, D. F., Butala, M. and Busby, S. J. W. (2019) 'Bacterial Transcription Factors: Regulation by Pick "N" Mix', *Journal of Molecular Biology*, 431(20), pp. 4067–4077. doi: 10.1016/j.jmb.2019.04.011.

Brundage, L.; Hendrick, J. P.; Schiebel, E.; Driessen, A. J. M.; Wickner, W. (1990) 'The purified *E. coli* integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation', *Cell*, 62, pp. 649–

657.

Burdette, L. A.; Leach, S. A.; Wong, H. T.; Tullman-Ercek, D. (2018) ‘Developing Gram-negative bacteria for the secretion of heterologous proteins’, *Microbial Cell Factories*, 17(1), pp. 1–16. doi: 10.1186/s12934-018-1041-5.

Cao, T. B. and Saier, M. H. (2003) ‘The general protein secretory pathway: Phylogenetic analyses leading to evolutionary conclusions’, *Biochimica et Biophysica Acta - Biomembranes*, 1609(1), pp. 115–125. doi: 10.1016/S0005-2736(02)00662-4.

Clare, D. K.; Vasishtan, D.; Stagg, S.; Quispe, J.; Farr, G. W.; Topf, M.; Horwich, A. L.; Saibil, H. R. (2012) ‘ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin’, *Cell*, 149(1), pp. 113–123. doi: 10.1016/j.cell.2012.02.047.

Clark, D. P. and Pazdernik, N. J. (2013) ‘Protein Synthesis’, in *Molecular Biology*. 2nd edn. Academic Press, pp. e250–e255. doi: 10.1016/b978-0-12-378594-7.00047-0.

Cline, K. (2015) ‘Mechanistic aspects of folded protein transport by the twin arginine translocase (Tat)’, *Journal of Biological Chemistry*, 290(27), pp. 16530–16538. doi: 10.1074/jbc.R114.626820.

Cohen, S. N.; Chang, A. C. Y.; Boyer, H. W.; Helling, R. B. (1973) ‘Construction of biologically functional bacterial plasmids in vitro’, *Proceedings of the National Academy of Sciences of the United States of America*, 70(11), pp. 3240–3244. doi: 10.1073/pnas.70.11.3240.

Cranford-Smith, T. and Huber, D. (2018) ‘The way is the goal: How SecA transports proteins across the cytoplasmic membrane in bacteria’, *FEMS Microbiology Letters*, 365(11), pp. 1–16. doi: 10.1093/femsle/fny093.

Cristóbal, S.; de Gier, J.; Nielsen, H.; von Heijne, G. (1999) ‘Competition between Sec- and TAT-dependent protein translocation in *Escherichia coli*’, 18(11), pp. 2982–2990.

Dalbey, R. E. and Kuhn, A. (2000) ‘Evolutionarily related insertion pathways of bacterial, mitochondrial, and thylakoid membrane proteins’, *Annual Review of Cell and Devlopmental Biology*, 16, pp. 51–87.

Darwin, A. J. and Stewart, V. (1996) ‘Nitrate and nitrite regulation of anaerobic gene expression’, in *Regulation of Gene Expression in Escherichia coli.* Springer, Boston, MA, p. pp 343-359.

Darwin, A. J. and Stewart, V. (1996) ‘The NAR Modulon Systems: Nitrate and Nitrite Regulation of Anaerobic Gene Expression’, in *Regulation of Gene Expression in Escherichia coli.* Boston: Springer, Boston, MA, pp. 343–359.

Delisa, M. P., Tullman, D. and Georgiou, G. (2003) ‘Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway’, *Pnas*, 100(10), pp. 6115–6120. doi: 10.1073/pnas.0937838100.

Dilks, K.; Rose, R. W.; Hartmann, E.; Pohlschroder, M. (2003) ‘Prokaryotic utilization of the twin-arginine translocation pathway: a genomic survey’, *Journal of Bacteriology*, 185, pp. 1478–1483.

Dilks, K. and Giménez, M. I. (2005) ‘Genetic and Biochemical Analysis of the Twin-Arginine Translocation Pathway in Halophilic Archaea Genetic and Biochemical Analysis of the Twin-Arginine Translocation Pathway in Halophilic Archaea’, 187(23), pp. 8104–8113. doi: 10.1128/JB.187.23.8104.

Ding, Z. and Christie, P. J. (2003) ‘*Agrobacterium tumefaciens* twin-arginine-dependent translocation is important for virulence, flagellation, and chemotaxis but not type IV secretion’, *Journal of Bacteriology*, 185(3), pp. 760–771. doi:

10.1128/JB.185.3.760-771.2003.

Doherty, A. J., Connolly, B. A. and Worrall, A. F. (1993) ‘Overproduction of the toxic protein, bovine pancreatic DNasel, in *Escherichia coli* using a tightly controlled T7-promoter-based vector’, *Gene*, 136(1–2), pp. 337–340. doi: 10.1016/0378-1119(93)90491-K.

Dow, J. M.; Gabel, F.; Sargent, F.; Palmer, T. (2013) ‘Characterization of a pre-export enzyme-chaperone complex on the twin-Arginine transport pathway’, *Biochemical Journal*, 452(1), pp. 57–66. doi: 10.1042/BJ20121832.

Driessen, A. J. M. and Nouwen, N. (2008) ‘Protein Translocation Across the Bacterial Cytoplasmic Membrane’, *Annual Review of Biochemistry*, 77(1), pp. 643–667. doi: 10.1146/annurev.biochem.77.061606.160747.

Economou, A.; Christie, P. J.; Fernandez, R. C.; Palmer, T. (2006) ‘Secretion by numbers: Protein traffic in prokaryotes’, *Molecular Microbiology*, 62(2), pp. 308–319. doi: 10.1111/j.1365-2958.2006.05377.x.

Eimer, E.; Fröbel, J.; Blümmel, A. S.; Müller, M. (2015) ‘TatE as a regular constituent of bacterial twin-arginine protein translocases’, *Journal of Biological Chemistry*, 290(49), pp. 29281–29289. doi: 10.1074/jbc.M115.696005.

Endermann, R., Hindenach, I. and Henning, U. (1978) ‘Major proteins of the *Escherichia coli* outer cell envelope membrane’, *Molecular Membrane Biology*, 88(1), pp. 71–74.

Fecker, T.; Galaz-Davison, P.; Engelberger, F.; Narui, Y.; Sotomayor, M.; Parra, L. P.; Ramírez-Sarmiento, C. A. (2018) ‘Active Site Flexibility as a Hallmark for Efficient PET Degradation by I. sakaiensis PETase’, *Biophysical Journal*, 114(6), pp. 1302–1312. doi: 10.1016/j.bpj.2018.02.005.

Figge, J.; Wright, C.; Collins, C. J.; Roberts, T. M.; Livingston, D. M. (1988)

‘Stringent regulation of stably integrated chloramphenicol acetyl transferase genes by *E. coli* lac repressor in monkey cells’, *Cell*, 52(5), pp. 713–722.

Flickinger, M. C.; Zhou, W.; Seth, G.; Guardia, M. J.; Hu, W. (2010) ‘Mammalian Cell Bioreactors’, *Encyclopedia of Industrial Biotechnology*, pp. 1–10. doi: 10.1002/9780470054581.eib394.

Frain, K. M.; Jones, A. S.; Schoner, R.; Walker, K. L.; Robinson, C. (2017) ‘The *Bacillus subtilis* TatAdCd system exhibits an extreme level of substrate selectivity’, *Biochimica et Biophysica Acta - Molecular Cell Research*, 1864(1), pp. 202–208. doi: 10.1016/j.bbamcr.2016.10.018.

Frain, K. M., Robinson, C. and van Dijl, J. M. (2019) ‘Transport of Folded Proteins by the Tat System’, *Protein Journal*, 38(4), pp. 377–388. doi: 10.1007/s10930-019-09859-y.

Fuchs, G. and Schlegel, H. G. (2006) *Allgemeine Mikrobiologie*. 8th edition. Thieme, ISBN 9783134446081

Gangl, D.; Zedler, J.; Włodarczyk, A.; Jensen, P. E.; Purton, S.; Robinson, C. (2015) ‘Expression and membrane-targeting of an active plant cytochrome P450 in the chloroplast of the green alga *Chlamydomonas reinhardtii*’, *Phytochemistry*, 110, pp. 22–28. doi: 10.1016/j.phytochem.2014.12.006.

Genest, O.; Ilbert, M.; Méjean, V.; Iobbi-Nivol, C. (2005) ‘TorD, an essential chaperone for TorA molybdoenzyme maturation at high temperature’, *Journal of Biological Chemistry*, 280(16), pp. 15644–15648. doi: 10.1074/jbc.M501119200.

Genevaux, P., Georgopoulos, C. and Kelley, W. L. (2007) ‘The Hsp70 chaperone machines of *Escherichia coli*: A paradigm for the repartition of chaperone functions’, *Molecular Microbiology*, 66(4), pp. 840–857. doi: 10.1111/j.1365-2958.2007.05961.x.

Gérard, F. and Cline, K. (2006) ‘Efficient twin arginine translocation (Tat) pathway transport of a precursor protein covalently anchored to its initial cpTatC binding site’, *Journal of Biological Chemistry*, 281(10), pp. 6130–6135. doi: 10.1074/jbc.M512733200.

Gérard, F. and Cline, K. (2007) ‘The thylakoid proton gradient promotes an advanced stage of signal peptide binding deep within the Tat pathway receptor complex’, *Journal of Biological Chemistry*, 282(8), pp. 5263–5272. doi: 10.1074/jbc.M610337200.

Gerlach, R.; Pop, O. and Müller, J. P. (2004) ‘Tat dependent export of *E. coli* phytase AppA by using the PhoD-specific transport system of *Bacillus subtilis*’, *Journal of Basic Microbiology*, 44(5). doi: 10.1002/jobm.200410423.

Ghosh, D.; Boral, D.; Vankudoth, K. R.; Ramasamy, S. (2019) ‘Analysis of haloarchaeal twin-arginine translocase pathway reveals the diversity of the machineries’, *Heliyon*, 5(5), p. e01587. doi: 10.1016/j.heliyon.2019.e01587.

Giacalone, M. J.; Gentile, A. M.; Lovitt, B. T.; Berkley, N. L.; Gunderson, C. W.; Surber, M. W. (2006) ‘Toxic protein expression in *Escherichia coli* using a rhamnose-based tightly regulated and tunable promoter system’, *BioTechniques*, 40(3), pp. 355–364. doi: 10.2144/000112112.

Glyakina, A. V.; Likhachev, I. V.; Balabaev, N. K.; Galzitskaya, O. V. (2018) ‘Comparative mechanical unfolding studies of spectrin domains R15, R16 and R17’, *Journal of Structural Biology*, 201(2), pp. 162–170. doi: 10.1016/j.jsb.2017.12.003.

Goosen, N. and van de Putte, P. (1995) ‘The regulation of transcription initiation by integration host factor’, *Molecular Microbiology*, 16, pp. 1–7.

Goosens, V. J.; Monteferrante, C. G. and Van Dijl, J. M. (2014) ‘The Tat system of Gram-positive bacteria’, *Biochimica et Biophysica Acta - Molecular Cell Research*,

1843(8), pp. 1698–1706. doi: 10.1016/j.bbamcr.2013.10.008.

Gouffi, K.; Gérard, F.; Santini, C. L.; Wu, L. F. (2004) 'Dual Topology of the *Escherichia coli* TatA Protein', *Journal of Biological Chemistry*, 279(12), pp. 11608–11615. doi: 10.1074/jbc.M313187200.

Green, E. R. and Mecsas, J. (2015) 'Bacterial Secretion Systems – An overview', *American society for Microbiology*, 4(1), pp. 1–32. doi: 10.1128/microbiolspec.VMBF-0012-2015.Bacterial.

Gronenborn, B. (1976) 'Overproduction of phage lambda repressor under control of the *lac* promotor of *Escherichia coli*', *Molecular and Generell Genetics*, 148, pp. 243–250. doi: 10.1007/BF00332898

Groß, C.; Hamacher, K.; Schmitz, K.; Jager, S. (2017) 'Cleavage Product Accumulation Decreases the Activity of Cutinase during PET Hydrolysis', *Journal of Chemical Information and Modeling*, 57(2), pp. 243–255. doi: 10.1021/acs.jcim.6b00556.

Gur, E.; Biran, D. and Ron, E. Z. (2011) 'Regulated proteolysis in Gram-negative bacteria — how and when?', *Nature Reviews Microbiology*, 9(12), pp. 839–848. doi: 10.1038/nrmicro2669.

Gur, E.; Ottofueling, R. and Dougan, D. A. (2013) Machines of destruction – AAA+ proteases and the adaptors that control them. Subcellula, *Subcellular Biochemistry*. Subcellula. Springer, Dordrecht. doi: 10.1007/978-94-007-5940-4_1.

Guzman, L. M.; Belin, D.; Carson, M. J.; Beckwith, J. (1995) 'Tight regulation, modulation, and high-level expression by vectors containing the arabinose p*bad* promoter', *Journal of Bacteriology*, 177, pp. 4121–4130. doi: 10.1128/jb.177.14.4121-4130.1995

Habersetzer, J.; Moore, K.; Cherry, J.; Buchanan, G.; Stansfeld, P. J.; Palmer, T.

(2017) ‘Substrate-triggered position switching of TatA and TatB during Tat transport in *Escherichia coli*’, *Open Biology*, 7(8). doi: 10.1098/rsob.170091.

Hartl, F. U. (2017) ‘Unfolding the chaperone story’, *Molecular Biology of the Cell*, 28(22), pp. 2919–2923. doi: 10.1091/mbc.E17-07-0480.

Hartl, F. U., Bracher, A. and Hayer-Hartl, M. (2011) ‘Molecular chaperones in protein folding and proteostasis’, *Nature*, 475(7356), pp. 324–332. doi: 10.1038/nature10317.

Hatahet, F., Boyd, D. and Beckwith, J. (2014) ‘Disulfide bond formation in prokaryotes: History, diversity and design’, *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1844(8), pp. 1402–1414. doi: 10.1016/j.bbapap.2014.02.014.

Hatzixanthis, K.; Clarke, T. A.; Oubrie, A.; Richardson, D. J.; Turner, R. J.; Sargent, F. (2005) ‘Signal peptide-chaperone interactions on the twin-arginine protein transport pathway’, *Proceedings of the National Academy of Sciences*, 102(24), pp. 8460–8465. doi: 10.1073/pnas.0500737102.

Haugen, S. P.; Ross, W. and Gourse, R. L. (2008) ‘Advances in bacterial promoter recognition and its control by factors that do not bind DNA’, *Nature Reviews Microbiology*, 6(7), pp. 507–519. doi: 10.1038/nrmicro1912.

Hicks, M. G.; De Leeuw, E.; Porcelli, I.; Buchanan, G.; Berks, B. C.; Palmer, T. (2003) ‘The *Escherichia coli* twin-arginine translocase: Conserved residues of TatA and TatB family components involved in protein transport’, *FEBS Letters*, 539(1–3), pp. 61–67. doi: 10.1016/S0014-5793(03)00198-4.

Higgins, E. (2010) ‘Carbohydrate analysis throughout the development of a protein therapeutic’, *Glycoconjugate Journal*, 27(2), pp. 211–225. doi: 10.1007/s10719-009-9261-x.

Hinsley, A. P.; Stanley, T.; Palmer, T.; Berks, B. C. (2001) 'A naturally occurring bacterial Tat signal peptide lacking one of the "invariant" arginine residues of the consensus targeting motif', *FEBS Letters*, 497(1), pp. 45–49. doi: 10.1016/S0014-5793(01)02428-0.

Hohn, B. (2001) 'Transfer of genetic information from *Agrobacterium tumefaciens* to plants', *Encyclopedia of Genetics*, 54(3), pp. 757–764. doi: 10.1006/rwgn.2001.1638.

Holland, I. B. (2004) 'Translocation of bacterial proteins - An overview', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1694(1-3 SPEC.ISS.), pp. 5–16. doi: 10.1016/j.bbamcr.2004.02.007.

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), pp. 7592–7605. doi: 10.1074/jbc.RA118.002205.

Huang, X.; Cao, L.; Qin, Z.; Li, S.; Kong, W.; Liu, Y. (2018) 'Tat-Independent Secretion of Polyethylene Terephthalate Hydrolase PETase in *Bacillus subtilis* 168 Mediated by Its Native Signal Peptide', *Journal of Agricultural and Food Chemistry*, 66(50), pp. 13217–13227. doi: 10.1021/acs.jafc.8b05038.

Imamoglu, R.; Balchin, D.; Hayer-Hartl, M.; Hartl, F. U. (2020) 'Bacterial Hsp70 resolves misfolded states and accelerates productive folding of a multi-domain protein', *Nature Communications*, 11(1). doi: 10.1038/s41467-019-14245-4.

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 I), pp. 3438–3441. doi:

10.1073/pnas.79.11.3438.

Ito, K. and Mori, H. (2009) Bacterial Secreted Proteins: Secretory Mechanisms and Role in Pathogenesis - Chapter 1. Edited by K. Wooldridge. Caister Academic Press.

Ize, B.; Stanley, N. R.; Buchanan, G.; Palmer, T. (2003) 'Role of the *Escherichia coli* Tat pathway in outer membrane integrity', *Molecular microbiology*, 48, pp. 1183–1193.

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), pp. 1801–1804. doi: 10.1128/JB.183.5.1801-1804.2001.

Jack, R. L.; Buchanan, G.; Dubini, A.; Hatzixanthis, K.; Palmer, T.; Sargent, F. (2004) 'Coordinating assembly and export of complex bacterial proteins', *EMBO Journal*, 23(20), pp. 3962–3972. doi: 10.1038/sj.emboj.7600409.

Johnson, I. S. (1983) 'Human Insulin from Recombinant DNA', *Science*, 219(4585), pp. 632–637.

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 - Molecular Cell Research*, 1863(12), pp. 3116–3124. doi: 10.1016/j.bbamcr.2016.09.006.

Joshi, M. V.; Mann, S. G.;Antelmann, H.; Widdick, D. A.; Fyans, J. K.; Chandra, G.; Hutchings, M. I.; Toth, I.; Hecker, M.; Loria, R.; Palmer, T. (2010) 'The twin arginine protein transport pathway exports multiple virulence proteins in the plant pathogen *Streptomyces scabies*', *Molecular Microbiology*, 77(1), pp. 252–271. doi:

10.1111/j.1365-2958.2010.07206.x.

Kadokura, H., Katzen, F. and Beckwith, J. (2003) ‘Protein Disulfide Bond Formation in Prokaryotes’, *Annual Review of Biochemistry*, 72(1), pp. 111–135. doi: 10.1146/annurev.biochem.72.121801.161459.

Kanehisa Laboratories (2010) *Protein export in prokaryotes*. Available at: https://www.genome.jp/kegg-bin/show_pathway?map=ko03060&show_description=show (Accessed: 20 March 2020).

Kawe, M., Horn, U. and Plückthun, A. (2009) ‘Facile promoter deletion in *Escherichia coli* in response to leaky expression of very robust and benign proteins from common expression vectors’, *Microbial Cell Factories*, 8, pp. 1–8. doi: 10.1186/1475-2859-8-8.

Kieselbach, T. and Funk, C. (2003) ‘The family of Deg/HtrA proteases: From *Escherichia coli* to *Arabidopsis*’, *Physiologia Plantarum*, 119(3), pp. 337–346. doi: 10.1034/j.1399-3054.2003.00199.x.

Kikuchi, Y.; Yoda, K.; Yamasaki, M.; Tamura, G. (1981) ‘The nudeotide sequence of the promoter and the ammo-terminal region of alkaline phosphatase structural gene (phoA) of *Escherichia coli*’, *Nucleic Acids Research*, 9(21), pp. 5671–5678. doi: 10.1093/nar/9.21.5671.

Kim, Y. E.; Hipp, M. S.; Bracher, A.; Hayer-Hartl, M.; Ulrich Hartl, F. (2013) Molecular chaperone functions in protein folding and proteostasis, *Annual Review of Biochemistry*. doi: 10.1146/annurev-biochem-060208-092442.

Kolter, R. (1951) ‘The Limitations of LB Medium Moselio (Elio) Schaechter Associate Bloggers Bloggers Emeriti In Memoriam’.

Komarudin, A. G. and Driessen, A. J. M. (2019) ‘SecA-Mediated Protein

Translocation through the SecYEG Channel', *Protein Secretion in Bacteria*, pp. 13–28. doi: 10.1128/microbiolspec.psib-0028-2019.

Kostakioti, M.; Lewman, C. L.; Thanassi, D. G.; Stathopoulos, C. (2005) 'Mechanisms of protein export across the bacterial outer membrane', *Journal of Bacteriology*, 187(13), pp. 4306–4314. doi: 10.1128/JB.187.13.4306-4314.2005.

Kostecki, J. S.; Li, H.; Turner, R. J.; DeLisa, M. P. (2010) 'Visualizing interactions along the *Escherichia coli* twin-arginine translocation pathway using protein fragment complementation', *PLoS ONE,* 5(2), pp. 21–23. doi: 10.1371/journal.pone.0009225.

Lanzer, M. and Bujard, H. (1988) 'Promoters largely determine the efficiency of repressor action', *Proceedings of the National Academy of Science USA*, 85, pp. 8973–8977. doi: 10.1073/pnas.85.23.8973.

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), pp. 15376–15381. doi: 10.1073/pnas.0806338105.

Lee, D. J.; Bingle, L. E .H.; Heurlier, K.; Pallen, M. J.; Penn, C. W.; Busby, S. J. W. Hobman, J. L. (2009) 'Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains.', *BMC microbiology*, 9, p. 252. doi: 10.1186/1471-2180-9-252.

Lee, H. C. and Bernstein, H. D. (2001) 'The targeting pathway of *Escherichia coli* presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal', *Proceedings of the National Academy of Sciences of the United States of America*, 98(6), pp. 3471–3476. doi: 10.1073/pnas.051484198.

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), pp. 5871–5879. doi: 10.1128/JB.184.21.5871-5879.2002.

Lee, S. Y. (1996) 'High cell-density culture of *Escherichia coli*', *Trends in Biotechnology*, 14(3), pp. 98–105. doi: 10.1016/0167-7799(96)80930-9.

Lee, Y. J.; Lee, R.; Lee, S. H.; Yim, S. S.; Jeong, K. J. (2016) 'Enhanced secretion of recombinant proteins via signal recognition particle (SRP)-dependent secretion pathway by deletion of rrsE in *Escherichia coli*', *Biotechnology and Bioengineering*, 113(11), pp. 2453–2461. doi: 10.1002/bit.25997.

Lens, J. and Evertzen, A. (1952) 'The difference between insulin from cattle and from pigs', *Biochimica et Biophysica Acta*, 83(s399), pp. 29–31. doi: 10.1016/0006-3002(52)90048-6.

Lim, H. K.; Mansell, T. J.; Linderman, S. W.; Fisher, A. C.; Dyson, M. R.; DeLisa, M. P. (2009) 'Mining mammalian genomes for folding competent proteins using Tat-dependent genetic selection in *Escherichia coli*', *Protein Science*, 18(12), pp. 2537–2549. doi: 10.1002/pro.262.

Lin, J. T. and Stewart, V. (1997) 'Nitrate assimilation by bacteria', *Advances in Microbial Physiology*, 39, pp. 1–30. doi: 10.1016/s0065-2911(08)60014-4.

Lingg, N. *et al.* (2012) 'The sweet tooth of biopharmaceuticals: Importance of recombinant protein glycosylation analysis', *Biotechnology Journal*, 7(12), pp. 1462–1472. doi: 10.1002/biot.201200078.

Liu, B.; He, L.; Wang, L.; Li, T.; Li, C.; Liu, H.; Luo, Y. (2018) 'Protein crystallography and site-direct mutagenesis analysis of the poly(Ethylene terephthalate) hydrolase petase from *Ideonella sakaiensis*', *ChemBioChem*, 19(14),

pp. 1471–1475. doi: 10.1002/cbic.201800097.

Liu, X.; Gupta, S. T. P.; Bhimsaria, D.; Reed, J. L.; Rodríguez-Martínez, J. A.; Ansari, A. Z.; Raman, S. (2019) 'De novo design of programmable inducible promoters', *Nucleic acids research*, 47(19), pp. 10452–10463. doi: 10.1093/nar/gkz772.

Luirink, J. and Sinning, I. (2004) 'SRP-mediated protein targeting: Structure and function revisited', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1694(1-3 SPEC.ISS.), pp. 17–35. doi: 10.1016/j.bbamcr.2004.03.013.

Luli, G. W. and Strohl, W. R. (1990) 'Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations', *Applied and Environmental Microbiology*, 56(4), pp. 1004–1011. doi: 10.1128/aem.56.4.1004-1011.1990.

Makrides, S. C. (1996) 'Strategies for achieving high-level expression of genes in *Escherichia coli*', *Microbiological Reviews*, 60(3), pp. 512–538.

Mallick, P.; Boutz, D. R.; Eisenberg, D.; Yeates, T. O.(2002) 'Genomic evidence that the intracellular proteins of archaeal microbes contain disulfide bonds', *Proceedings of the National Academy of Sciences of the United States of America*, 99(15), pp. 9679–9684. doi: 10.1073/pnas.142310499.

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), pp. 2533–2542. doi: 10.1002/bit.24535.

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), pp. 281–290. doi: 10.1002/btpr.1858.

McGovern, P. E.; Zhang, J.; Tang, J.; Zhang, Z.; Hall, G. R.; Moreau, R. A.; Nuñez, A.; Butrym, E. D.; Richards, M. P.; Wang, C. S.; Cheng, G.; Zhao, Z.; Wang, C. (2004) 'Fermented beverages of pre- and proto-historic China', *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), pp. 17593–17598. doi: 10.1073/pnas.0407921102.

Mendel, S. and Robinson, C. (2007) Targeting of Proteins by the Twin-Arginine Translocation System in Bacteria and Chloroplasts, *Enzymes*. Elsevier Masson SAS. doi: 10.1016/S1874-6047(07)25003-6.

Merrimack Pharmaceuticals, I. (2016) 'FDA Grants Merrimack Fast Track Designation for Seribantumab (MM-121) in Non-small Cell Lung Cancer (NASDAQ:MACK)', *Merrimack*. Available at: http://investors.merrimack.com/releasedetail.cfm?releaseid=978326.

Merz, F.; Boehringer, D.; Schaffitzel, C.; Preissler, S.; Hoffmann, A.; Maier, T.; Rutkowska, A.; Lozza, J.; Ban, N.; Bukau, B.; Deuerling, E. (2008) 'Molecular mechanism and structure of Trigger Factor bound to the translating ribosome', *EMBO Journal*, 27(11), pp. 1622–1632. doi: 10.1038/emboj.2008.89.

Miot, M. and Betton, J. M. (2004) 'Protein quality control in the bacterial periplasm', *Microbial Cell Factories*, 3, pp. 1–13. doi: 10.1186/1475-2859-3-4.

Miroux, B. and Walker, J. E. (1996) 'Over-production of Proteins in *Escherichia coli*: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels', *Journal of Molecular Biology*, 260(3), pp. 289–298.

Mogensen, J. E. and Otzen, D. E. (2005) 'Interactions between folding factors and

bacterial outer membrane proteins', *Molecular Microbiology*, 57(2), pp. 326–346. doi: 10.1111/j.1365-2958.2005.04674.x.

Mohamed N. Baeshen (2014) 'Production of biopharmaceuticals in *E. coli*: current scenario and future perspectives', *Journal of microbiolgy and biotechnology*, 25(7), pp. 1–24. doi: 10.4014/jmb.1405.05052.

Money, N. P. (2016) 'Fungi and Biotechnology', in *The Fungi: Third Edition*. Elsevier Ltd., pp. 401–424. doi: 10.1016/B978-0-12-382034-1.00012-8.

Morán Luengo, T., Mayer, M. P. and Rüdiger, S. G. D. (2019) 'The Hsp70–Hsp90 Chaperone Cascade in Protein Folding', *Trends in Cell Biology*, 29(2), pp. 164–177. doi: 10.1016/j.tcb.2018.10.004.

Mori, H. and Cline, K. (1998) 'A signal peptide that directs non-Sec transport in bacteria also directs efficient and exclusive transport on the thylakoid Delta pH pathway', *Journal of Biological Chemistry*, 273(19), pp. 11405–11408. doi: 10.1074/jbc.273.19.11405.

Mould, R. M. and Robinson, C. (1991) 'A proton gradient is required for the transport of two lumenal oxygen-evolving proteins across the thylakoid membrane', *Journal of Biological Chemistry*, 266(19), pp. 12189–12193. doi: 10.1016/S0021-9258(18)98879-4

Movva, N. R., Nakamura, K. and Inouye, M. (1980) 'Amino acid sequence of the signal peptide of ompA protein, a major outer membrane protein of *Escherichia coli*', *Journal of Biological Chemistry*, 255(1), pp. 27–29. doi: 10.1016/S0021-9258(19)86257-9.

Mujacic, M. and Cooper, K. W. (1999) 'Cold-inducible cloning vectors for low-temperature protein expression in *Escherichia coli*: application to the production of a toxic and proteolytically sensitive fusion protein', 238, pp. 325–332. doi:

10.1016/S0378-1119(99)00328-5.

Müller, M; Koch, H. G.; Beck, K.; Schafer, U.. (2001) ‘Protein traffic in bacteria: multiple routes from the ribosome to and across the membrane’, *Nucleic Acids Research*, 66, pp. 107–157. doi: 10.1016/S0079-6603(00)66028-2.

Müller, M. and Klösgen, R. B. (2005) ‘The Tat pathway in bacteria and chloroplasts’, *Molecular Membrane Biology*, 22(1–2), pp. 113–121. doi: 10.1080/09687860500041809.

Nakamoto, H. and Bardwell, J. C. A. (2004) ‘Catalysis of disulfide bond formation and isomerization in the *Escherichia coli* periplasm’, *Biochimica et Biophysica Acta - Molecular Cell Research*, 1694(1-3 SPEC.ISS.), pp. 111–119. doi: 10.1016/j.bbamcr.2004.02.012.

Natale, P., Brüser, T. and Driessen, A. J. M. (2008) ‘Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane-Distinct translocases and mechanisms’, *Biochimica et Biophysica Acta - Biomembranes*, 1778(9), pp. 1735–1756. doi: 10.1016/j.bbamem.2007.07.015.

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), pp. 241–250. doi: 10.1016/0168-1656(96)01514-3.

Nouwen, N.; Piwowarek, M.; Berrelkamp, G.; Driessen, A. J. M. (2005) ‘The large first periplasmic loop of SecD and SecF plays an important role in SecDF functioning’, *Journal of Bacteriology*, 187(16), pp. 5857–5860. doi: 10.1128/JB.187.16.5857-5860.2005.

Nyquist, K. and Martin, A. (2014) ‘Marching to the beat of the ring: Polypeptide translocation by AAA+ proteases’, *Trends in Biochemical Sciences*, 39(2), pp. 53–60. doi: 10.1016/j.tibs.2013.11.003.

Oates, J.; Mathers, J.; Mangels, D.; Kühlbrandt, W.; Robinson, C.; Model, K. (2003) 'Consensus structural features of purified bacterial TatABC complexes', *Journal of Molecular Biology*, 330(2), pp. 277–286. doi: 10.1016/S0022-2836(03)00621-1.

Oliver, D. B. (1993) 'SecA protein: autoregulated ATPase catalysing preprotein insertion and translocation across the *Escherichia coli* inner membrane', *Molecular Microbiology*, 7(2), pp. 159–165. doi: 10.1111/j.1365-2958.1993.tb01107.x.

Öztürk, S., Ergün, B. G. and Çalık, P. (2017) 'Double promoter expression systems for recombinant protein production by industrial microorganisms', *Applied Microbiology and Biotechnology*, 101(20), pp. 7459–7475. doi: 10.1007/s00253-017-8487-y.

Paetzel, M.; Karla, A.; Strynadka, N.; Dalbey, R. (2002) 'Signal peptidases', *Chemical reviews*, 102(12), pp. 4549–4580. doi: 10.1021/cr010166y.

Palmer, T. and Berks, B. C. (2012) 'The twin-arginine translocation (Tat) protein export pathway', *Nature Reviews Microbiology*, 10(7), pp. 483–496. doi: 10.1038/nrmicro2814.

Palmer, T., Berks, B. C. and Sargent, F. (2010) 'Analysis of Tat Targeting Function and Twin-Arginine Signal Peptide Activity in *Escherichia coli*', in Economou, A. (ed.) *Protein Secretion: Methods and Protocols*. Totowa, NJ: Humana Press, pp. 191–216. doi: 10.1007/978-1-60327-412-8_12.

Palmer, T., Sargent, F. and Berks, B. C. (2005) 'Export of complex cofactor-containing proteins by the bacterial Tat pathway', *Trends in Microbiology*, 13(4), pp. 175–180. doi: 10.1016/j.tim.2005.02.002.

Palmer, T., Sargent, F. and Berks, B. C. (2010) 'The Tat protein export pathway', *American Society for Microbiology*.

Panahandeh, S.; Maurer, C.; Moser, M; DeLisa, M. P.; Müller, M. (2008) 'Following

the path of a twin-arginine precursor along the TatABC translocase of *Escherichia coli*', *Journal of Biological Chemistry*, 283(48), pp. 33267–33275. doi: 10.1074/jbc.M804225200.

Papadopoulou, A., Hecht, K. and Buller, R. (2019) 'Enzymatic PET degradation', *Chimia*, 73(9), pp. 743–749. doi: 10.2533/chimia.2019.743.

Papanikou, E., Karamanou, S. and Economou, A. (2007) 'Bacterial protein secretion through the translocase nanomachine', *Nature Reviews Microbiology*, 5(11), pp. 839–851. doi: 10.1038/nrmicro1771.

Patel, R., Smith, S. M. and Robinson, C. (2014) 'Protein transport by the bacterial Tat pathway', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1843(8), pp. 1620–1628. doi: 10.1016/j.bbamcr.2014.02.013.

PET Resin Association (2015) *About PET*. Available at: http://www.petresin.org/aboutpet.asp (Accessed: 3 April 2020).

Phan, T. T. P., Nguyen, H. D. and Schumann, W. (2012) 'Development of a strong intracellular expression system for *Bacillus subtilis* by optimizing promoter elements', *Journal of Biotechnology*, 157(1), pp. 167–172. doi: 10.1016/j.jbiotec.2011.10.006.

Plastics Insight (2019) *Polyethylene Terephthalate (PET): Production, Price, Market and its Properties*. Available at: https://www.plasticsinsight.com/resin-intelligence/resin-prices/polyethylene-terephthalate/ (Accessed: 3 April 2020).

Du Plessis, D. J. F., Nouwen, N. and Driessen, A. J. M. (2011) 'The Sec translocase', *Biochimica et Biophysica Acta - Biomembranes*, 1808(3), pp. 851–865. doi: 10.1016/j.bbamem.2010.08.016.

Pogliano, J. A. and Beckwith, J. (1994) 'SecD and SecF facilitate protein export in *Escherichia coli*.', *The EMBO Journal*, 13(3), pp. 554–561. doi: 10.1002/j.1460-

2075.1994.tb06293.x.

Pop, O.; Martin, U.; Abel, C.; Müller, J. P. (2002) 'The twin-arginine signal peptide of PhoD and the TatAd/Cd proteins of *Bacillus subtilis* form an autonomous tat translocation system', *Journal of Biological Chemistry*, 277(5), pp. 3268–3273. doi: 10.1074/jbc.M110829200.

Powers, E. T., Powers, D. L. and Gierasch, L. M. (2012) 'FoldEco: A Model for Proteostasis in *E. coli*', *Cell Reports*, 1(3), pp. 265–276. doi: 10.1016/j.celrep.2012.02.011.

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

Pugsley, A. P. (1993) 'The complete general secretory pathway in gram-negative bacteria.', *Microbiological Reviews*, 57(1), pp. 50 LP – 108. Available at: http://mmbr.asm.org/content/57/1/50.abstract.

Qiu, J., Swartz, J. R. and Georgiou, G. (1998) 'Expression of active human tissue-type plasminogen activator in *Escherichia coli*', *Applied and Environmental Microbiology*, 64(12), pp. 4891–4896. doi: 10.1128/aem.64.12.4891-4896.1998.

Quax, W. J. (1997) 'Merits of Secretion of Heterologous Proteins from Industrial Microorganisms', *Folia Microbiologica*, 42(2), pp. 99–103. doi: 10.1007/bf02898715.

Ranade, V. V (2010) Biotechnology: Pharmaceutical Aspects, *American Journal of Therapeutics*. doi: 10.1097/mjt.0b013e31817c947a.

Randall, L. and Hardy, 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, pp. 921–928.

Randall, L. and Hardy, S. (2002) 'SecB, one small chaperone in the complex milieu of the cell.', *Cellular and Molecular Life Sciences*, 59(10), pp. 1617–1623.

Richardson, D. J. and Watmough, N. J. (1999) 'Inorganic nitrogen metabolism in bacteria', *Current Opinion in Chemical Biology*, 3(2), pp. 207–219. doi: 10.1016/S1367-5931(99)80034-9.

Richter, S.; Lindenstrauss, U.; Lücke, C.; Bayliss, R,; Brüser, T. (2007) 'Functional tat transport of unstructured, small, hydrophilic proteins', *Journal of Biological Chemistry*, 282(46), pp. 33257–33264. doi: 10.1074/jbc.M703303200.

Richter, S. and Brüser, T. (2005) 'Targeting of unfolded PhoA to the TAT translocon of *Escherichia coli*', *Journal of Biological Chemistry*, 280(52), pp. 42723–42730. doi: 10.1074/jbc.M509570200.

Robinson, C.; Matos, C. F. R. O.; Beck, D.; Ren, C.; Lawrence, J.; Vasisht, N.; Mendel, S. (2011) 'Transport and proofreading of proteins by the twin-arginine translocation (Tat) system in bacteria', *Biochimica et Biophysica Acta - Biomembranes*, 1808(3), pp. 876–884. doi: 10.1016/j.bbamem.2010.11.023.

Robinson, C. and Bolhuis, A. (2004) 'Tat-dependent protein targeting in prokaryotes and chloroplasts', *Biochimica et Biophysica Acta - Molecular Cell Research*, 1694(1-3 SPEC.ISS.), pp. 135–147. doi: 10.1016/j.bbamcr.2004.03.010.

Rollauer, S. E.; Tarry, M. J.; Graham, J. E.; Jääskeläinen, M.; Jäger, F.; Johnson, S.; Krehenbrink, M.; Liu, S. M.; Lukey, M. J.; Marcoux, J.; McDowell, M. A.; Rodriguez, F.; Roversi, P; Stansfeld, P. J.; Robinson, C. V.; Sansom, M. S.P.; Palmer, T.; Högbom, M.; Berks, B. C.; Lea, S. M. (2012) 'Structure of the TatC core of the twin-arginine protein transport system', *Nature*, 492(7428), pp. 210–214. doi: 10.1038/nature11683.

Ronkvist, Å. M.; Xie, W.; Lu, W.; Gross, R. A. (2009) 'Cutinase-Catalyzed Hydrolysis of Poly(ethylene terephthalate)', *Macromolecules*, 42(14), pp. 5128–5138. doi: 10.1021/ma9005318.

Rosano, G. L. and Ceccarelli, E. A. (2014) 'Recombinant protein expression in *Escherichia coli*: Advances and challenges', *Frontiers in Microbiology*, 5(APR). doi: 10.3389/fmicb.2014.00172.

Rose, P.; Fröbel, J.; Graumann, P. L.; Müller, M. (2002) 'Adaptation of protein secretion to extremely high-salt conditions by extensive use of the twin-arginine translocation pathway', *Molecular microbiology*, 45, pp. 943–950.

Rossiter, A. E.; Browning, D. F.; Leyton, D. L.; Johnson, M. D.; Godfrey, R. E.; Wardius, C. A.; Desvaux, M.; Cunningham, A. F.; Ruiz-Perez, F.; Nataro, J. P. Busby, S. J.W.; Henderson, I. R. (2011) 'Transcription of the plasmid-encoded toxin gene from Enteroaggregative *Escherichia coli* is regulated by a novel co-activation mechanism involving CRP and Fis', *Molecular Microbiology*, 81(1), pp. 179–191. doi: 10.1111/j.1365-2958.2011.07685.x.

Rossiter, A. E.; Godfrey, R. E.; Connolly, J. A.; Busby, S. J.W.; Henderson, I. R. Browning, D. F. (2015) 'Expression of different bacterial cytotoxins is controlled by two global transcription factors, CRP and Fis, that co-operate in a shared-recruitment mechanism', *Biochemical Journal*, 466, pp. 323–335. doi: 10.1042/BJ20141315.

Rusch, S. L. and Kendall, D. A. (2007) 'Interactions That Drive Sec-Dependent Bacterial Protein Transport', *Biochemistry*, 46(34), pp. 9665–9673.

Sabra, W.; Quitmann, H.; Zeng, A. P.; Dai, J. Y.; Xiu, Z. L. (2011) Microbial Production of 2,3-Butanediol. Second Edi, *Comprehensive Biotechnology, Second Edition*. Second Edi. Elsevier B.V. doi: 10.1016/B978-0-08-088504-9.00161-6.

Sachelaru, I.; Petriman, N. A.; Kudva, R; Kuhn, P.; Welte, T.; Knapp, B.; Drepper,

F.; Warscheid, B.; Koch, H. G. (2013) ‘YidC occupies the lateral gate of the SecYEG translocon and is sequentially displaced by a nascent membrane protein’, *Journal of Biological Chemistry*, 288(23), pp. 16295–16307. doi: 10.1074/jbc.M112.446583.

Sachelaru, I.; Winter, L.; Knyazev, D. G.; Zimmermann, M.; Vogt, A.; Kuttner, R.; Ollinger, N.; Siligan, C.; Pohl, P.; Koch, H. G. (2017) ‘YidC and SecYEG form a heterotetrameric protein translocation channel’, *Scientific Reports*, 7(1), pp. 1–15. doi: 10.1038/s41598-017-00109-8.

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), pp. 1–7. doi: 10.1186/s12934-016-0437-3.

Sanger, F. (1958) ‘The chemistry of insulin’, *Nobel Lectures*. Available at: https://www.nobelprize.org/prizes/chemistry/1958/sanger/lecture/.

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*’, *EMBO Journal*, 17(1), pp. 101–112. doi: 10.1093/emboj/17.1.101.

Sargent, F.; Bogsch, E. G.; Stanley, N. R.; Wexler, M.; Robinson, C.; Berks, B. C.; Palmer, T. (1998) ‘Overlapping functions of components of a bacterial Sec-independent protein export pathway’, *EMBO Journal*, 17(13), pp. 3640–3650. doi: 10.1093/emboj/17.13.3640.

Sargent, F. (2007) ‘The twin-arginine transport system : moving folded proteins across membranes’, *Biochemical Society Transactions*, 35(5), pp. 835–847. doi: 10.1042/BST0350835.

Sauer, R. T. and Baker, T. A. (2011) ‘AAA+ Proteases: ATP-fueled machines of protein destruction’, *Annual Review of Biochemistry*, 80, pp. 587–612. doi:

10.1146/annurev-biochem-060408-172623.

Schuhmann, H. and Adamska, I. (2012) ‘Deg proteases and their role in protein quality control and processing in different subcellular compartments of the plant cell’, *Physiologia Plantarum*, 145(1), pp. 224–234. doi: 10.1111/j.1399-3054.2011.01533.x.

Seo, H.; Kim, S.; Son, H. F.; Sagong, H. Y.; Joo, S.; Kim, K. J. (2019) ‘Production of extracellular PETase from *Ideonella sakaiensis* using sec-dependent signal peptides in *E. coli*’, *Biochemical and Biophysical Research Communications*, 508(1), pp. 250–255. doi: 10.1016/j.bbrc.2018.11.087.

Sjöström, M.; Wold, S.; Wieslander, A.; Rilfors, L. (1987) ‘Signal peptide amino acid sequences in *Escherichia coli* contain information related to final protein localization. A multivariate data analysis.’, *The EMBO Journal*, 6(3), pp. 823–831. doi: 10.1002/j.1460-2075.1987.tb04825.x.

Srivastava, A. K. and Gupta, S. (2011) Fed-Batch Fermentation - Design Strategies. Second Edi, *Comprehensive Biotechnology, Second Edition*. Second Edi. Elsevier B.V. doi: 10.1016/B978-0-08-088504-9.00112-4.

Stanley, N. R.; Findlay, K. and Berks, B. C. (2001) ‘*Escherichia coli* strains blocked in Tat-dependent protein export exhibit pleiotropic defects in the cell envelope’, 183(1), pp. 139–144. doi: 10.1128/JB.183.1.139.

Strauch, K. L.; Johnson, K. and Beckwith, J. (1989) ‘Caracterization of degP, a gene required for proteolysis in the cell envelope and essential for growth of *E. coli* at high temperature ’, 171(5), pp. 2689–2696. doi: 10.1128/jb.171.5.2689-2696.1989.

Studier, F. W. (1991) ‘Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system’, *Journal of Molecular Biology*, 219(1), pp. 37–44. doi: 10.1016/0022-2836(91)90855-Z.

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', *Journal of Biological Chemistry*, 293(18), pp. 6672–6681. doi: 10.1074/jbc.RA117.000880.

Swartz, J. R. (2001) 'Advances in *Escherichia coli* production of therapeutic proteins', *Current Opinion in Biotechnology*, 12(2), pp. 195–201. doi: 10.1016/S0958-1669(00)00199-3.

Tang, Y. C.; Chang, H. C.; Roeben, A.; Wischnewski, D.; Wischnewski, N.; Kerner, M. J.; Hartl, F. U.; Hayer-Hartl, M. (2006) 'Structural Features of the GroEL-GroES Nano-Cage Required for Rapid Folding of Encapsulated Protein', *Cell*, 125(5), pp. 903–914. doi: 10.1016/j.cell.2006.04.027.

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), pp. 13284–13289. doi: 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), pp. 211–222. doi: 10.1007/s00253-006-0465-8.

Thermo Fisher Scientific (2020) *IPTG Induction*. Available at: https://www.thermofisher.com/uk/en/home/life-science/cell-culture/microbiological-culture/iptg-induction.html (Accessed: 24 October 2019).

Thomason, L. C.; Sawitzke, J. A.; Li, X.; Costantino, N.; Court, D. L. (2014) Recombineering: Genetic engineering in bacteria using homologous recombination, *Current Protocols in Molecular Biology*. doi: 10.1002/0471142727.mb0116s106.

Trivedi, H. K. and Ju, L. -K (1994) 'Study of Nitrate Metabolism of *Escherichia coli* Using Fluorescence', *Biotechnology Progress*, 10(4), pp. 421–427. doi: 10.1021/bp00028a012.

Tseng, T.-T.; Gratwick, K. S.; Kollman, J.; Park, D.; Nies, D. H.; Goffeau, A.; Saier Jr., M. H. (1999) 'The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins', *Journal of Molecular Microbiology Biotechnology*, 1, pp. 107–125. PMID: 10941792

Tsukazaki, T. (2018) 'Structure-based working model of SecDF, a proton-driven bacterial protein translocation factor', *FEMS Microbiology Letters*, 365(12), pp. 1–9. doi: 10.1093/femsle/fny112.

Tsukazaki, T. and Nureki, O. (2011) 'The mechanism of protein export enhancement by the SecDF membrane component', *Biophysics*, 7, pp. 129–133. doi: 10.2142/biophysics.7.129.

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), pp. 8309–8316. doi: 10.1074/jbc.M610507200.

Turner, R. J., Papish, A. L. and Sargent, F. (2004) 'Sequence analysis of bacterial redox enzyme maturation proteins (REMPs)', *Canadian Journal of Microbiology*, 50(4), pp. 225–238. doi: 10.1139/w03-117.

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(October), pp. 1–8. doi: 10.3389/fendo.2018.00613.

Verma, A.; Agrahari, S.; Rastogi, S.; Singh, A. (2011) 'Biotechnology in the realm of history', *Journal of Pharmacy and BioAllied Science*, 3(3), pp. 321–323.

Waegeman, H. and Soetaert, W. (2011) 'Increasing recombinant protein production in *Escherichia coli* through metabolic and genetic engineering', *Journal of Industrial Microbiology and Biotechnology*, 38(12), pp. 1891–1910. doi: 10.1007/s10295-011-1034-4. doi: 10.1007/s10295-011-1034-4.

Walker, M. S. and DeMoss, J. A. (1992) 'Role of alternative promoter elements in transcription from the nar promoter of *Escherichia coli*', *Journal of Bacteriology*, 174(4), pp. 1119–1123. doi: 10.1128/jb.174.4.1119-1123.1992.

Wang, J. H.; Chen, T. and Tsai, C. (2012) 'In Search of an Innovative State: The Development of the Biopharmaceutical Industry in Taiwan, South Korea and China', *Development and Change*, 43(2). Available at: https://doi.org/10.1111/j.1467-7660.2012.01769.x.

Wang, N.; Guan, F.; Lv, X.; Han, D.; Zhang, Y.; Wu, N.; Xia, X.; Tian, J. (2020) 'Enhancing secretion of polyethylene terephthalate hydrolase PETase in *Bacillus subtilis* WB600 mediated by the SPamy signal peptide', *Letters in Applied Microbiology*, 71(3), pp. 235–241. doi: 10.1111/lam.13312.

Weickert, M. J.; Doherty, D. H.; Best, E. A.; Olins, P. O. (1996) 'Optimization of heterologous protein production in *Escherichia coli*', *Current Opinion in Biotechnology*, 7(5), pp. 494–499. doi: 10.1016/S0958-1669(96)80051-6.

Werner, R. and Noé, W. (1993) 'Mammalian cell cultures. Part II: Genetic engineering, protein glycosylation, fermentation and process control', *Arzneim-Forch/Drug Res*, 43, pp. 1242–1249. PMID: 8292072

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), pp. 16717–16722. doi: 10.1074/jbc.M000800200.

Wood, D. W.; Setubal, J. C.; Kaul, R.; Monks, D. E.; Kitajima, J. P.; Okura, V. K.; Zhou, Y.; Chen, L.; Wood, G. E.; Almeida, J.; Woo, L.; Chen, Y.; Paulsen, I. T.; Eisen, J. A.; Karp, P. D.; Bovee D., S.; Chapman, P.; Clendenning, J.; Deatherage, ; . Gillet, W.; Grant, C.; Kutyavin, T.; Levy, R.; Li, M. J.; McClelland, E.; Palmieri, A.; Raymond, C.; Rouse, G.; Saenphimmachak, C.; Wu, Z.; Romero, P.; Gordon, D.; Zhang, S.; Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, B.; Liao, L.; Kim, S.; Hendrick, C.; Zhao, Z. Y.; Dolan, M.; Chumley, F.; Tingey, S. V.; Tomb, J. F.; Gordon, M. P.; Olson, M. V.; Nester, E. W. (2001) ‘The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58’, *Science*, 294(5550), pp. 2317–2323. doi: 10.1126/science.1066804.

Yamanè, T. and Shimizu, S. (2005) ‘Fed-batch techniques in microbial processes’, *Bioprocess Parameter Control*, pp. 147–194. doi: 10.1007/bfb0006382.

Yang, S; Xu, H.; Wang, J.; Liu, C.; Lu, H.; Liu, M.; Zhao, Y.; Tian, B.; Wang, L.; Hua, Y. (2016) ‘Cyclic AMP Receptor Protein Acts as a Transcription Regulator in Response to Stresses in *Deinococcus radiodurans*’, *PLoS ONE*, 11(5). doi: 10.1371/journal.pone.0155010

Yee, L. and Blanch, H. W. (1992) ‘Recombinant Protein Expression in High cell density fed-bantch culrutes of *Escherichia coli*. Nature Biotechnology 10, pp. 1550 – 1556. doi: 10.1038/nbt1292-1550

Yoshida, S.; Hiraga, K.; Takehana, T.; Taniguchi, I.; Yamaji, H.; Maeda, Y.; Toyohara, K.; Miyamoto, K.; Kimura, Y.; Oda, K. (2016) ‘A bacterium that

degrades and assimilates poly(ethylene terephthalate)', *Science*, 351(6278), pp. 1196–1199. doi: 10.1126/science.aaf8305.

Younus, H. (2018) 'Therapeutic potentials of superoxide dismutase.', *International journal of health sciences*, 12(3), pp. 88–93. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29896077%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5969776.

8. Appendix

8.1. Alignment of TatABC *E. coli* to TatABC *A. tumefaciens*

TatA alignment

```
43.3% identity in 67 residues overlap; Score: 122.0; Gap frequency: 6.0%

Eco             1 MGGISIWQLLIIAVIVVLLFGTKKLGSIGSDLGASIKGFKKAMSDDEPKQDKTSQDADFT
Atu             1 MGSFSVWHWLIVLVIVLVLFGRGKIPELMGDVAKGIKSFKKGMADE----DQTPPPADVN
                  **  * *  **  ***  ***  *      *    ** *** * *     * *   **

Eco            61 AKTIADK
Atu            57 AKTVDHK
                  ***   *
```

TatB alignment

```
31.2% identity in 93 residues overlap; Score: 135.0; Gap frequency: 0.0%

Eco             1 MFDIGFSELLLVFIIGLVVLGPQRLPVAVKTVAGWIRALRSLATTVQNELTQELKLQEFQ
Atu             1 MFDIGWSELLVIAVVLIVVVGPKDLPPMIRAFGKTMAGLRKMAGDFRTQFDEALKEADMD
                  ***** ****       ** **  **           **  *          **

Eco            61 DSLKKVEKASLTNLTPELKASMDELRQAAESMK
Atu            61 DVRQTISDVRNLNPTNSLRDAMNPLRQLGNEIK
                  *          * *  *   *  ***     *
```

TatC alignment

```
35.2% identity in 256 residues overlap; Score: 349.0; Gap frequency: 4.7%

Eco             7 QPLITHLIELRKRLLNCIIAVIVIFLCLVYFANDIYHLVSAPLIKQLPQG------STMI
Atu            10 QPLIEHLMELRTRLIWSLGAFFVAFIACFAVAKHLFNLLVIPYKWAVIWAGLDVAKSSLI
                  **** ** *** **     *  * *      *     *   *              *  *

Eco            61 ATDVASPFFTPIKLTFMVSLILSAPVILYQVWAFIAPALYKHERRLVVPLLVSSSLLFYI
Atu            70 YTAPQEFFFTQIKVAMFGAMVISFPVIASQLYKFVAPGLYKNERAAFLPFLIASPILFLI
                   *     *** **         * ***  *   * ** *** **    * *  *  ** *

Eco           121 GMAFAYFVVFPLAFGF------LANTAPEGVQVSTDIASYLSFVMALFMAFGVSFEVPVA
Atu           130 GAALVYFFFTPMVMWFFLAMQQLPVDGEVAISLMPKVSEYLSLIMTLVLSFGLVFQLPVV
                  * *  **   *    *      *                 ***  * *   **  *  **

Eco           175 IVLLCWMGITSPEDLRKKRPYVLVGAFVVGMLLTPPDVFSQTLLAIPMYCLFEIGVFFSR
Atu           190 TTLLARVGILTSDWLREKRKFAIVMAFVVAAVLTPPDPMSQIGLALPAIILYEISIYMAR
                    **   **     ** **    * ****   *****  **  ** *   * **     *

Eco           235 FYVGKGRNREEENDAE
Atu           250 LVERKRAAESKSTELE
                      *          *
```

Figure 61: Here shown is the alignment of TatABC *E. coli* (eco) to TatABC *A. tumefaciens* (atu). For alignment web.expasy.org has been used.

8.2. Base pair and amino acid sequence of codon optimised TatABC *A. tumefaciens* operon used for pDoc-K

Base pair sequence:

ATGATGGGTTCTTTCTCTGTTTGGCACTGGCTGATCGTTCTGGTTATCGTTCTGGTTCTGTT
CGGTCGTGGTAAAATCCCGGAACTGATGGGTGACGTTGCTAAAGGTATCAAATCTTTCAAAA
AAGGTATGGCTGACGAAGACCAGACCCCGCCGCCGGCTGACGCTAACACCAAAACCGTTGAC
CACAAAGCTGACGAAATCAAATAAATGTTCGACATCGGTTGGTCTGAACTGCTGGTTATCGC
TGTTGTTCTGATCGTTGTTGTTGGTCCGAAAGACCTGCCGCCGATGATCCGTGCTTTCGGTA
AAACCATGGCTGGTCTGCGTAAAATGGCTGGTGACTTCCGTACCCAGTTCGACGAAGCTCTG
AAAGAAGCTGACATGGACGACGTTCGTCAGACCATCTCTGACGTTCGTAACCTGAACCCGAC
CAACTCTCTGCGTGACGCTATGAACCCGCTGCGTCAGCTGGGTAACGAAATCAAATCTGACC
TGCAGAAAGCTACCACCCCGCCGGAAGGTCTGTCTTCTTCTCCGGCTACCTCTACCGCTGCT
CCGGCTACCTCTGAACCGGTTGCTCCGCTGGTTTCTATCCCGGAACCGGAAATGAAACTGCC
GGACACCCCGCCGGTTACCGCTTCTGCTCCGGTTGCTGCTCCGGTTGTTCCGGCTGAAAAAC
CGAAACGTGCTCGTGCTAAATCTATCGCTACCGTTGAAGCTGAAGCTGTTGCTGCTAAACCG
AAACGTGCTTCTCGTGCTAAAGCTGTTGCTGCTTCTGAACCGGTTGCTGTTAACTCTACCGA
AACCTCTGCTGCTAAACCGACCGTTAAAGCTGTTGCTCGTAAAGCTGCTGCTCGTAAAGTTG
CTGCTGAAAAACCGGTTGTTGCTGCTGACGCTAAACCGGTTAAACCGGCTCGTACCAAAGCT
GCTAAACCGAAAAAAGACGAGGCTTAAATGTCTGGTGACATCGAAGACAAACCGCAGCCGCT
GATCGAACACCTGATGGAACTGCGTACCCGTCTGATCTGGTCTCTGGGTGCTTTCTTCGTTG
CTTTCATCGCTTGCTTCGCTGTTGCTAAACACCTGTTCAACCTGCTGGTTATCCCGTACAAA
TGGGCTGTTCTGTGGGCTGGTCTGGACGTTACCAAATCTTCTCTGATCTACACCGCTCCGCA
GGAATTCTTCTTCACCCAGATCAAAGTTGCTATGTTCGGTGCTATGGTTATCTCTTTCCCGG
TTATCGCTTCTCAGCTGTACAAATTCGTTGCTCCGGGTCTGTACAAAAACGAACGTGCTGCT
TTCCTGCCGTTCCTGGTTGCTTCTCCGATCCTGTTCCTGATCGGTGCTGCTCTGGTTTACTT
CTTCTTCACCCCGATGGTTATGTGGTTCTTCCTGGCTATGCAGCAGCTGCCGGAAGACGGTG
AAGTTGCTATCCATCATCACCATCACCATTAACTCGA

Amino acid sequence:

MMGSFSVWHWLIVLVIVLVLFGRGKIPELMGDVAKGIKSFKKGMADEDQTPPPADANTKTVD
HKADEIK-
MFDIGWSELLVIAVVLIVVVGPKDLPPMIRAFGKTMAGLRKMAGDFRTQFDEALKEADMDDV
RQTISDVRNLNPTNSLRDAMNPLRQLGNEIKSDLQKATTPPEGLSSSPATSTAAPATSEPVA
PLVSIPEPEMKLPDTPPVTASAPVAAPVVPAEKPKRARAKSIATVEAEAVAAKPKRASRAKA
VAASEPVAVNSTETSAAKPTVKAVARKAAARKVAAEKPVVAADAKPVKPARTKAAKPKKDEA
-
MSGDIEDKPQPLIEHLMELRTRLIWSLGAFFVAFIACFAVAKHLFNLLVIPYKWAVLWAGLD
VTKSSLIYTAPQEFFFTQIKVAMFGAMVISFPVIASQLYKFVAPGLYKNERAAFLPFLVASP
ILFLIGAALVYFFFTPMVMWFFLAMQQLPEDGEVAIHHHHHH-L

8.3. Base pair and amino acid sequence of Clone 1, containing the gene encoding the *tatABC* operon of *A. tumefaciens*

Base pair sequence:

ATGATGGGTTCTTTCTCTGTTTGGCACTGGCTGATCGTTCTGGTTATCGTTCTGGTTCTGTT
CGGTCGTGGTAAAATCCCGGAACTGATGGGTGACGTTGCTAAAGGTATCAAATCTTTCAAAA
AAGGTATGGCTGACGAAGACCAGACCCCGCCGCCGGCTGACGCTAACACCAAAACCGTTGAC
CACAAAGCTGACGAAATCAAATAAATGTTCGACATCGGTTGGTCTGAACTGCTGGTTATCGC
TGTTGTTCTGATCGTTGTTGTTGGTCCGAAAGACCTGCCGCCGATGATCCGTGCTTTCGGTA
AAACCATGGCTGGTCTGCGTAGAATGGCTGGTGACTTCCGTACCCAGTTCGACGAAGCTCTG
AAAGAAGCTGACATGGACGACGTTCGTCAGACCATCTCTGACGTTCGTAACCTGAACCCGAC
CAACTCTCTGCGTGACGCTATGAACCCGCTGCGTCAGCTGGGTAACGAAATCAAATCTGACC
TGCAGAAAGCTACCACCCCGCCGGAAGGTCTGTCTTCTTCTCCGGCTACCTCTACCGCTGCT
CCGGCTACCTCTGAACCGGTTGCTCCGCTGGTTTCTATCCCGGAACCGGAAATGAAACTGCC
GGACACCCCGCCGGTTACCGCTTCTGCTCCGGTTGCTGCTCCGGTTGTTCCGGCTGAAAAAC
CGAAACGTGCTCGTGCTAAATCTATCGCTACCGTTGAAGCTGAAGCTGTTGCTGCTAAACCG
AAACGTGCTTCTCGTGCTAAAGCTGTTGCTGCTTCTGAACCGGTTGCTGTTAACTCTACCGA
AACCTCTGCTGCTAAACCGACCGTTAAAGCTGTTGCTCGTAAAGCTGCTGCTCGTAAAGTTG
CTGCTGAAAAACCGGTTGTTGCTGCTGACGCTAAACCGGTTAAACCGGCTCGTACCAAAGCT
GCTAAACCGAAAAAAGACGAGGCTTAAATGTCTGGTGACATCGAAGACAAACCGCAGCCGCT
GATCGAACACCTGATGGAACTGCGTACCCGTCTGATCTGGTCTCTGGGTGCTTTCTTCGTTG
CTTTCATCGCTTGCTTCGCTGTTGCTAAACACCTGTTCAACCTGCTGGTTATCCCGTACAAA
TGGGCTGTTCTGTGGGCTGGTCTGGACGTTACCAAATCTTCTCTGATCTACACCGCTCCGCA
GGAATTCTTCTTCACCCAGATCAAAGTTGCTATGTTCGGTGCTATGGTTATCTCTTTCCCGG
TTATCGCTTCTCAGCTGTACAAATTCGTTGCTCCGGGTCTGTACAAAAACGAACGTGCTGCT
TTCCTGCCGTTCCTGGTTGCTTCTCCGATCCTGTTCCTGATCGGTGCTGCTCTGGTTTACTT
CTTCTTCACCCCGATGGTTATGTGGTTCTTCCTGGCTATGCAGCAGCTGCCGGAAGACGGTG
AAGTTGCTATCCATCATCACCATCACCATTAACTCGA

Amino acid sequence:

MMGSFSVWHWLIVLVIVLVLFGRGKIPELMGDVAKGIKSFKKGMADEDQTPPPADANTKTVD
HKADEIK-
MFDIGWSELLVIAVVLIVVVGPKDLPPMIRAFGKTMAGLRRMAGDFRTQFDEALKEADMDDV
RQTISDVRNLNPTNSLRDAMNPLRQLGNEIKSDLQKATTPPEGLSSSPATSTAAPATSEPVA
PLVSIPEPEMKLPDTPPVTASAPVAAPVVPAEKPKRARAKSIATVEAEAVAAKPKRASRAKA
VAASEPVAVNSTETSAAKPTVKAVARKAAARKVAAEKPVVAADAKPVKPARTKAAKPKKDEA
-
MSGDIEDKPQPLIEHLMELRTRLIWSLGAFFVAFIACFAVAKHLFNLLVIPYKWAVLWAGLD
VTKSSLIYTAPQEFFFTQIKVAMFGAMVISFPVIASQLYKFVAPGLYKNERAAFLPFLVASP
ILFLIGAALVYFFFTPMVMWFFLAMQQLPEDGEVAIHHHHHH-L

8.4. Base pair and amino acid sequence of Clone 2, containing the gene encoding the *tatABC* operon of *A. tumefaciens*

Base pair sequence:

```
ATGATGGGTTCTTTCTCTGTTTGGCACTGGCTGATCGTTCTGGTTATCGTTCTGGTTCTGTT
CGGTCGTGGTAAAATCCCGGAACTGATGGGTGACGTTGCTAAAGGTATCAAATCTTTCAAAA
AAGGTATGGCTGACGAAGACCAGACCCCGCCGCCGGCTGACGCTAACACCAAAACCGTTGAC
CACAAAGCTGACGAAATCAAATAAATGTTCGACATCGGTTGGTCTGAACTGCTGGTTATCGC
TGTTGTTCTGATCGTTGTTGTTGGTCCGAAAGACCTGCCGCCGATGATCCGTGCTTTCGGTA
AAACCATGGCTGGTCTGCGTAAAATGGCTGGTGACTTCCGTACCCAGTTCGACGAAGCTCTG
AAAGAAGCTGACATGGACGACGTTCGTCAGACCATCTCTGACGTTCGTAACCTGAACCCGAC
CAACTCTCTGCGTGACGCTATGAACCCGCTGCGTCAGCTGGGTAACGAAATCAAATCTGACC
TGCAGAAAGCTACCACCCCGCCGGAAGGTCTGTCTTCTTCTCCGGCTACCTCTACCGCTGCT
CCGGCTACCTCTGAACCGGTTGCTCCGCTGGTTTCTATCCCGGAACCGGAAATGAAACTGCC
GGACACCCCGCCGGTTACCGCTTCTGCTCCGGTTGCTGCTCCGGTTGTTCCGGCTGAAAAAC
CGAAACGTGCTCGTGCTAAATCTATCGCTACCGTTGAAGCTGAAGCTGTTGCTGCTAAACCG
AAACGTGCTTCTCGTGCTAAAGCTGTTGCTGCTTCTGAACCGGTTGCTGTTAACTCTACCGA
AACCTCTGCTGCTAAACCGACCGTTAAAGCTGTTGCTCGTAAAGCTGCTGCTCGTAAAGTTG
CTGCTGAAAAACCGGTTGTTGCTGCTGACGCTAAACCGGTTAAACCGGCTCGTACCAAAGCT
GCTAAACCGAAAAAAGACGAGGCTTAAATGTCTGGTGACATCGAAGACAAACCGCAGCCGCT
GATCGAACACCTGATGGAACTGCGTACCCGTCTGATCTGGTCTCTGGGTGCTTTCTTCGTTG
CTTTCATCGCTTGCTTCGCTGTTGCTAAACACCTGTTCAACCTGCTGGTTATCCCGTACAAA
TGGGCTGTTCTGTGGGCTGGTCTGGACGTTACCAAATCTTCTCTGATCTACACCGCTCCGCA
GGAATTCTTCTTCACCCAGATCAAAGTTGCTATGTTCGGTGCTATGGTTATCTCTTTCCCGG
TTATCGCTTCTCAGCTGTACAAATTCGTTGCTCCGGGTCTGTACAAAAACGAACGTGCTGCT
TTCCTGCCGTTCCTGGTTGCTTCTCCGATCCTGTTCCTGATCGGTGCTGCTCTGGTTTACTT
CTTCTTCACCCCGATGGTTATGTGGTTCTTCCTGGCTATGCAGCAGCTGCCGGAAGACGGTG
AAGTTGCTATCCATCATCACCATCACCATTAACTCGA
```

Amino acid sequence:

```
MMGSFSVWHWLIVLVIVLVLFGRGKIPELMGDVAKGIKSFKKGMADEDQTPPPADANTKTVD
HKADEIK-
MFDIGWSELLVIAVVLIVVVGPKDLPPMIRAFGKTMAGLRKMAGDFRTQFDEALKEADMDDV
RQTISDVRNLNPTNSLRDAMNPLRQLGNEIKSDLQKATTPPEGLSSSPATSTAAPATSEPVA
PLVSIPEPEMKLPDTPPVTASAPVAAPVVPAEKPKRARAKSIATVEAEAVAAKPKRASRAKA
VAASEPVAVNSTETSAAKPTVKAVARKAAARKVAAEKPVVAADAKPVKPARTKAAKPKKDEA
-
MSGDIEDKPQPLIEHLMELRTRLIWSLGAFFVAFIACFAVAKHLFNLLVIPYKWAVLWAGLD
VTKSSLIYTAPQEFFFTQIKVAMFGAMVISFPVIASQLYKFVAPGLYKNERAAFLPFLVASP
ILFLIGAALVYFFFTPMVMWFFLAMQQLPEDGEVAIHHHHHH-L
```

8.5. Alignment of Clone 1 insertion to the *A. tumefaciens tatABC* operon

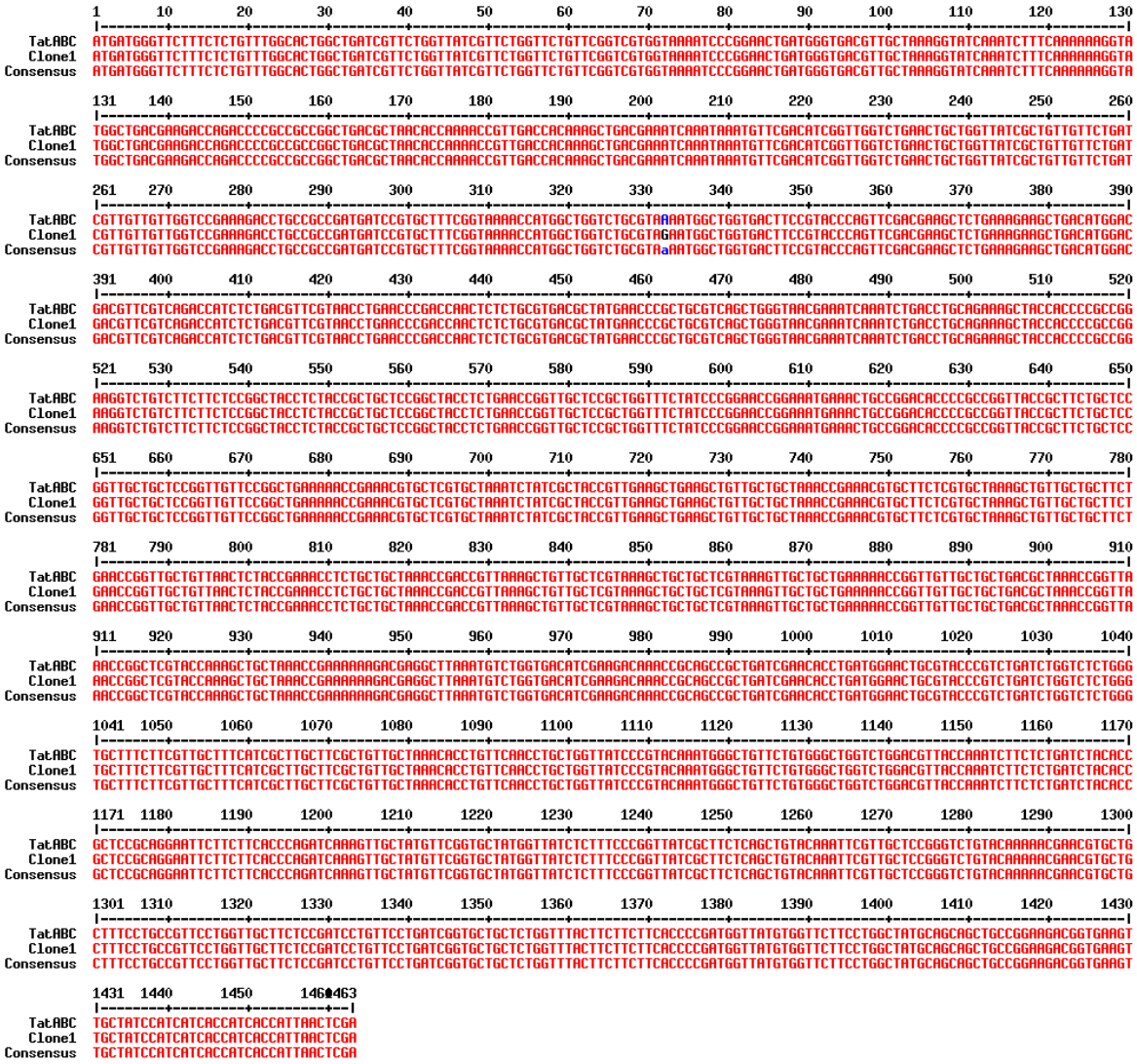


Figure 62: **Here shown is the alignment of the *A. tumefaciens tatABC* sequence to the inserted sequence in Clone 1.** The sequence has been aligned using multalin (http://multalin.toulouse.inra.fr/multalin/multalin.html).

8.6. Alignment of Clone 2 insertion to the *A. tumefaciens tatABC* operon

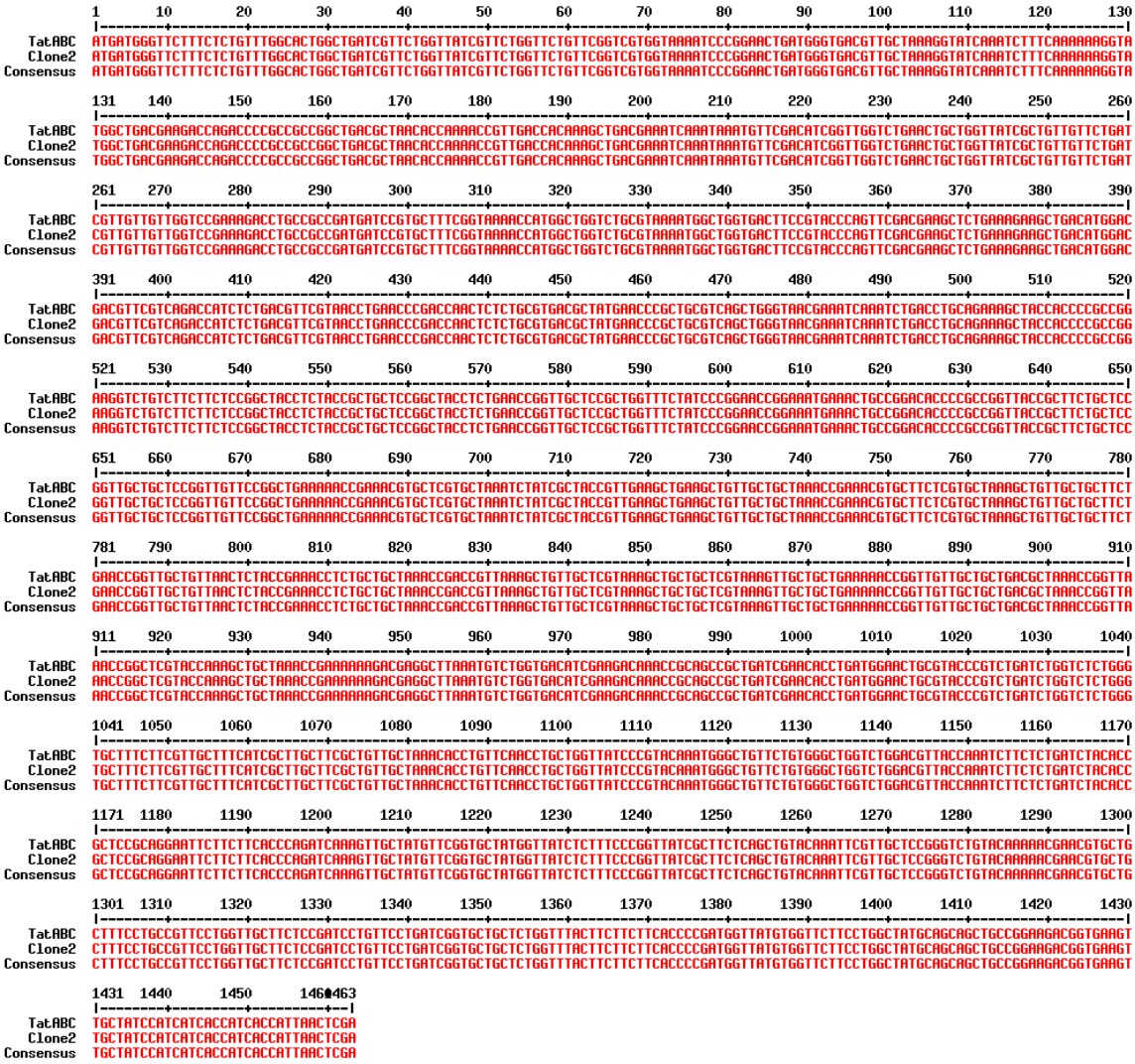


Figure 63: **Here shown is the alignment of the *A. tumefaciens tatABC* sequence to the inserted sequence in Clone 2.** The sequence has been aligned using multalin (http://multalin.toulouse.inra.fr/multalin/multalin.html).