Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Sigurd Tømmerberg Humlebrekk

Physiological and Metabolic Consequences of Expression Vector Presence in *Escherichia coli*

Master's thesis in Industrial chemistry and biotechnology
Supervisor: Per Bruheim
Co-supervisor: Laura García Calvo
February 2022

NTNU
Norwegian University of
Science and Technology

Sigurd Tømmerberg Humlebrekk

Physiological and Metabolic Consequences of Expression Vector Presence in *Escherichia coli*

Master's thesis in Industrial chemistry and biotechnology
Supervisor: Per Bruheim
Co-supervisor: Laura García Calvo
February 2022

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of
Science and Technology

Abstract

Escherichia coli is one of the most commonly used hosts for recombinant protein production, although the bacterium has some drawbacks that can negatively impact protein yields. Expression vectors place a metabolic burden on the host, leading to physiological and metabolic consequences. These consequences complicate the development and optimization of recombinant protein production processes. The metabolic burden on the host caused by expression vectors presence was studied using recombinant *E. coli* BL21 transformed with expression vectors with two different strengths of the inducible XylS/*Pm* expression system and two variations in the plasmid copy numbers. For the different combinations of expression system strength and plasmid copy number studied, there was one strain expressing the reporter protein mCherry and a negative control strain lacking the mCherry encoding gene. Batch cultivations in bench-scale bioreactors were performed to study the growth characteristics of the different strains. The extracellular concentration of substrates and products throughout the cultivations were quantified using high-performance liquid chromatography. Reverse phase liquid chromatography-tandem mass spectrometry was used to quantify the intracellular concentrations of amino acids and TCA cycle intermediates to investigate how the expression vectors affected the intracellular pools of these metabolites. Finally, microscopy was used to compare cellular lengths between the different strains.

The results demonstrated that the metabolic burden expression vectors placed on the host lead to physiological and metabolic consequences. It was observed that the theoretical degree of metabolic burden implied by increased plasmid copy number, expression system strength, and expression of the reporter protein mCherry, was reflected in the severity of these consequences. The recombinant strains transformed with expression vectors having a higher theoretical degree of metabolic burden had a significantly greater reduction in growth rate after induction and increased secretion of overflow metabolites compared to the other strains. The strain harboring the expression vector with the highest degree of theoretical metabolic burden was observed to have the most severe physiological and metabolic consequences. The strain had complete growth inhibition after induction of mCherry expression, decreased biomass formation, increased cell filamentation, an adaption of the central carbon metabolism towards energy-generating pathways, and increased secretion of the overflow metabolites acetic acid and lactic acid.

Sammendrag

Escherichia coli er en av de mest brukte vertsorganismene for rekombinant proteinproduksjon, selv om bakterien har ulemper som kan redusere proteinutbytte. Ekspresjonsvektorer kan påføre produksjonsorganismen en metabolsk byrde, som kan føre til fysiologiske og metabolske konsekvenser. Disse konsekvensene kompliserer utviklingen og optimaliseringen av rekombinante bioprosesser. Den metabolske byrden ekspresjonsvektorer påfører produksjonsorganismen ble studert ved hjelp av rekombinant *E. coli* BL21 transformert med ekspresjonsvektorer med to forskjellige styrker av det induserbare XylS/*Pm* ekspresjonssystemet og to variasjoner i plasmidkopinumre. For de ulike kombinasjonene av ekspresjonsvektorstyrkene og plasmidkopinumrene studert var det en stamme som produserte reporterprotein mCherry og en kontrollstamme som manglet det mCherry-kodende genet. Batch kultiveringer i labskala bioreaktorer ble utført for å studere vekstegenskapene til de ulike stammene. Den ekstracellulære konsentrasjonen av substrat og produkter gjennom kultiveringen ble kvantifisert ved hjelp av høypresisjonsvæskekromatografi. Revers fase væskekromatografi-tandem massespektrometri ble brukt til å kvantifisere den intracellulære konsentrasjonen av aminosyrer og TCA syklus metabolitter for å studere hvordan ekspresjonsvektorene påvirket de intracellulære konsentrasjonene av metabolittene. Mikroskopering ble brukt til å sammenligne de cellulære lengdene mellom de ulike stammene.

Resultatene demonstrerte at den metabolske byrden ekspresjonsvektorene påførte vertsorganismen førte til fysiologiske og metabolske konsekvenser. Det ble observert at den teoretiske mengden metabolsk byrde implisert av økende plasmidkopitall, styrke på ekspresjonssystemet og produksjon av reporterproteinet mCherry ble gjenspeilet i graden av fysiologiske og metabolske konsekvenser. De rekombinante stammene transformert med ekspresjonsvektorer med høy grad av teoretisk byrde hadde signifikant mer reduserte veksthastigheter og økt produksjon av overstrømningsmetabolitter etter induksjon sammenlignet med de andre stammene. Stammen som var transformert med ekspresjonsvektoren som hadde høyest grad av teoretisk metabolsk byrde ble observert til å ha høyest grad av fysiologiske og metabolske konsekvenser. Stammen hadde fullstendig vekststopp etter indusering av mCherry produksjon, redusert biomassekonsentrasjon, større grad av filamentøse celler, tilpasning av metabolismen mot energi-produserende reaksjonsspor, og økte ekstracellulær konsentrasjon av overstrømmingsmetabolittene eddiksyre og melkesyre.

Acknowledgements

This thesis was conducted at the Depertment of Biotechnology and Food Science at the Norwegian University of Science and Technology (NTNU).

I would like to express my gratitude to my main supervisor Prof. Per Bruheim, for letting me work on the MetaboProt project and supervising me throughout the project. Your constructive feedback and quick replies have been much appreciated. I would like to thank my co-supervisor Dr. Laura García Calvo for all the valuable days and late evenings you spent with me in the lab, for your guidance throughout the project and for passing on invaluable knowledge. My appreciation also go out the other members of the Bruheim lab. To Nikalet Everson and Lise Femanger Mathiassen for being great lab partners. To Lisa Marie Røst and Mi Jang for performing the masspectrometry analysis and helping me interpret the data. Last but not least, I would like to thank my friends, family, and fellow students for the support and encouragements throughout the project.

Contents

Abbreviations

Abbreviations	Explanation
5'-UTR	5'-Untranslated region of mRNA
AA	Amino acid
ACN	Acetonitrile
ANOVA	Analysis of variance
BSA	Bovine serum albumin
CapIC	Capillary ion chromatography
DCW	Dry cell weight
DO	Dissolved oxygen
EDAC	N-(3-Dimethylaminopropyl)-N'-ethylcarbodimide hydrochloride
ESTD	External standard
EtOH	Ethanol
H_0	Null hypothesis
H_1	Alternative hypothesis
HPLC	High performance liquid chromatography
ISTD	Internal standard
MeOH	Methanol
MQ	Milli-Q
MS/MS	Tandem mass spectrometry
O-BHA	Ortho-benzylhydroxylamine
OA	Organic acid
OD_{600}	Optical density at 600 nm
OD_{595}	Optical density at 595 nm
PBS	Phosphate buffered saline
PC	Principal component
PCA	Principal component analysis
PCN	Plasmid copy number
PITC	Phenyl isothiocyanate
ppGpp	Guanosine tetra- or pentaphosphate
PPP	Pentose phosphate pathway
QC	Quality control
RP	Recombinant protein
RP LC	Reverse phase liquid chromatography
RPP	Recombinant protein production
RT	Room temperature
S/N	Signal-to-noise ratio
TCA	Tricarboxylic acid
WT	Wild type

* Metabolite abbreviations are listed in Appendix A

Chapter 1

Introduction

1.1 Recombinant protein production

Recombinant proteins (RP) have been defined by The Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine as proteins originating from recombinant DNA technology (Ganten *et al.*, 2006). RPs are procured by cloning a gene sequence of the protein of interest into an appropriate expression vector, then transforming it into an expression host. The expression host will then have the ability to produce the protein of interest (Meagher *et al.*, 1977; Rosano *et al.*, 2019). Thus, recombinant DNA technology makes it possible for expression hosts to produce proteins they would not be able to produce under natural circumstances (Waegeman and Soetaert, 2011).

The first reported production of a functional RP was in 1977 when *Escherichia coli* (*E. coli*) was used to produce the human hormone somatostatin (Itakura *et al.*, 1977). Shortly after, in 1982, human insulin was produced in *E. coli* was the first RP drug to be accepted by the Food and Drug Administration and launched on the marked (Johnson, 1983). Since then, recombinant protein production (RPP) has revolutionized biochemistry and become one of the key components supporting the progress of biotechnology over the last decades. Technological advancements have made it possible to produce and purify protein on a large scale (Kirk 2002), genetically engineer host strains to optimize RPP (Nielsen, 2001) and develop tailor-made proteins optimized for specific processes by protein engineering (Kirk *et al.*, 2002).

1.1.1 Importance and applications of recombinant protein production

The development of RPP has significantly revolutionized the biopharmaceutical and bioprocessing industries. Prior to RPP, proteins were obtained from animals and plants. Large amounts of tissue or biological fluids were required to produce proteins at a meager yield with laborious downstream processing (Waegeman and Soetaert, 2011; Rosano *et al.*, 2019). RPP using microbial hosts has made it possible to produce and purify proteins at much greater quantities with simpler methods (Rosano *et al.*, 2019).

Recombinant proteins have a wide range of applications within a broad set of industries. Enzymes (lipases, amylase, proteases, etc.) produced by recombinant protein production are used to produce food, beverages, and animal feed, convert starch into fuel, as components in textile products, cleaning supplies, cosmetics, biopolymers, waste management, and so on (Kirk *et al.*, 2002; Puetz and Wurm, 2019). In the biopharmaceutical industry, recombinant protein production is used to produce biopharmaceutical drugs, including hormones, human growth factors, cytokines, and antibodies (Walsh, 2018). As of 2018, over 300 protein-based drugs have been approved in Europe and USA, and almost a third of all pharmaceuticals in development are protein-based (Walsh, 2018). In biological research, the majority of commonly used enzymes such as thermostable DNA polymerases and restriction endonucleases are produced using recombinant protein production (Wyre, 2015). The market for recombinant proteins is steadily increasing, with an estimated market value at USD 2.2 billion in 2020, predicted to reach USD 4.1 billion in 2025 (Markets and Markets, 2020).

1.1.2 *Escherichia coli* as expression host

There are many different expression hosts available, including bacterial (Terpe, 2006), yeast (Mattanovich *et al.*, 2012), filamentous fungi (Punt *et al.*, 2002), transgenic plants (Kusnadi *et al.*, 1997), and animal cells such as insect and mammalian cells (Wurm, 2004; Kost *et al.*, 2005). Bacterial expression hosts are preferable for the production of many recombinant proteins from both bacterial and eukaryotic origin due to their cost-effectiveness, well-characterized genetics, and a large number of available expression systems (Terpe, 2006). One of the most commonly used bacterial expression hosts is the gram-negative, facultative anaerobic, rod-shaped bacterium *Escherichia coli* (Tenaillon *et al.*, 2010; Waegeman and Soetaert, 2011).

E. coli had a central role in the development of recombinant protein production. Therefore, a wide selection of expression systems and genetically engineered *E. coli* strains optimized for recombinant protein production are available. Other advantages *E. coli* has as an expression host is that the bacterium has fast growth kinetics, and it can be grown to high cell densities in inexpensive media (Rosano and Ceccarelli, 2014; Tripathi and Shrivastava, 2019). It has been observed that *E. coli* has doubling times around 20 min when grown in optimal conditions (Sezonov *et al.*, 2007), and can grow to cell densities up to 190 g/L cell dry weight (CDW) (Nakano *et al.*, 1997; Shiloach and Fass, 2005). Additionally, the bacterium is well-characterized genetically, physiologically, and metabolically, and there are many molecular biology tools and techniques available to engineer novel *E. coli* strains (Waegeman and Soetaert, 2011; Overton, 2014).

However, *E. coli* also has some major drawbacks as an expression host. The bacterium cannot perform many of the post-translational modifications found in higher organisms, including glycosylation and formation of disulfide bonds, which is crucial for correct folding of complex proteins (Waegeman and Soetaert, 2011). *E. coli* also has poor secretion capabilities, leading to aggregation of proteins inside the cell (Wyre, 2015). Other disadvantages include plasmid

instability, codon bias, formation of inclusion body, and endotoxin issues (Palomares *et al.*, 2004; Waegeman and Soetaert, 2011; Tripathi and Shrivastava, 2019). These limitations have been widely researched, and many strategies have been developed (Palomares *et al.*, 2004; Francis and Page, 2010; Gupta and Shukla, 2016; Kaur *et al.*, 2018).

1.2 Expression vector

1.2.1 RK2 Plasmid

Plasmids are extrachromosomal cytoplasmic DNA elements that can replicate autonomously. They are commonly used as molecular vehicles to transport recombinant genes (Palomares *et al.*, 2004). The RK2 plasmid belongs to the incompatibility group IncP, which is known for its large host range, being able to replicate and maintain in most gram-negative bacteria, including *E. coli* (Fang and Helinski, 1991). The RK2 plasmid is a self-transmissible plasmid, which means the plasmid carries the genes needed for its transfer by conjugation (Fang and Helinski, 1991; Smillie *et al.*, 2010). The replication of RK2 plasmids are controlled by two locus on the plasmid, the origin of replication *oriV*, and the *trans*-acting *trfA* gene encoding TrfA proteins that binds and activates *oriV* to initiate replication (Durland *et al.*, 1990).

RK2 plasmids replicate at a defined plasmid copy number (PCN) (Toukdarian and Helinski, 1998), meaning RK2 plasmids replicates to a specific number of functional plasmids in the cell (Plotka *et al.*, 2017). The PCN is maintained by a mechanism called handcuffing, which is initiated when the number of plasmids reaches or exceeds the characteristic PCN (Toukdarian and Helinski, 1998). All Trfa-bound plasmids will then be reversibly coupled at their origins, causing steric hindrance of origin activity (Blasina *et al.*, 1996). The PCN of wild type (WT) RK2 plasmids in *E. coli* has been estimated to be five to seven per chromosome. The PCN can be increased by introducing point mutations in the *trfa* gene, with observed PCN up to 24-fold of the WT PCN in *E. coli* (Blatny *et al.*, 1997).

1.2.2 The *Pm/xylS* expression system

The XylS/*Pm* regulator/promoter system originates from the TOL plasmid pWW0, nativly found in *Pseudomonas putida* (Gawin *et al.*, 2017). The plasmid carries genes encoding toluene and xylene degrading enzymes (Greated *et al.*, 2002). The *Pm/xylS* expression system consists of the positively regulated promoter *Pm*, the regulator gene *xylS*, and a 5'-untranslated region (5'-UTR) (Gawin *et al.*, 2017). *xylS* encodes the transcription factor XylS, which in the presence of benzoates induces transcription from *Pm* Winther-Larsen *et al.* (2000); Gawin *et al.* (2017). It has been observed that the expression system is capable of producing proteins in at industrial levels at high cell density cultures of *E. coli* (Sletta *et al.*, 2004, 2007). The advantages of the *Pm/xylS* system are that the inducer is cheap, it has a broad host range, and the expression levels can easily be find tuned (Balzer *et al.*, 2013). The wild-type textitPm/xylS

system is strong, but the expression levels can be greatly increased by introducing mutations in the *Pm*, *xylS* gene or 5'-UTR (Zwick *et al.*, 2012).

1.2.3 Ampicillin resistance as selection marker

Recombinant protein production imposes a large metabolic burden on the producing cells, leading to decreased growth rates. To prevent plasmid-free nonproducing cells from outgrowing the producing cells, positive selection marker genes are cloned into the expression vector (Miki and McHugh, 2004; Palomares *et al.*, 2004; Rosano *et al.*, 2019). The most common strategy at the laboratory scale is to use antibiotic resistance genes as selection markers. The antibiotic can then be supplemented to the media, thereby creating a selective pressure where only producing cells carrying the expression vector can grow (Rosano *et al.*, 2019). The *bla* gene, which gives resistance to the antibiotic ampicillin (Amp), is a commonly used selection marker for protein expression in *E. coli* (Feizollahzadeh *et al.*, 2017). Amp is a β-lactam antibiotic that binds to the enzyme transpeptidase, thereby inhibiting the bacterial cell wall synthesis (Rafailidis *et al.*, 2007). The *bla* gene encodes the periplasmic enzymes β-lactamase that inactivates β-lactam antibiotics. Thus, cells carrying the *bla* gene will be able to grow in the presence of Amp (Rosano and Ceccarelli, 2014).

1.2.4 The reporter protein mCherry

Reporter genes have easily identifiable and measurable phenotypes that can be used to indicate the presence or absence of the gene product (Naylor, 1999). They can be used to differentiate transformed and non-transformed organisms or material (Miki and McHugh, 2004). Therefore, it is crucial that the reporter gene-phenotype is easily distinguishable from the background of endogenous proteins (Naylor, 1999). Reporter proteins differ from selection markers as they do not give transformed cells selective advantages (Miki and McHugh, 2004). The choice of reporter gene depends on the sensitivity, convenience, dynamic range, and reliability in assays (Naylor, 1999). Genes encoding fluorescence proteins have been widely used as reporter genes to study gene products and their *in vivo* dynamics (Heppert *et al.*, 2016). One such reporter protein is the monomeric red fluorescent protein mCherry, which belongs to the mFruit series (Shaner *et al.*, 2005; Day and Davidson, 2009). mCherry has peak excitation at 587 nm and emission maxima at 610 nm, and has shown to have good brightness, photostability, and performance in fusions (Shaner *et al.*, 2005; Day and Davidson, 2009; Davidson and Campbell, 2009). These properties makes mCherry suitable for both *in vivo* and *in vitro* studies (Lagendijk *et al.*, 2010).

1.3 Physiological and metabolic consequences due to plasmid-imposed metabolic burden

The presence of expression vectors imposes a burden on the cell, as maintenance of the plasmid and expression of the plasmid-encoded genes drain the cell of its energy and metabolic precursors. This drainage of cellular resources has been termed metabolic burden (Glick, 1995; Hoffmann and Rinas, 2004). The metabolic burden imposed by expression vectors increases with the expression vector's size, PCN, and expression cassette strength (Hoffmann and Rinas, 2004; Zwick *et al.*, 2012). The drainage of cellular resources may lead to physiological consequences, including growth inhibition (Ow *et al.*, 2006), reduced cell viability (Silva *et al.*, 2012), cell filamentation (Jeong and Lee, 2003), expression vector loss (Palomares *et al.*, 2004), induction of stress responses (Haddadin and Harcum, 2005) and metabolic alterations (Hoffmann and Rinas, 2004), which can have negative effects on recombinant production yields and quality (Glick, 1995; Hoffmann and Rinas, 2004).

1.3.1 Growth inhibition

Reduction in growth rate and decreased biomass yield are the most notable consequences of expression vector presence (Hoffmann and Rinas, 2004). The expression of housekeeping and growth-related genes is reduced as the competition over cellular resources increases as a consequence of the metabolic burden imposed by the expression vector (Neubauer *et al.*, 2003). This increased competition over resources leads to reduced growth rates, culturability, biomass yields, and in the worst case, cell death (Hoffmann and Rinas, 2004; Silva *et al.*, 2012). The energy demand of the expression vector presence leads to enhanced respiration to generate ATP, which results in loss of carbon at the expense of biomass formation (Glick, 1995). Additionally, high-level recombinant protein production can lead to the induction of a stringent-like stress response, which will cause the cells to enter complete growth arrest. This will be further elaborated in Subsection 1.4.2.

1.3.2 Filamentation of *E. coli*

Filamentation is used to describe the abnormal growth where bacterial cells continue growing without dividing, resulting in elongated cells with multiple chromosomal copies (Jaimes-Lizcano *et al.*, 2014). Cell filamentation has been reported to be a consequence of high-level recombinant protein production (Jeong and Lee, 2003; Lederer *et al.*, 2011). As filamentation is a consequence of cell division blockage, it has been hypothesized that metabolic burden may lead to inactivation of essential cell division proteins (Silva *et al.*, 2012). Ow *et al.* (2006) observed that genes associated with cell wall synthesis and septal division are down-regulated in cells carrying expression vectors. Additionally, filamentation can be a result of the SOS stress response (Silva *et al.*, 2012), which will be further elaborated in Subsection 1.4.3. Cell filamentation leads to decreased growth rates or complete growth arrest and decreased pro-

ductivity of the recombinant protein (Jeong and Lee, 2003; Silva *et al.*, 2012).

1.3.3 Loss of expression vectors

The metabolic burden expression vectors place on the host leads to plasmid-free cells lacking this burden to outgrow the cells containing expression vectors(Waegeman and Soetaert, 2011). Several mechanisms may lead to reduced numbers of functional plasmids, including plasmid segregation instability, which occurs upon cell division when plasmids are unevenly divided between daughter cells(Paulsson and Ehrenberg, 2001). Reduced plasmid numbers over time can result in the formation of plasmid-free cells or cells with low PCN, which will outgrow the cells harboring expression vectors and thus significantly reduce product yields (Palomares *et al.*, 2004). Several strategies have been developed to prevent plasmid loss during recombinant protein production. A commonly used strategy is to incorporate a resistance gene to a selection marker (e.g. an antibiotic) into the expression vector, then add that selection marker to the culture media. The selection marker will then create a selection pressure that favors cells harboring the expression vector (Palomares *et al.*, 2004; Feizollahzadeh *et al.*, 2017). The use of antibiotics (e.g. ampicillin) as selection markers has shown to be ineffective if the antibiotic is inactivated, degraded, or if the antibiotic resistance enzyme leaks out of the cells harboring expression vectors (Palomares *et al.*, 2004)

1.3.4 Metabolic flux adjustments

The metabolic burden imposed by expression vectors increases the energy demand in the cell. The cell adjusts its metabolic fluxes towards energy-generating pathways to meet the high energy demands (Shiloach and Fass, 2005). How the cells adjust their metabolism depends on the cultivation conditions, especially the access to glucose, but increased respiration for ATP regeneration is observed in all cases (Hoffmann and Rinas, 2004; Shiloach and Fass, 2005). The increased fluxes towards energy-generating pathways lead to lower flux towards anabolic pathways, which results in lower biomass formation (Hoffmann and Rinas, 2004).

Under anaerobic conditions, *E. coli* generates energy by performing mixed acid fermentation where products like acetic acid, lactic acid, succinic acid, and ethanol are produced and secreted (Thakker *et al.*, 2012). This way of generating energy is less efficient than aerobic respiration, where the substrate is completely oxidized by the glycolysis and TCA cycle (Szenk *et al.*, 2017). When *E. coli* are cultivated in conditions with an excess concentration of glucose (e.g. batch cultivation), the cells get their energy by a combination of aerobic respiration and mixed acid fermentation when oxygen is plentiful, even though incomplete oxidation be fermentation is less efficient (Basan *et al.*, 2015; Szenk *et al.*, 2017). This phenomenon has been termed overflow metabolism, and the metabolites excreted are called overflow metabolites (Basan *et al.*, 2015). It has been reported that cells producing recombinant proteins have increased rates of overflow metabolism and excrete higher amounts of overflow metabolites, especially acetate. It has been reported that accumulation of acetate and other overflow metabolites can

inhibit growth and recombinant protein production (Hoffmann and Rinas, 2004; Shiloach and Fass, 2005).

1.4 Stress responses induced by heterologous protein expression

High-level recombinant protein production interferes with the host cell's physiology Hoffmann and Rinas (2004). Plasmid maintenance and expression of plasmid-encoded genes impose a metabolic burden, leading to a high degree of physiological stress in the host (Hoffmann and Rinas, 2004; Carneiro *et al.*, 2013). The physiological stress may elicit cellular stress responses similar to the ones induced by environmental stress, including the heat shock, stringent, and SOS stress response (Hoffmann and Rinas, 2004; Silva *et al.*, 2012).

1.4.1 The heat shock response

The heat shock response is a cellular stress response that allows cells to survive stressful conditions caused by environmental and metabolic changes (Arsène *et al.*, 2000). The stress response has been shown to be induced by a wide range of stress conditions, including sudden temperature upshift, exposure to metabolically harmful substances, and complex metabolic processes (Arsène *et al.*, 2000; Hoffmann and Rinas, 2004). The heat shock response is positively regulated by the transcription factor σ^{32} in *E. coli* (Arsène *et al.*, 2000; Guisbert *et al.*, 2008). The σ^{32} levels are low under steady-state conditions, hence limiting transcription of heat shock proteins (Arsène *et al.*, 2000). The heat shock response is initiated when σ^{32} levels increase rapidly and shut off with declining σ^{32} levels or inhibition of σ^{32} activity. This mode of σ^{32} regulation is rapid and allows fast response and adaption to sudden stress (Arsène *et al.*, 2000; Guisbert *et al.*, 2008).

The heat shock response upregulates the expression of heat shock proteins that are mainly composed of chaperones and proteases. The chaperons (E.g. Clp, DnaK/J and GroEL/S) help proteins fold properly, and proteases (e.g. Lon, ClpP, clpC, HslU/V, ClpQ, and FtsH) degrade unfolded and damaged proteins (Arsène *et al.*, 2000; Guisbert *et al.*, 2008; Silva *et al.*, 2012). The heat shock proteins act as a protein quality folding system (Hoffmann and Rinas, 2004). Previous findings have shown that induction of high-level recombinant protein expression in *E. coli* has lead to increased transcription (Hoffmann and Rinas, 2004; Haddadin and Harcum, 2005; Singha *et al.*, 2017), synthesis (Hoffmann and Rinas, 2004) and accumulation of heat shock proteins (Hoffmann and Rinas, 2004). The presence of high copy DNA plasmids carrying a selection marker and non-expressing 1.8 kb DNA has also been shown to have upregulated transcription of common heat shock proteins (Ow *et al.*, 2006). It has been hypothesized that activation of heat shock response during RPP may be caused by an accumulation of misfolded proteins (Hoffmann and Rinas, 2004). High-level expression of proteins at rates exceeding the capacity of the folding machinery of *E. coli* can lead to accumulation of misfolded

proteins (Vera *et al.*, 2007), which has been reported to trigger heat shock response (Guisbert *et al.*, 2008). The degree of heat shock response during RPP varies with the expression system, the protein expressed, expression rate, and cultivation conditions (Hoffmann and Rinas, 2004).

1.4.2 Starvation and the stringent stress response

When nutrients become limiting, *E. coli* cells reprogram their gene expression from supporting growth towards the stationary phase. This reprogramming of the gene expression has been termed the stringent response and ensures prolonged survival in nutrient limiting environments (Traxler *et al.*, 2008). The reprogramming is triggered by the accumulation of the alarmones guanosine tetra- and pentaphosphate (ppGpp and pppGpp, respectively), collectively referred to as (p)ppGpp. The (p)ppGpp levels are mainly controlled by the enzymes RelA and SpoT (Ferullo and Lovett, 2008; Traxler *et al.*, 2008). Upon starvation RelA, also called (p)ppGpp synthase I, synthesizes (p)ppGpp in response to uncharged tRNA binding RelA-associated ribosomes (Ferullo and Lovett, 2008). SpoT has a weak (p)ppGpp synthase activity reported to be initiated by a wide range of signals, including carbon, iron, and fatty acid starvation (Traxler *et al.*, 2008). Additionally, SpoT has a weak ppGpp hydrolase activity that mediates (p)ppGpp turnover (Ferullo and Lovett, 2008; Traxler *et al.*, 2008). The reaction equation of (p)ppGpp formation is shown in Equation 1.1. Stringent response leads to downregulation of the majority of genes encoding proteins associated with transcription and translation machinery (Hoffmann and Rinas, 2004), and approximately half of the genes involved in amino acid synthesis (Chang *et al.*, 2002), activate expression of specific proteases (Harcum and Bentley, 1999; Silva *et al.*, 2012), and complete growth arrest (Chang *et al.*, 2002).

$$\mathrm{ATP} + \mathrm{GDP/GTP} \xrightarrow{\mathrm{RelA/SpoT}} \mathrm{ppGpp/pppGpp} + \mathrm{AMP} \tag{1.1}$$

Protein expression is one of the most energy-consuming processes in the cell (Shiloach and Fass, 2005). As previously described, expression vector presence drains the cell of metabolic precursors and energy, including amino acids (Carneiro *et al.*, 2013). High-level recombinant protein expression can lead to depletion of amino acid pools, especially if the amino acid composition of the protein expressed is considerably different the average composition of protein in *E. coli*, and can lead to induction of stringent-like responses (Hoffmann and Rinas, 2004). Previous studies have observed that high-level recombinant protein production led to regulation of several genes associated with the stringent stress response (Haddadin and Harcum, 2005), and also increased (p)ppGpp concentration (Cserjan-Puschmann *et al.*, 1999). It has also been observed that strains carrying a mutated *relA* and therefore not able to initiate a stringent-like stress response by amino acid depletion produces higher yields of recombinant proteins, compared to the control strain with an active *relA* gene (Harcum and Bentley, 1999). Kumar *et al.* (2020) performed a study where they supplemented recombinant producing cells with limiting amino acids. They observed using transcriptomics and proteomics that

cell cultures supplemented with limiting amino acids had downregulation of several genes associated with global stress responses, thereby indicating that amino acid depletion may be a major contributor to metabolic stress in recombinant protein producing cells.

1.4.3 The SOS response

The SOS response is initiated in response to damaged DNA or interference with DNA replication (Hoffmann and Rinas, 2004). Physiological stress has also been shown to induce SOS response, including change in pH, starvation, or transition from growth to stationary phase (Majchrzak *et al.*, 2006). The SOS response is regulated by the interaction between the co-protease RecA and the transcriptional repressor LexA. RecA is activated when it binds to single-stranded DNA, where it assists in the cleaving of the transcriptional repressor LexA, thereby allowing transcription of SOS genes (Singha *et al.*, 2017). The SOS response stops cell proliferation (Silva *et al.*, 2012), and induces a complex enzyme cascade that promotes repair of damaged DNA and error-prone DNA synthesis (Hoffmann and Rinas, 2004; Aertsen and Michiels, 2005). Transcriptome studies have observed that transcription of RecA and LexA is upregulated during recombinant protein expression in *E. coli*, indicating that RPP can induce SOS-like stress responses (Hoffmann and Rinas, 2004).

1.5 Aim of study

E. coli is a commonly used expression host for RPP, despite its drawbacks that can lead to reduced protein yields. There is a lack of in-depth knowledge about the metabolic burden associated with expression vector presence in *E. coli*. Combined with the industry's pressure to produce results rapidly, this lack of knowledge has made it more efficient for the industry to utilize "Trial-and-error" approaches to optimize RPP processes. These approaches are not only costly but can also be unpredictable and time-consuming.

The work in this thesis was performed as a part of the MetaboProt project, which aims to provide more knowledge about the metabolic burden imposed by recombinant protein production. The goal of the project is to identify bottlenecks that can be addressed to improve recombinant protein efficiency. In this thesis, the metabolic burden imposed by expression vectors with different combinations of expression system strength, PCN, and plasmid size was studied with the aim of identifying how these factors affected the metabolic burden imposed on the host, and to study the physiological and metabolic consequences caused by the metabolic burden.

Chapter 2

Material and Methods

2.1 Bacterial strains

The bacterial strains studied in this master thesis were *E. coli* BL21 (New England BioLabs, genotype *fhuA2 [lon] ompT gal [dcm]* Δ*hsdS*) and six recombinant *E. coli* BL21 strains transformed with different versions of the pVB expression vector described in Subsection 2.1.1. *E. coli* BL21 is a common laboratory strain for recombinant protein production (Jeong *et al.*, 2015). The strain is deficient in the proteases Lon and OmpT, which leads to less degradation of foreign proteins (Waegeman and Soetaert, 2011). The strains also has a hsdS mutation which disrupts DNA methylation and degradation, and thereby limit degradation of expression vectors (Rosano *et al.*, 2019).

2.1.1 Expression Vectors

The pVB expression vectors studied in this master thesis were genetically modified RK2 plasmids, with two variations of the expression cassette, wild type (A) and mutated XylS/*Pm*/5'-UTR (Z). The Z-variants have a wild type expression control element *xylS* and mutated *Pm* promoter and 5'-UTR. These mutations give the Z-variants a stronger expression cassette. The negative control strains' (A2, Z1, and Z2) expression vectors do not carry the gene encoding the reporter protein mCherry. All expression vectors contain the *bla* gene, encoding ampicillin resistance. The name of the expression vectors, strain abbreviations, expression cassettes, whether it contains the mCherry encoding gene or not, plasmid size, and PCN are listed in Table 2.1. The plasmid maps are given in Appendix B. Vectron BioSolutions AS provided the expression vectors.

Table 2.1 Name, strain abbreviation, *trfA* gene, mCherry encoding gene, plasmid size [bp] and PCN of the expression vectors used in this master thesis. The expression vectors were provided by Vectron BioSolutions AS.

Name of the Expression Vector	Strain abbreviation	*trfA* gene	mCherry gene	Plasmid size [bp]	PCN
pVB-1A0B1	A2	271C mutant	No	8172	20
pVB-1Z0A1	Z1	Wild type	No	8249	5
pVB-1Z0B1	Z2	271C mutant	No	8249	20
pVB-1A0B1-mCherry	A2-mCh	271C mutant	Yes	8930	20
pVB-1Z0A1-mCherry	Z1-mCh	Wild type	Yes	8928	5
pVB-1Z0B1-mCherry	Z2-mCh	271C mutant	Yes	8928	20

2.1.2 Transformation and preparation of freeze stock *E. coli* cultures

E. coli BL21 were transformed with the expression vectors described in Subsection 2.1.1, using a heat shock transformation method adapted from (Hanahan, 1983). Lyophilized vector DNA were centrifuged (6000 rcf, 1 min, 4°C), reconstituted in sterile MQ-H_2O, and mixed by vortex (~1 min). Aliquots of 100 µL competent *E. coli* BL21 were thawed on ice, then 1 µL plasmid solution was added and the tubes were mixed by inversion. The cells were stored on ice for 30 min, then transferred to a water bath at 37°C for 1.5 min. The cells were placed back on ice for 2 min, followed by the addition of 900 µL SOC medium (Subsection 2.2.2.4) at 37°C, and incubation (37°C, 200 rpm, 1 h). The transformation mixes were plated out in volumes of 200 µL onto pre-warmed LB agar plates (Table 2.4) with ampicillin (0.1 mg/mL). The plates were incubated overnight (37°C), then used to prepare glycerol stocks as described below.

The following method was used to prepare freeze stocks of the wild type (WT) *E. coli* BL21, and the recombinant strains described in Subsection 2.1.1. Overnight cultures were prepared by inoculating LB medium (Table 2.3) with single cell colonies from the plates prepared the day before. Ampicillin (0.1 mg/mL) was added to recombinant strains, and the cultures were incubated overnight (37 °C, 200 rpm, 16 ± 1 h). The overnight cultures were transferred to sterile 50 mL centrifuge tubes, centrifuged (4000 rcf, 5 min, RT), and the supernatants were partially discarded to reduce the volumes to 12 mL. The pellets were resuspended in the remaining supernatant, and glycerol (60%, VWR, 24387.292) was added to a concentration of 20% (v/v) glycerol. The tubes were mixed by inversion, and aliquoted in 1 mL volues into sterile cryotubes. The glycerol stocks were stored at -80°C until use.

2.2 Chemicals and stock solutions

2.2.1 Stock solutions

2.2.1.1 Trace element stock solutions

Trace element stock solutions were made separately by dissolving the compounds in MQ-H_2O to the concentrations given in Table 2.2. The solutions were then filter-sterilized through syringe filters (0.2 µm PTFE membrane, 13 mm, VWR, 514-4122) into individual sterile containers and stored at room temperature (RT).

Table 2.2 The composition and concentration [g/L] of the trace element stock solutions, and supplier and product number of the compounds. The compounds were dissolved in MQ-H_2O, filter-sterilized, and stored at RT.

Stock solution (concentration)	Concentration [g/L]	Supplier	Product number
$Zn(CH_33COO)_2 \cdot 2H_2O$ (500x)	4	*	-
Fe(III) citrate hydrate (100x)	6	Sigma-Aldrich	F3388-250G
$MnCl_2 \cdot 4H_2O$ (666.67x)	10	Sigma-Aldrich	M5005-100G
$CuCl_2 \cdot 2H_2O$ (10 000x)	15	Sigma-Aldrich	C3279-100G
$CoCl_2 \cdot 6H_2O$ (10 000x)	25	Sigma-Aldrich	C8661-25G
$Na_2Mo_4O_4 \cdot 2H_2O$ (10 000x)	25	Sigma-Aldrich	M1651-100G
H_3BO_3 (10 000x)	30	Sigma-Aldrich	B6768-500G
EDTA $\cdot 2H_2O$ (10 000x)	84	Sigma-Aldrich	E6635-100G
$MgSO_4 \cdot H_2O$ (400x)	246.47	Sigma-Aldrich	M5921-500G

*Provided by Vectron BioSolutions AS

2.2.1.2 Glucose (40%) stock solution

Glucose (40%) stock solution was prepared by dissolving D(+)-Glucose anhydrous (VWR, 101176K-2.5KG) in MQ-H_2O to a concentration of 400 g/L. The solution was autoclaved (121°C, 20 min) and stored at RT.

2.2.1.3 Ampicillin stock solution (1000x)

Ampicillin was the selectable marker used during the cultivations of recombinant strains. Ampicillin stock solution (1000x, 100 mg/mL) was prepared by dissolving ampicillin sodium salt (BioChemica, A0839,0100) in MQ-H_2O to a concentration of 100 mg/mL. The solution was filter-sterilized through a syringe filter (0.2 µm PTFE membrane, 13 mm, VWR, 514-4122), aliquoted to sterile Eppendorf tubes in volumes of 0.5 and 1 mL, and stored at -20°C.

2.2.1.4 *m*-Toluate 0.5 M stock solution (500x)

m-Toluate was used to induce the pVB expression vectors (Table 2.1) during experiments. The *m*-toluate stock solution (500x, 0.5 M) was prepared by dissolving *m*-toluic acid (Sigma-Aldrich, T3,660-9) in ethanol (absolute, VWR, 20821.310) to a concentration of 68.075 mg/mL. The solution was filter-sterilized through a syringe filter (0.2 µm PTFE membrane, 13 mm, VWR, 514-4122), aliquoted to sterile Eppendorf tubes in volumes of 0.5 mL, and stored at -20°C.

2.2.2 Media

2.2.2.1 LB medium and agar plates

LB medium

Lysogeny broth medium (LB medium) was used in the primary precultures for all experiments. LB-medium was made by dissolving the components in MQ-H_2O according to Table 2.3. The medium was autoclaved (121 °C, 20 min) and stored at RT.

Table 2.3 The composition of LB medium, and the components' concentration [g/L], supplier, and product number. The components were dissolved in MQ-H_2O, autoclaved (121 °C, 20 min), and stored at RT.

Component	Concentrations [g/L]	Supplier	Product number
Tryptone	10	Sigma-Aldrich	T9410-250G
Yeast extract	5	Sigma-Aldrich	92144-500G-F
NaCl	5	VWR Chemicals	27810.295

LB agar plates

LB agar plates were made by dissolving the components in MQ-H_2O according to in Table 2.4. The LB agar medium was autoclaved (121 °C, 20 min), cooled to 55°C, and aliquated into Petri dishes with 25 mL per plate. The LB agar plates were left to solidify at RT, then stored at 4°C until use.

Table 2.4 The composition of LB agar plates and the components' concentration [g/L], supplier, and product number. The components were dissolved in MQ-H_2O and autoclaved (121 °C, 20 min).

Component	Concentrations [g/L]	Supplier	Product number
Tryptone	10	Sigma-Aldrich	T9410-250G
Yeast extract	5	Sigma-Aldrich	92144-500G-F
NaCl	5	VWR Chemicals	27810.295
Bacteriological agar	15	OXOID	LP0011

2.2.2.2 Hi medium

Hi medium is a defined mineral medium that was used in the cultures for microscopy experiments, and secondary precultures for bioreactor cultivations. The medium was prepared in two steps to avert precipitation. A Hi 2x basis medium was made by dissolving the components in MQ-H_2O according to Table 2.5. The basis medium was autoclaved (121 °C, 20 min) and stored at RT. The final Hi medium was made according to Table 2.6, by combining Hi 2x basis medium (Table 2.5), trace element stock solutions (Table 2.2), glucose (40%) stock solution (Subsection 2.2.1.2), and MQ-H_2O using sterile technique. Ampicillin stock solution (1000x, 100 mg/mL) (Subsection 2.2.1.3) was added to a 0.1 mg/mL concentration to cultures with recombinant strains.

Table 2.5 The composition of 2x Hi Basis medium, and components' concentration, supplier and product number. The components were dissolved in MQ-H_2O, autoclaved (121 °C, 20 min) and stored at RT.

Compound	Concentrations [g/L]	Supplier	Product number
$Na_2HPO_4 \cdot 2H_2O$	17.20	Honeywell Fluka	30435-1KG
KH_2PO_4	6.00	Sigma-Aldrich	P5655-100G
NH_4Cl	2.00	Sigma-Aldrich	A9434-500G
NaCl	1.00	VWR Chemicals	27810.295

Table 2.6 The composition of Hi medium, in addition to the amount of stock solutions added [mL/L] and the final concentration of the components [g/L and mM].

Components	Amount added [mL/L]	Concentration [g/L]	Concentration mM
Hi 2x basis medium	500	-	-
$Na_2HPO_4 \cdot 2H_2O$	-	8.60	48.317
KH_2PO_4	-	3.00	22.717
NH_4Cl	-	1.00	18.695
NaCl	-	0.50	8.556
$Zn(CH_3COO)_2 \cdot 2H_2O$ (500x)	2	0.008	0.036
Fe(III) citrate hydrate stock solution (100x)	10	0.060	0.245
$MnCl_2 \cdot 4H_2O$ stock solution (666.67x)	1.5	0.015	0.076
$CuCl_2 \cdot 2H_2O$ stock solution (10 000x)	0.1	0.0015	0.009
$CoCl_2 \cdot 6H_2O$ stock solution (10 000x)	0.1	0.0025	0.011
$Na_2Mo_4O_4 \cdot 2H_2O$ stock solution (10 000x)	0.1	0.0025	0.010
H_3BO_3 stock solution (10 000x)	0.1	0.003	0.049
EDTA $\cdot 2H_2O$ stock solution (10 000x)	0.1	0.0084	0.023
$MgSO_4 \cdot H_2O$ stock solution (400x)	2.5	0.62	2.434
Glucose (40%) stock solution	50	20.00	55.506
MQ-H_2O	433.5	-	-

2.2.2.3 Hf.1 medium

Hf.1 medium is a defined mineral medium used for bioreactor cultivations. The medium was prepared in two steps to avert precipitation. Hf.1 basis medium was made in the bioreactor vessel by dissolving the components in MQ-H_2O according to Table 2.7, and autoclaved (121 °C, 20 min). The final Hf.1 medium was prepared according to Table 2.8 by adding trace element stock solutions (Table 2.2), glucose (40%) stock solution (Subsection 2.2.1.2), and MQ-H_2O to the Hf.1 basis medium using sterile technique. Ampicillin stock solution (1000x, 100 mg/mL) (Subsection 2.2.1.3) was added to a 0.1 mg/mL concentration to cultures with recombinant strains.

Table 2.7 The composition of the H1.f basis medium, and the component's concentration, supplier and product number. The components were mixed in MQ-H_2O and autoclaved (121 °C, 20 min).

Compound	**Concentrations [g/700 mL]**	**Supplier**	**Product number**
KH_2PO_4	14.94	Sigma-Aldrich	P5655-100G
$(NH_4)_2HPO_4$	3.6	Merck	101207.0500
Citric acid monosodium salt, 99%	1.89	Acros organics	47690010

Table 2.8 The composition of Hf.1 medium, in addition to the amount of component added [mL/900 mL] and the concentration of the compounds [g/900 mL and mM].

Component	**Amount added [mL/900 mL]**	**Concentration [g/900mL]**	**Concentration [mM]**
Hf.1 basis medium	700	-	-
KH_2PO_4	-	14.940	121.978
$(NH_4)_2HPO_4$	-	3.600	30.289
Citric acid monosodium salt, 99%	-	1.890	9.284
$Zn(CH_33COO)_2 \cdot 2H_2O$ (500x)	2.250	0.009	0.046
Fe(III) citrate hydrate stock solution (100x)	11.250	0.0675	0.306
$MnCl_2 \cdot 4H_2O$ stock solution (666.67x)	1.692	0.01692	0.095
$CuCl_2 \cdot 2H_2O$ stock solution (10 000x)	0.114	0.00171	0.011
$CoCl_2 \cdot 6H_2O$ stock solution (10 000x)	0.112	0.00279	0.013
$Na_2Mo_4O_4 \cdot 2H_2O$ stock solution (10 000x)	0.112	0.00279	0.013
H_3BO_3 stock solution (10 000x)	0.114	0.00342	0.061
$EDTA \cdot 2H_2O$ stock solution (10 000x)	0.113	0.00945	0.028
$MgSO_4 \cdot H_2O$ stock solution (400x)	5.473	1.35	6.086
Glucose (40%) stock solution	45	22.5	138.766
MQ-H_2O	133.771	-	-

2.2.2.4 SOC medium

Super Optimal Catabolite-repression (SOC) basis medium was made by dissolving the components in MQ-H_2O according to Table 2.9. A 20% glucose solution was made by dissolving D(+)-Glucose anhydrous (VWR, 101176K 2.5KG) to a concentration of 200 g/L. The SOC basis medium and 20% glucose solution were autoclaved (121°C, 20 min) and cooled down to 60°C. Then 2 mL 20% glucose solution was mixed into 100 mL of the SOC basis medium, and

the SOC medium was aliquoted in volumes of 10 mL into sterile 15 mL centrifuge tubes and stored at -20°C.

Table 2.9 The composition of SOC basis medium and the component's concentration [g/100 mL], supplier, and product number. The components were dissolved in MQ-H_2O, autoclaved (121 °C, 20 min), and stored at -20°C.

Component	Concentrations [g/L]	Supplier	Product number
Tryptone	10	Sigma-Aldrich	T9410-250G
Yeast extract	5	Sigma-Aldrich	92144-500G-F
NaCl	5	VWR Chemicals	27810.295
KCl	0.19	Merck	31248
$MgSO_4$·7 H_2O	4.92	VWR	25167

2.3 Batch cultivation in bench-scale bioreactors

2.3.1 Bioreactor set up

The batch cultivations were performed with the Applikon 1000 mL autoclavable bioreactor system, controlled with Applikon my-Control system. Hf.1 basis medium (Table 2.7) was made in the bioreactor vessel. The pH electrode was calibrated using commercial buffers with pH 4.01 and 7.00 (Mettler Toledi, 5130004 and 51350006, respectively). The dissolved oxygen (DO) electrode was calibrated to 0% oxygen using nitrogen gas, and the calibrated electrodes were mounted in the bioreactor. Filters (0.2 µm PTFE membrane, 50 mm, Pall, 4250) were attached the inlet and outlet air ports, and the bioreactor system was autoclaved (121 °C, 20 min). The rest of the Hf.1 medium was added to the autoclaved bioreactor as detailed in Subsection 2.2.2.3. The bioreactor system was assembled as described in the instruction manual, and the bioreactor was primed to pH 7 and the temperature adjusted to 30 °C.

Precultures were prepared in two stages. Primary precultures were made by inoculating 50 mL LB medium (Table 2.3) with 125 µL glycerol stock in 250 mL baffled shake flasks. Ampicillin stock solution (1000x, 100 mg/mL) was added to a 0.1 mg/mL concentration to precultures with recombinant strains, and the precultures were incubated (200 rpm, 30°C, 10±2 h). The secondary precultures were made by inoculating 100 mL Hi medium with 150 µL primary preculture in 500 mL baffled shake flasks. Ampicillin stock solution (1000x, 100 mg/mL) was added to a 0.1 mg/mL concentration to cultures with recombinant strains, and the precultures were incubated (200 rpm, 30 °C, 12±2 h).

The reactors were inoculated to a starting OD_{600} of 0.1 with secondary preculture in the late exponential growth phase (OD_{600} ~ 3-4). Ampicillin stock solution (1000x, 100 mg/mL) was added to a 0.1 mg/mL concentration to the recombinant strains. The DO was immediately calibrated to 100%. Once biomass concentration reached the wanted OD_{600}, inducer *m*-Toluate stock solution (500x, 0.5 M) was added to a concentration of 1 mM, and the temperature was set to 25 °C.

2.3.2 Cultivation conditions

The DO, pH and temperature were monitored with the BioXpert V2 software throughout the cultivations. The air flow was set to 33 L/h, and the DO was continuously adjusted to 40% with an agitation cascade (200-800 rpm). The pH was set to 7 and continuously adjusted with 4 M NH_4OH (Sigma Aldrich, 30501-1L-M). The temperature was set to 30 °C at the start of the cultivation, and lowered to 25 °C after the inducer *m*-toluate was added. A mass spectrometer (Prima BT process mass spectrometer, Thermo Fisher Scientific) with the GasWorks software was used to analyze the off-gas throughout the cultivations. A volume of 100 µL 10% diluted antifoam (Sigma-Aldrich, 85390-500ML) was added when excess foaming was observed.

2.4 Monitoring and calibration of bacterial growth and mCherry production

2.4.1 Monitoring bacterial growth and mCherry production

Bacterial growth was monitored by measuring the OD_{600} with a spectrophotometer (VWR, Vis V-1200). Samples with OD_{600} surpassing the linear range of the spectrophotometer were diluted with MQ-H_2O to be within the linear range. The OD_{600} measurements were blanked with the medium the cells were grown in, diluted with the same ratio as the sample.

The mCherry production was monitored by measuring the fluorescence using a microplate reader (Tecan Spark®. 20M, Bergman Diagnostika) with bottom reading at gain 66. The fluorescence was measured in technical triplicates by pipetting volumes of 200 µL sample into their allocated wells in a transparent 96-well flat bottom plate (VWR, 734-2327). The cells' autofluorescence were corrected by subtracting the fluorescence of mCherry producing strains with the fluorescence of the corresponding control or WT strain at similar biomass concentration.

2.4.2 Biomass calibration: Correlation between OD_{600} and cell dry weight

The calibration between OD_{600} and cell dry weight (DCW) was performed on the strains *E. coli* BL21 WT, Z2 and Z2-mCh. Precultures were made by inoculating 10 mL LB medium (Table 2.3) with 100 µL glycerol stocks in 125 mL baffled shake flasks. Ampicillin stock solution (1000x, 100 mg/mL) was added to a concentration of 0.1 mg/mL to the recombinant strains, and the precultures were incubated (200 rpm, 30°C, 12 ±2 h). The cultures were made in biological quadruplicates by inoculating 100 mL Hi medium (Table 2.6) with 125 µL preculture in four 500 mL baffled shake flasks. Ampicillin stock solution (1000x, 100 mg/mL) was added to a concentration of 0.1 mg/mL to the recombinant strains, and the cultures were incubated overnight (30 °, 200 rpm, 14 ± 2h).

The overnight cultures of the same strain were divided into eight sterile 50 mL centrifuge tubes and centrifuged (4000 rcf, RT, 10 min), after they had reached late exponential growth phase (OD_{600} ~ 3-4). The supernatants were discarded, and the pellets were resuspended in Hi medium (Table 2.6). The cell suspensions with the same strain were combined and diluted to OD_{600} 20, which was used to make a dilution series with six concentrations ranging from OD_{600} 20 to 0.5.

The OD_{600} of the dilution series was measured in technical duplicates using a spectrophotometer (VWR, Vis V-1200). DCW samples were taken in technical triplicates simultaneously as the OD_{600} was measured. The DCW was sampled with a fast filtration protocol adapted from (Kvitvang and Bruheim, 2015). The filtration setup was composed of a filtration manifold (Pall Corporation®Vacuum manifold), vacuum pump (Vacuubrand®Diaphragm pump ME 4R NT), vacuum controller (Vacuubrand®CVC 3000), and a waste collector. Filters (Durapore®, 0.45 µm PVDF membrane, HVLP04700) were dried over two days and weighted. These filters were positioned on the filtration manifold and moistened with mQ-H_2O. Samples with biomass corresponding to 5 mL at OD_{600} 1 were taken from the bioreactor and transferred to the filter. The samples were exposed to 200 mbar pressure and washed with 10 mL RT MQ-H_2O. The filtration procedure was also performed on Hi medium (Table 2.6) as a blank. The filters were dried for approximately 48 h at 120°C and weighted. The correlation between OD_{600} and DCW was found by performing a linear regression of the average DCW as a function of the average OD_{600}. The blank (Hi medium) was subtracted from the samples' DCW. The results from the biomass calibrations are given in Appendix D.1.

2.4.3 mCherry calibration: Correlation between fluorescence and mCherry concentration

The mCherry calibration was performed in two stages. First, a calibration curve for mCherry concentration as a function of fluorescence was prepared as described in Subsection 2.4.3.1. Second, a calibration curve for fluorescence of lysed cell suspension as a function of fluorescence of cell suspensions with whole cells was prepared, as described in Subsection 2.4.3.2. The second calibration curve was made as it was uncertain if the presence of whole cells affected the fluorescence measurements. The mCherry is located intracellularly, thus the fluorescence excitation might be partly absorbed by the cell membrane and cell wall.

2.4.3.1 Preparation of mCherry calibration curve

The mCherry calibration was performed by measuring the fluorescence of two independent mCherry dilution series. Commercial mCherry stock (100 µg, Antibodies, ABIN412973) was used to prepare two independent standard curves with nine concentrations of mCherry ranging from 100 to 0.39 ng/µL, diluted with Hi medium (Table 2.6). The standards were pipetted into their allocated wells in a transparent 96-well flat bottom plate (VWR, 734-2327) in volumes of 200 µL. The wells between the standards were filled with Hi medium (Table 2.6)

to prevent cross-talk in the fluorescence measurements. The fluorescence was measured as described in Subsection 2.4.1. The correlation between mCherry concentration and fluorescence was found by performing a linear regression of mCherry concentration as a function of fluorescence. The standards outside of the linear range between mCherry concentration and fluorescence were removed prior to the linear regression analysis. The results are given in Appendix D.2.

2.4.3.2 Comparison of fluorescence between cell suspensions with whole and lysed cells

The following experiment was conducted on *E. coli* BL21 Z1-mCh to compare fluorescence between cell suspensions with lysed and whole cells. A preculture was made by inoculating 50 mL LB medium (Table 2.3) containing ampicillin (0.1 mg/mL) with 100 µL glycerol stock, in a 250 mL baffled shake flask, followed by incubation (30°C, 200 rpm, 13 h). The culture was made by inoculating 100 mL Hi medium (Table 2.6) containing ampicillin (0.1 mg/mL) with 125 µL preculture, in a 500 mL baffled shake flask, followed by incubation (30°C, 200 rpm, 13 h). When the culture reached OD_{600} 1, mCherry expression was induced by addition of *m*-tolutae to a 1 mM concentration, and the culture was incubated (25 °C, 200 rpm).

Once the culture reached OD_{600} 2.5, it was transferred to two sterile 50 mL centrifuge tubes and centrifuged (4000 rcf, 25°C, 5 min). The supernatants were discarded, and the pellets were resuspended in Hi medium (Table 2.6). The cell suspension was used to make a dilution series with seven dilutions ranging from undiluted to a 1:64 dilution, using Hi medium (Table 2.6) to dilute. The dilution series was mixed by vortex and divided between two previously weighed sterile Eppendorf tubes. One set of dilution series was used to measure the fluorescence of whole cells, and the other for lysed cells, prepared as described below.

The fluorescence of the whole-cell dilution series was measured as described in Subsection 2.4.1. The other dilution series was lysed with Bacterial Protein Extraction Reagent (B-PER) (Thermo Fisher Scientific, 78243) as described in the instruction manual. The dilution series was centrifuged (15000 rcf, 5 min, RT) and the supernatants were discarded. B-PER was added to a concentration of 4 µL/mg wet cell pellet, and thoroughly mixed with a pipette. The dilution series was incubated (15 min, RT) with mixing every 5 min by pipetting. The cell suspensions were observed with a microscope to confirm that the cell lysis was successful.

The lysed cell suspensions were diluted with Hi medium (Table 2.6) to the same biomass concentration as the whole cell dilution series. The lysed dilution series was thoroughly mixed by vortex, centrifuged (15000 rcf, 5 min, RT), and the supernatant was collected. The pellets were resuspended in Hi-medium (Table 2.6) to the same volume as described above. The fluorescence of the supernatants and resuspended pellets were measured as described in Subsection 2.4.1. These fluorescence measurements were added together to represent the total fluorescence of the lysed cell suspensions. The correlation in fluorescence between lysed and whole cells was found by performing a linear regression with the fluorescence of the lysed cells as a function of the fluorescence of whole cells. The results are given in Appendix D.2.

2.5 Quantification of extracellular substrates and products

Extracellular substrates and products were quantified to study how their concentrations changed throughout the bioreactor cultivations, and to compare their concentrations between the different strains. The samples were prepared by harvesting a known volume of culture broth, which was divided into technical triplicates and centrifuged (8000 rcf, 5 min, 4°C). The supernatants were collected, snap freezed in liquid nitrogen, and stored at -80°C until further analysis. The extracellular concentration of glucose and some organic acids were quantified by high-performance liquid chromatography (HPLC), as described in Subsection 3.1.4, and the extracellular protein concentration was quantified by Bradford protein assay, as described in Subsection 2.5.2.

2.5.1 Quantification of extracellular glucose and organic acids by HPLC

The method was used to quantify glucose, acetic acid, citric acid, lactic acid, and succinic acid by HPLC. An external standard-mix (ESTD-mix) was prepared by mixing analytical grade standards in MQ-H_2O. The ESTD-mix was filtered through a syringe filter (0.2 µm PTFE membrane, 13 mm, VWR, 514-0068), then used to make a dilution series composed of five dilutions. The range of the ESTD series for each compound is listed in Table 2.10, together with the supplier and product number of the standards.

Table 2.10 The compounds used as ESTD to quantify the extracellular concentration of glucose, acetic acid, citric acid, lactic acid, and succinic acid by HPLC, together with their concentration range in the standard curve, supplier, and product number.

Compound	ESTD concentration range [g/L]	Supplier	Product number
Glucose	0.63 - 5.00	VWR	101176K-2.5KG
Acetic acid	0.06 - 0.50	Supelco	1.00063.1011
Citric acid	0.06 - 0.50	Sigma-Aldrich	C7129-100G
Lactic acid	0.25 - 2.00	Sigma-Aldrich	L1500-100G
Succinic acid	0.06 - 0.50	Sigma-Aldrich	S3674-100G

Supernatant samples described in Subsection 2.5 were filtered through syringe filters (0.2 µm PTFE membrane, 13 mm, VWR, 514-0068). The samples were analyzed both diluted and undiluted to ensure that the compounds concentrations were not above the range of their standard curves. MQ-H_2O was used as blank. The samples were analyzed using an Alliance HPLC (Waters) with a Hi-Plex column of the dimensions 300 x 7.7 mm (Agilent Technologies), detected with a refractive index (RI) and a UV/Vis detector. The compounds were eluted by a mobile phase composed of 0.05 M H_2SO_4 in MQ-H_2O with a 0.6 mL/min flow rate. The column oven and RI detector temperatures were set to 45 and 35°C, respectively. All data are given as the average of two technical replicas.

2.5.2 Quantification of extracellular proteins by Bradford protein assay

A protein assay kit (BioRad, 500-0001) was used to analyze the supernatant samples described in Subsection 2.5. The protein assay can be used to quantify proteins by colorimetry according to the Bradford method (Bradford, 1976). The kit contains a dye reagent and bovine serum albumin (BSA) to be used as standard. BSA was used to prepare two independent standard dilution series with ten concentrations in the range of 0 to 10 µg/µL BSA. The supernatant samples were analyzed in four different dilutions to ensure the optical densities were within the linear range with the protein concentrations. The samples and standards were transferred to Eppendorf tubes in volumes of 800 µL, mixed with 200 µL dye reagent by vortex, and incubated for at least 5 min in the dark at RT. The mixtures were transferred to cuvettes, and the optical density was measured at 595 nm with a spectrophotometer (VWR, Vis V-1200). The protein concentration standard curve was found by performing a linear regression analysis on the BSA standards, with BSA concentration as a function of the OD_{595}. The linear regression was performed with the python script given in Appendix D.3. The linear regression resulted in the function y=21.241x - 0.026, with a R^2 of 0.991. The function was used to find the concentration of extracellular protein in the supernatant samples, with corrections for dilutions in the OD_{595} measurements.

2.6 Endometabolome Analysis

Endometabolome analysis was performed on the strains *E. coli* BL21 WT, A2, A2-mCh, Z2, and Z2-mCh. The analysis was carried out to study how the central carbon metabolism of the different strains changed throughout the cultivation, and to compare these changes between the strains. The recombinant strains were analyzed with one biological replica, while the WT strain was studied in biological triplicates to examine the repeatably between cultivations. The strains were cultivated in bench-scale bioreactors as described in Section 2.3. Samples were taken at different cell densities and processed as described in Subsection 2.6.1. Three additional replicas were sampled for each strain at each time point, as described in Subsection 2.4.2, to determine the DCW at the sampling points. Reverse phase liquid chromatography-tandem mass spectrometry (RP LC-MS/MS) was used to quantify amino acids, as described in Subsection 2.6.2, and organic acids, as described in Subsection 2.6.3. The resulting endometabolome data were analyzed as described in Subsection 2.6.5. Phosphorylated metabolites and TCA cycle intermediates were supposed to be quantified by capillary ion chromatography-tandem mass spectrometry (capIC-MS/MS) as described in Subsection 2.6.4. However, this could not be done as the equipment was under maintenance.

2.6.1 Endometabolome sampling and sample processing

Metabolites are unstable and susceptible to degradation by both internal and external factors (Gil *et al.*, 2015). Therefore, it is crucial that the sampling procedure is fast to prevent turnover of metabolites. The sampling procedure selected was therefore adapted from (Kvitvang and Bruheim, 2015), which utilizes fast filtration and immediate quenching of the metabolism. The samples were processed by extracting the metabolites from the cells, then lyophilization by freeze-drying, resulting in a dry metabolite extract. Upon analysis, the metabolite extracts were reconstituted in cold MQ-H_2O and spin filtered to remove impurities.

2.6.1.1 Fast filtration and quenching of the metabolism

Samples were taken in at least 4 technical replicas at three different cell densities in the cultivation, at OD_{600} ~ 2.5 (T1) right before induction, OD_{600} ~ 7 (T2) and OD_{600} ~ 15 (T3), for all strains except Z2-mCh. It was previously observed that the biomass concentration of Z2-mCh cultures had a decreasing trend after induction of mCherry expression. Therefore, Z2-mCh was sampled at OD_{600} ~ 2.5 (T1) as the other strains, 30 min after induction (T2), and 4.5 h after induction (T3), which was the same time frame as to when T2 was sampled for the corresponding control strain Z2.

The fast filtration setup consisted of a filtration manifold (Pall Corporation®Vacuum manifold), vacuum pump (Vacuubrand®Diaphragm pump ME 4R NT), vacuum controller (Vacuubrand®CVC 3000), and a waste collector. A filter (Durapore®, 0.45 µm PVDF membrane, HVLP04700) was placed on the filtration manifold and moistened with room tempered MQ-H_2O. Samples with biomass corresponding to 5 mL at OD_{600} 1 were taken from the bioreactor and transferred to the filter. The volume was adapted to ensure similar filtration times between the different sampling points. The samples were exposed to 200 mbar pressure, and after around 80 % of the growth medium had been filtered away, the remaining growth medium was washed with 10 mL MQ-H_2O at 30°C. When the filter was sufficiently dry, but not overdried, it was quickly transferred to a 50 mL tube containing 10 mL quenching solution (MeOH (VWR, 1.06035.2500) : ACN(VWR, 83640.320) : MQ-H_2O, 2:3:5, -20°C), and the tube was immediately placed in liquid nitrogen for complete quenching of the metabolism.

2.6.1.2 Preparation of metabolite extracts

The tubes containing the endometabolome samples were thawed in an ethanol bath (-20°C, 30 min). The tubes were left in the ethanol bath for an additional 30 min with three rounds of vigorous mixing by vortex to extract the metabolites from the cells. The filters were removed, and the tubes were centrifuged (4500 rcf, 5 min, -9°C) to clear out cell debris. A volume of 9 mL supernatant was transferred to new 50 mL centrifuge tubes with a hole in the lid, which was immediately transferred to a liquid nitrogen bath to snap-freeze. The metabolites were concentrated by freeze-drying (VirTis BenchTop Pro with Omnitronics™) for approximately 48 h, after which they were stored standing in racks at -80°C. Prior to analysis, the dried metabo-

lite extracts were redissolved in 500 µL cold MQ-H_2O and centrifuged (4500 rcf, 5 min, 0°C). The metabolite extracts were transferred to 10 kD spin filter tubes (VWR, 516 0228) and centrifuged (10 000 rcf, 10 min, 0°C).

2.6.2 Quantification of amino acids by RP LC-MS/MS

Amino acids (AA) were quantified as described in (Røst *et al.*, 2020). The method quantifies proteinogenic amino acids by RP LC-MS/MS with electron spray ionization in positive mode. The metabolite extracts were derivatized with phenyl isothiocyanate (PITC) as described in Subsection 2.6.2.1 to increase the amino acids' affinity with the column. The RP LC-MS/MS analysis was performed using an ACQUITY I-Class UPLC coupled to a Xevo TQ-XS triple quadrupole mass spectrometer (Waters). The UPLC was equipped with a Waters Aquity BEH C18 2.1×100 mm column with a pore size of 1.7µm (186002352, Waters). The RP LC-MS/MS analysis was performed following the procedure described in (Røst *et al.*, 2020).

2.6.2.1 Amino acid derivatization procedure

The derivatization was performed in 96-well plates (Waters, 186002643, 96-well sample collection plate, 350 µL). A volume of 10 µL internal standard-mix (ISTD) (Cambridge isotope laboratories, MSK-A2-1.2) was pipetted to all wells, excluding the blank wells. An ESTD-mix (Sigma-Aldrich, AAS18) was used to make an ESTD dilution series with eight concentrations ranging from 300 to 0.05 µM, diluted in MQ-H_2O. The ESTDs were pipetted into their assigned well in volumes of 150 µL, and the 25 µM ESTD was pipetted into every 20^{th} well as quality control (QC). The metabolite extracts (Subsection 2.6.1.2) were pipetted into their assigned wells in volumes of 150 µL. The plates were transferred to a Speedvac™(4 Torr, 60°C, ccc, 3h) to concentrate the extracts. Once dry, the plates were stored overnight at -80°C.

Derivatization solution (EtOH:pyridine:MQ-H_2O, 1:1:1 v/v) was prepared by mixing 96% EtOH (VWR, 20824.365), pyridine (Sigma-Aldrich, 270970-250ML), and MQ-H_2O. PITC (Sigma-Aldrich, 78780-25ML) was added to the concentration 5% (v/v), and the mixture was vortex until transparent. The solution was immediately added to all wells in volumes of 50 µL, except the blank wells, and the plates were shaken (500 rpm, 20 min, RT) (Biosan, PST-60HL-4). During this step, PITC reacts with amine groups (-NH_2) in alkaline conditions, which results in the formation of phenylthiocarbamyl derivatives (PTC-amino acids) (Røst *et al.*, 2020). The PTC-amino acids were concentrated in a Speedvac™(4 Torr, 45°C, ccc, 1.5 h), then stored overnight at -80°C, when dry.

Extraction solvent (MeOH:MQ-H_2O, 1:1 v/v, with 5 mM ammonium acetate) was prepared by mixing methanol (MeOH) (VWR, 1.06035.2500), MQ-H_2O and ammonium acetate (Sigma-Aldrich, 73594). Extraction solvent was added to all the wells except the blank wells, and the plates were shaken (500 rpm, 30 min, RT) (Biosan, PST-60HL-4). The plates were then centrifuged (400 rcf, 5 min, RT), and 100 µL supernatant was transferred from each well to

their assigned wells in new plates (Waters, 186005837, 96-well sample collection plate, 700 µL round well). The blank well was filled with 500 µL extraction solvent.

2.6.3 Quantification of organic acids by RP LC-MS/MS

Organic acids (OA) were quantified using a method adapted from (Tan *et al.*, 2014), with modifications described in (Røst *et al.*, 2020). The method quantifies low molecular weight (LMW) OA by RP LC-MS/MS with electron spray ionization in positive mode. The metabolite extracts were derivatized with *ortho*-benzylhydroxylamine (*o*-BHA) and N-(3-Dimethylaminopropyl)-N'-ethylcarbodimide hydrochloride (EDAC), as described in Subsection 2.6.3.1. The derivatization procedure increases the LMW OAs' affinity with the column, as *o*-BHA and EDAC cause an oxime formation in molecules containing a keto group (Røst *et al.*, 2020). The RP LC-MS/MS analysis was performed using the UPLC-TQ-XS setup described in Subsection 2.6.2, according to the procedure described in (Røst *et al.*, 2020).

2.6.3.1 Organic acid derivatization procedure

The derivatization was performed in 96-well plates (Waters, 186005837, 96-well sample collection plate, 700 µL round well). ISTD-mix was prepared by mixing equal volumes of ^{13}C ISTD-mix (Cambridge isotope laboratories, MSK-OA-1) and ^{13}C Lac+Pyr ISTD-solution (Cambridge isotope laboratories, 2440-0.5 and 1579-0.5, respectively). The finished ISTD-mix was pipetted into all wells in volumes of 20 µL, excluding the blank wells. Two ESTD-mixes were used to prepare two independent ESTD dilution series, one for all LMW OAs and one for citrate (Cit), lactate (Lac) and pyruvate (Pyr) with higher concentrations. The ESTD-mix for all LMW OA (Analytical grade standards from Sigma-Aldrich) was used to make a dilution series with twelve concentrations ranging from 60 to 0.01 µM. The high concentration ESTD-mix of Cit, Lac and Pyr (Analytical grade standards from Sigma-Aldrich) was used to make a dilution series with eight concentrations ranging from 1000 to 7.81 µM. The ESTDs were pipetted to their assigned wells in volumes of 80 µL, and the 5 µM ESTD was pipetted to every 20^{th} well as QC.

Pyridine buffer solution was made by mixing 1.35 mL HCl (12 M, Sigma-Aldrich, 30721-2.5L-M), 2.2 mL pyridine (Sigma-Aldrich, 270970-250ML), and 22 mL MQ-H_2O. The *o*-BHA and EDAC derivatization solutions were made by dissolving *o*-BHA (Sigma-Aldrich, B22984-25G) and EDAC (Sigma-Aldrich, 03450-25G) in pyridine buffer to a concentration of 1 M *o*-BHA (160 mg/mL) and EDAC (196 mg/mL). The derivatization solutions were mixed thoroughly and kept at 37 °C until use, to prevent precipitation. Both derivatization solutions were added to all wells in volumes of 50 µL, excluding the blank wells. The plates were shaken (500 rpm, 1 h, RT), and ethyl acetate (VWR, 83621.320) was added to all wells in volumes of 300 µL, except the blank wells. The plates were shaken (500 rpm, 1 h, RT) (Biosan, PST-60HL-4), after which, all organic compounds in the mixture had risen to the top of the well. Therefore, 100 µL of the top layers in the wells were transferred to their assigned wells in new 96-well

plates (Waters, 186005837, 96-well sample collection plate, 700 µL round well). The plates were placed in a Speedvac™(4 torr, 60 °C, ccc, ~1 h) to concentrate the LMW OA derivatives. The plates were stored at -80°C, once dry. Upon analysis, 50 µL methanol solution (MeOH (VWR, 1.06035.2500) : MQ-H_2O, 1:1 v/v) was added to all wells, excluding the blank. The plates were shaken (500 rpm, 5 min, RT) (Biosan, PST-60HL-4), and 500 µL of the methanol solution was added to the blank well.

2.6.4 Quantification of phosphorylated metabolites and TCA cycle intermediates by CapIC-MS/MS

Phosphorylated metabolites and TCA cycle intermediates were supposed to be quantified using a method adapted from (Kvitvang *et al.*, 2014), with modifications described in (Stafsnes *et al.*, 2018). An ISTD and ESTD-mix were prepared as described in (Stafsnes *et al.*, 2018). The ESTD-mix was used to make an ESTD dilution series with eight concentrations ranging from 10 000 to 10 nM. The metabolite extracts and ESTDs were added to HPLC vial inserts in volumes of 80 µL, together with 20 µL ISTD-mix, and the inserts were placed into HPLC vials with pre-slit lids. Blank vials were filled with 100 µL MQ-H_2O. The samples and ESTDs were prepared, but could not be analyzed due to equipment malfunction. The samples will be analyzed at a later date using a Xevo TQ-XS triple quadrupole mass spectrometer (Waters, Milford, MA, USA), following the procedure described in (Stafsnes *et al.*, 2018).

2.6.5 Processing of endometabolome data

The endometabolome data was processed as described in (Røst *et al.*, 2020), using the TargetLynx™ application manager in MassLynx 4.1 software (Waters). The metabolites were automatically quantified by interpolation from the calibration curves, which were generated using least-squares regression with 1/x weighting. The response factors of the ESTD and processed samples were corrected using the response factor of the corresponding ISTD-isotopologue. If necessary, the quantified data was manually adjusted by modifying the area and baseline of the integrated peaks. The concentrations were corrected for dilutions performed during sample processing and converted to DCW basis (µmol/ g DCW) using the DCW values from the endometabolome sampling.

Missing values were estimated by the minimum measured value of the respective metabolite. Dixon's Q-test was used to identify and reject potential outliers in the sub-sample sets, as shown in Appendix E. The sample sets were normalized by sum. The python script given in Appendix F.2 was used to autoscale the data sets with mean and variance equal to 0 and 1, respectively, and to perform principle component analysis (PCA). The script generated 2D-scores plots. The metabolic data was used generate bar charts that visualize the overall concentrations of the metabolite classes in the different samples. Metabolic data normalized by sum was used to make heat maps that visualize the changes at individual metabolite levels.

2.7 Comparison of cellular length by microscopy

Microscopy was performed to compare the cell length between the strains *E. coli* BL21 WT, A2, A2-mCh, Z2, and Z2-mCh. The experiment was conducted to examine how the metabolic stress derived from recombinant protein production affects the cells' morphology. The strains were studied in biological duplicates from different glycerol stocks, by examining the length of 100 cells per replica, before and after addition of the inducer *m*-toluate.

2.7.1 Preparation of Agarose Pads

The method used to prepare the agarose pads was adapted from (Skinner *et al.*, 2013). Six microscope slides were cleaned with MQ-H_2O and ethanol (96%, VWR, 20824.365), then dried with Kimwipe. Five of the slides were stacked as shown in Figure 2a in (Skinner *et al.*, 2013). The agarose solution was prepared by mixing agarose (VWR, 0710-100G) in 1x PBS (Thermo Fisher, BR0014G) to a concentration of 1% (w/v) agarose. The mixture was heated in a microwave at low power in short increments (10-15 s) until the agarose was dissolved. The solution was poured onto the slide stack and covered with the remaining slide. A weight (~200 g) was placed on top of the slide stack, and the agarose solution was left to solidify for 20 min at RT. When solidified, the four slides on the sides of the stack were removed, and the excess agarose was cut off with a scalpel. The agarose pad was stored in a fridge between two slides in a plastic ziplock bag until use.

2.7.2 Microscopy protocol

Precultures were made by inoculating 10 mL LB medium (Table 2.3) with 50 µL glycerol stock in 125 mL baffled shake flasks, followed by incubation (30 °C, 200 rpm, 10±2 h). Ampicillin stock solution (1000x, 100 mg/mL) was added to a 0.1 mg/mL concentration to the recombinant strains. The cultures were made by inoculating 50 mL Hi medium (Table 2.6) with 80 µL preculture in 250 mL baffled shake flasks, followed by incubation overnight (30 °C, 200 rpm). Ampicillin stock solution (1000x, 100 mg/mL) was added to a 0.1 mg/mL concentration to the recombinant strains. When the cultures reached OD_{600} 1, 500 µL of the cultures were transferred to sterile 1.5 mL Eppendorf tubes. The remaining cultures were induced by the addition of *m*-Toluate stock solution (500x, 0.5 M) to a concentration of 1 mM and incubated (25 °C, 200 rpm).

Pieces (1 cm x 1 cm) of the agarose pad (Subsection 2.7.1) were cut out using a scalpel and transferred to clean microscopy slides. The uninduced cultures in the Eppendorf tubes were diluted to OD_{600} 0.5 with sterile 1x PBS. The diluted cell suspensions were mixed thoroughly by vortex and 5 µL was pipetted onto agarose pad pieces and covered with coverslips. A Zeiss Axio Imager Z2 microscope equipped with Axiocam MR R3 (ZEISS) was used to observe and take images of cells. The images were taken with fluorescence and phase-contrast imaging. Fluorescence images were captured using a channel customized for mCherry. The procedure

was repeated when the induced cell suspension reached OD_{600} 2.5, for all strains except Z2-mCh. The biomass concentration of Z2-mCh decreases after induction, hence Z2-Ch would not reach OD_{600} 2.5. Therefore, Z2-mCh was studied 8 hours after induction, which was the time frame the negative control strain Z2 used to reach OD_{600} 2.5 after induction. The length of the cells were measured using Zen 2.3 Pro software with calibrated scale bars. At least 100 cells were measured for every biological replica. The statistical tests performed on the microscopy data as described in Appendix G.

Chapter 3

Results

As a part of the research project MetaboProt, the work performed in this thesis aimed to study the metabolic burden expression vectors imposes on *E. coli*. The recombinant strains shown in Table 3.1 were selected for the experiments conducted in this thesis. These strains were selected based on the different combinations of expression cassette strength and PCN the expression vectors they were transformed with, as these combinations may give insight into how these factors contribute to the degree of metabolic burden the expression vector places on the host. The degree of metabolic burden is expected to be reflected by physiological and metabolic consequences observed during the cultivation of the strains. There are two variations of the expression cassette, wild type XylS/*Pm*/5'-UTR (A) and a mutated version (Z) with a higher expression cassette strength. The expression vectors have two variations in PCN 5 (1) and 20 (2) plasmids. For each combination of expression cassette version and PCN, there is a mCherry expressing strain (-mCh) and a negative control strain that lacks the mCherry encoding gene. More information on the expression vectors is given in Subsection 2.1.1 and the plasmid maps are shown in Appendix B.

Table 3.1 Strain and the expression cassette type, mCherry encoding gene, plasmid size [bp], and PCN of the expression vectors they were transformed with. The expression vectors were provided by Vectron BioSolutions AS.

Strain	Expression cassette	mCherry gene	Plasmid size [bp]	PCN
A2	Wild type	No	8172	20
Z1	Mutated	No	8249	5
Z2	Mutated	No	8249	20
A2-mCh	Wild type	Yes	8930	20
Z1-mCh	Mutated	Yes	8928	5
Z2-mCh	Mutated	Yes	8928	20

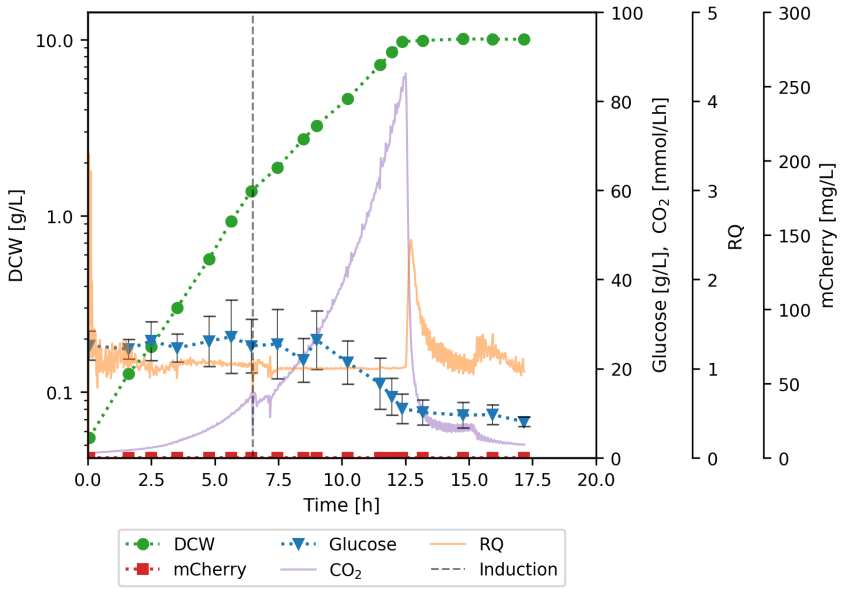
3.1 Bioreactor cultivations with quantification of extracellular substrates and products

Batch cultivations in bench-scale bioreactors were conducted to study how the different expression vectors affected the growth rate and biomass formation of the strains. Additionally, samples of the culture broth were taken throughout the cultivations. The extracellular concentration of substrates and products were quantified in these samples to examine and compare the substrate consumption and product secretion between the strains.

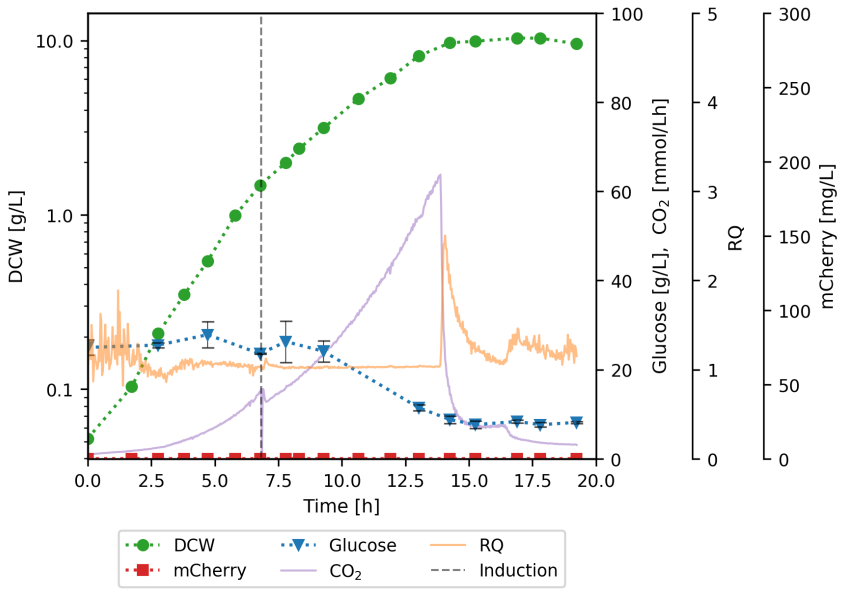
3.1.1 Growth characterisation of WT *E. coli* BL21 and recombinant strains

It was hypothesized that the recombinant strains would have different growth characteristics, as they harbor expression vectors with different combinations of plasmid size, PCN, and expression cassette strength. The growth curves (Figure 3.1) and determined growth rates (Table 3.2) show the strains reacted differently to induction and lowering of temperature.

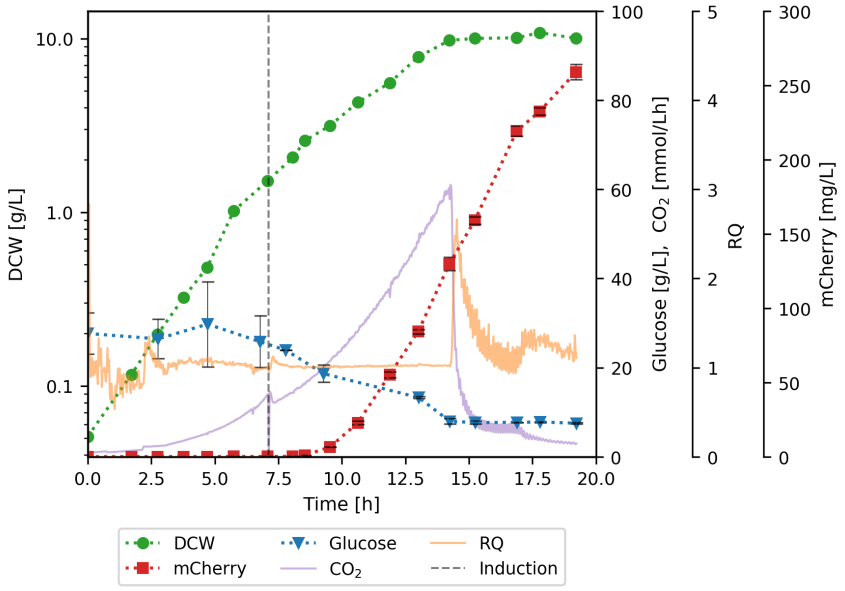
The WT strain had higher growth rates than the recombinant strains, especially after induction, where the growth rate had decreased by only 34%. The higher growth rate of WT is expected as it does not harbor expression vectors that require metabolic resources. The recombinant strains had similar growth rates before induction, but the growth rates differed after. The growth rate of Z1 decreased by only 38% after induction, which is comparable with WT, indicating a low degree of metabolic stress. A2's and A2-mCh's growth rates were approximately halved, while Z1-mCh and Z2 decreased around 60%. Induction had the most severe effect on Z2-mCh, as observed by the growth inhibition after induction. Z2-mCh reaches a maximum biomass formation at approximately 1.5 g DCW (OD_{600} ~ 3), around an hour after induction, after which the biomass concentrations had a decreasing trend with time. In contrast, the biomass concentration of the other strains continued to grow after induction at slower growth rates, reaching a maximum around 9 to 11 g CDW (OD_{600} ~ 20). In summery, the growth rates reflects the theoretical degree of metabolic burden described above. The strains carrying expression vectors with combinations of the stronger mutated expression cassette, higher PCN, or the mCherry encoding gene or both (Z2, Z1-mCh, and Z2-mCh) were more severely affected by induction.



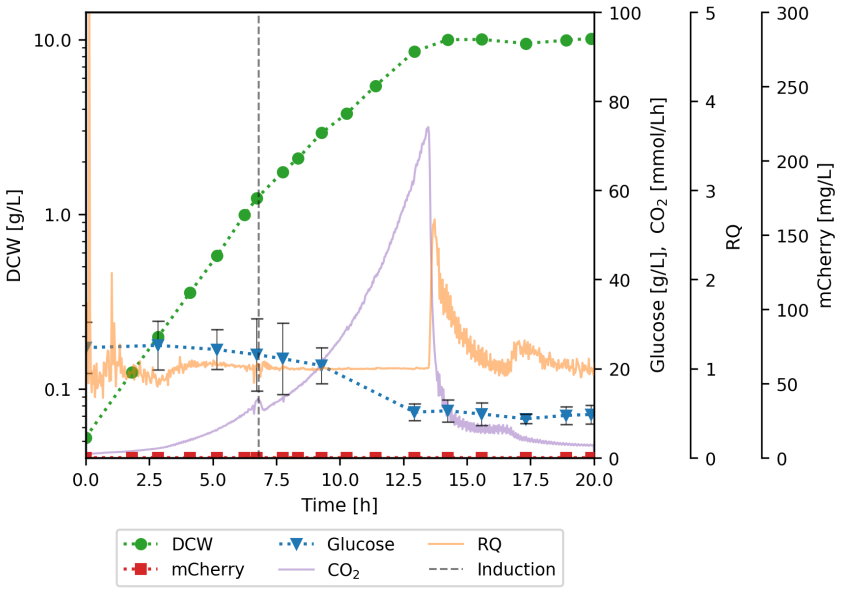
(a) *E. coli* BL21 WT



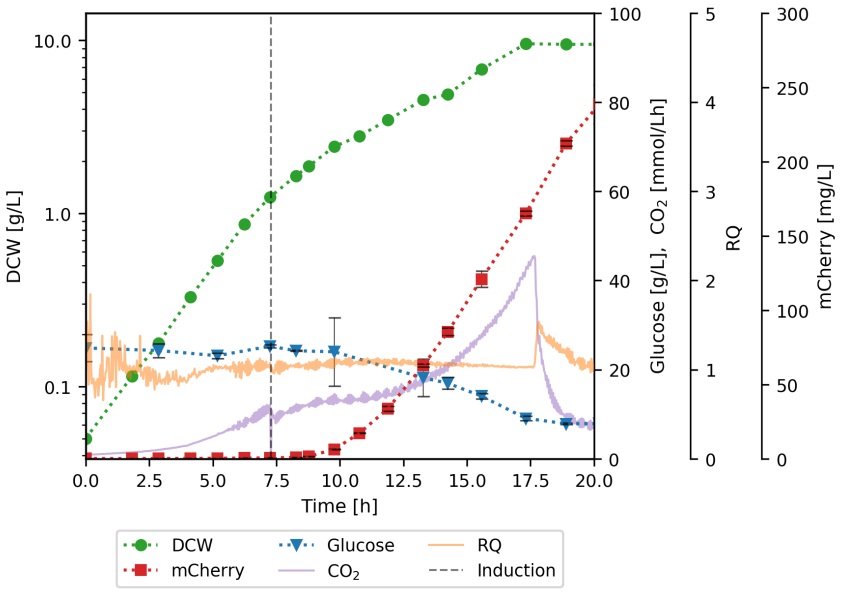
(b) *E. coli* BL21 A2



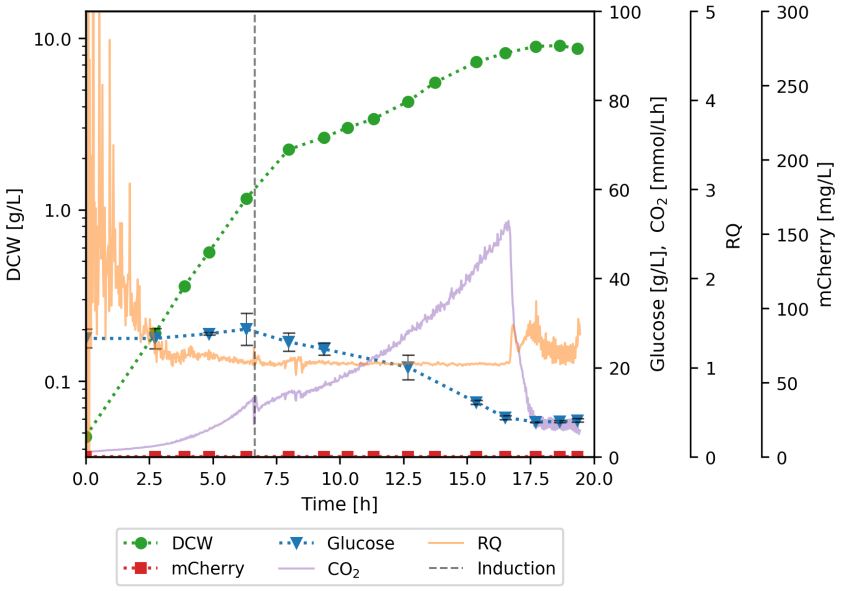
(c) *E. coli* BL21 A2-mCh



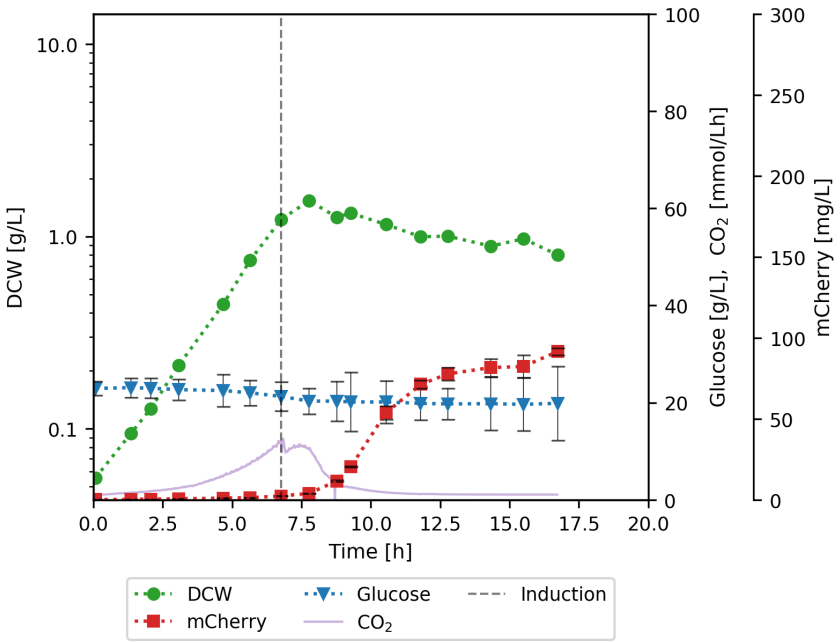
(d) *E. coli* BL21 Z1



(e) *E. coli* BL21 Z1-mCh



(f) *E. coli* BL21 Z2



(g) *E. coli* BL21 Z2-mCh. The RQ was excluded from the plot as the data had high variations, possibly due to the early stationary phase.

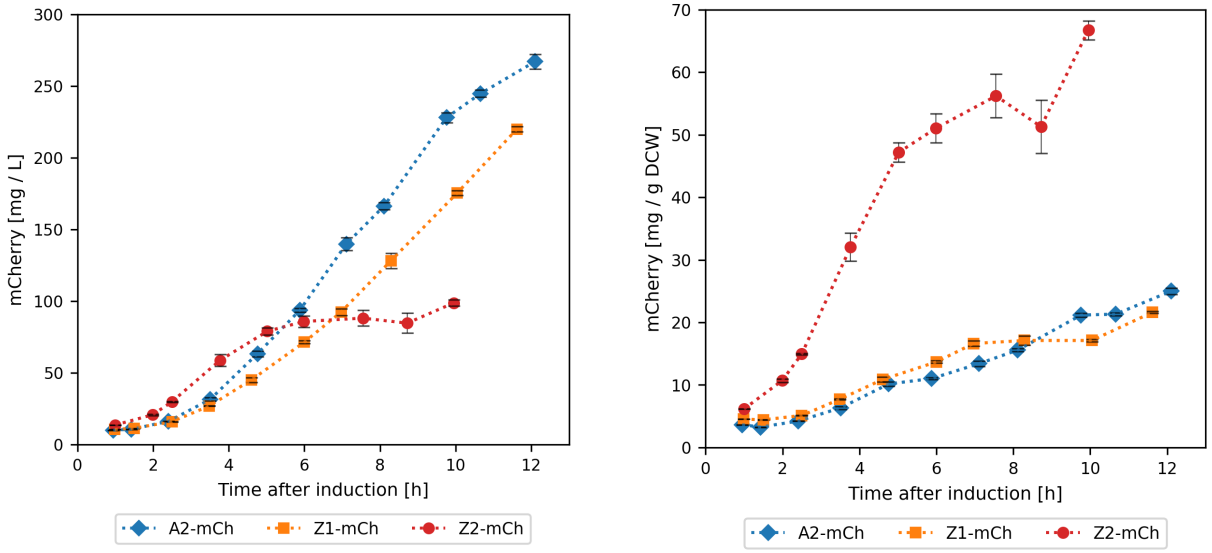
Figure 3.1: Growth curves (DCW [g/L]) were obtained during batch cultivations in bench-scale bioreactors, together with the glucose and mCherry concentration [mg/L], CO_2 production, and RQ throughout the cultivation. The strains were cultivated in Hf.1 medium (Table 2.8) at 30°C, with continuous adjustments of pH to 7 and DO to 40%. The culture media with recombinant strains contained ampicillin (0.1 mg/mL). Once the bacterial concentration reached approximately 1.3 g DCW/L (OD_{600} ~ 2.5), inducer *m*-Toluate (1 mM) was added, and the temperature was lowered to 25 °C. The point of induction is indicated with a black dashed line. The strain corresponding to the growth curve is given in the subfigure captions.

Table 3.2 Summery of the growth curves shown in Figure 3.1. The highest measured biomass concentration [g DCW/L], and the strains growth rate (μ) before and after induction with *m*-toluate (1 mM) and lowering of temperature, and the % reduction of growth rate after induction.

Strain	**Highest biomass concentration [g DCW/L]**	**Growth rate, μ[h^{-1}]**		
		pre-induction	**post-induction**	**% reduction**
WT	10.11	0.53	0.35	34
A2	10.32	0.50	0.25	50
A2-mCh	10.84	0.48	0.25	48
Z1	10.17	0.48	0.30	38
Z1-mCh	9.65	0.47	0.18	62
Z2	9.08	0.47	0.19	60
Z2-mCh	1.55	0.48	0.00	100

3.1.2 mCherry production in A2-mCh, Z1-mCh and Z2-mCh

mCherry production is illustrated in Figure 3.2 as the concentration of mCherry in mg/L and mg/g DCW basis, plotted against time after induction. It can be observed that A2-mCh and Z1-mCh produce more mCherry overall (Figure 3.2a), however, Z2-mCh produces significantly more mCherry on g/ g DCW basis (Figure 3.2b). This indicates that high-level mCherry expression may be one of the factors contributing to the growth inhibition of Z2-mCh observed after induction. A2-mCh and Z1-mCh produce similar amounts of mCherry on a DCW basis. The strains' overall mCherry yield cannot be compared, as the cultivations were stopped before the mCherry concentration plateaued.



(a) mCherry concentration [mg/L]

(b) mCherry concentration [mg/g DCW]

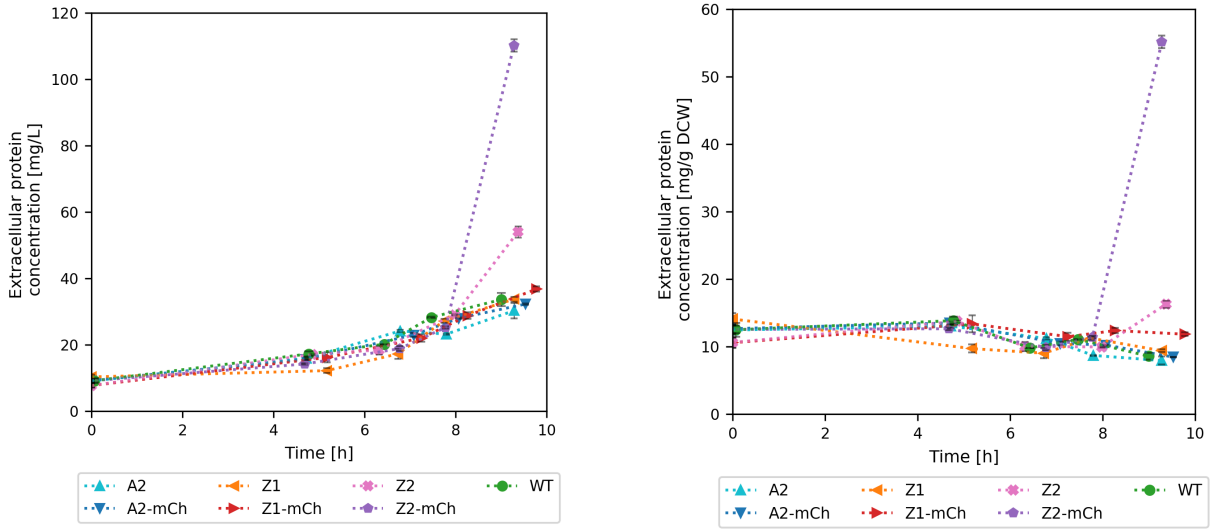
Figure 3.2: Production of mCherry by the strains *E. coli* Bl21 A2-mCh, Z1-mCh and Z2-mCh. The mCherry production is shown as concentration of mCherry plotted against time after induction, expressed as both mg/L (a) and mg/g DCW (b).

3.1.3 The biomass concentration of Z2-mCh decreases after induction due to cell lysis

During the growth characterization cultivations described in the previous subsection, it was observed that Z2-mCh had a decreasing trend in biomass concentration after induction. It was hypothesized that this trend was caused by cell lysis due to the metabolic stress derived from the expression vector. One way to indicate cell lysis during cultivations is to measure protein leakage (Newton *et al.*, 2016). The extracellular protein concentration in supernatant samples from different points in the cultivations was quantified by Bradford protein assay. It should be noted that this is a rough estimation used to compare the protein concentrations between strains and not to quantify the protein concentration. Bovine serum albumin (BSA) was used as a standard to quantify the protein concentrations. BSA and proteins found in *E. coli* are expected to have different amino acid compositions, which will therefore not give ac-

curate quantifications. However, it should give sufficient insight to compare the extracellular protein concentration between the strains.

The extracellular protein concentration is significantly higher in Z2-mCh compared to the other strains (Figure 3.3). The protein concentration is given in g/ g DCW basis (Figure 3.3b) shows that the extracellular protein concentration in Z2-mCh is almost six times higher than the other strains. Another interesting observation is that the protein concentration is also high for Z2, higher than the other strains except for Z2-mCh (Figure 3.3a). During the cultivations described in Subsection 3.1.1, it was observed that Z2's highest measured biomass concentration was lower than what was observed for WT, A2, A2-mCh, Z1, and Z1-mCh (Table 3.2). The results from the protein assay may indicate that cell lysis may have contributed to the low growth rate and biomass formation of Z2 and Z2-mCh after induction.



(a) Extracellular protein concentration [mg/L]

(b) Extracellular protein concentration [mg/g DCW]

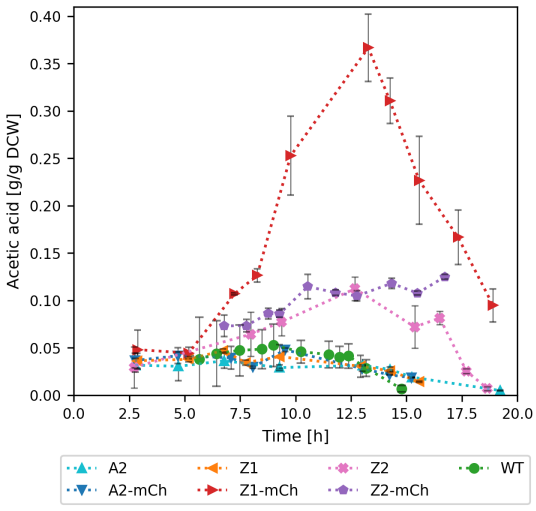
Figure 3.3: Extracellular protein concentration plotted against cultivation time for WT *E. coli* BL21 and the recombinant strains A2, A2-mCh, Z1, Z1-mCh, Z2, and Z2-mCh. The protein concentration is expressed as mg/L (a) and mg/g DCW (b). The protein concentration was quantified in cultivation broth supernatant samples by colorimetry, using a protein assay following the Bradford method, as described in Subsection 2.5.2.

3.1.4 Strains with a higher theoretical metabolic burden produced more overflow metabolites

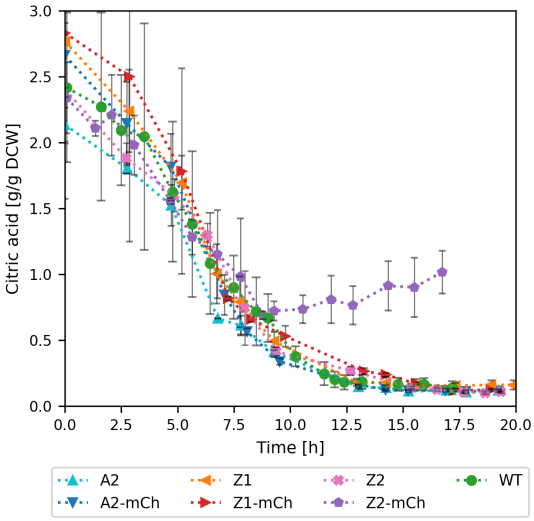
The extracellular concentration of substrates and products were quantified in supernatant samples using HPLC. The glucose quantifications had high variability between technical replicas, as can be observed by the large error bars (Figure 3.1). The high variability in glucose concentration makes calculate substrate based yields inaccurate. As an alternative method to visualize product formation, the product concentrations were plotted against cultivation time in g/g DCW basis.

It was observed in Subsection 3.1.1 that the growth of Z1-mCh, Z2, and Z2-mCh were more

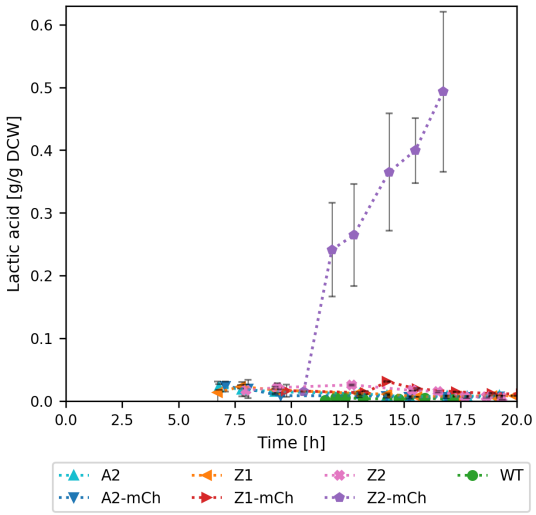
severely affected by induction than the other strains. From Figure 3.4, it can be observed that these strains produced more overflow metabolites than the other strains. Z1-mCh, Z2 and Z2-mCh produced significantly more acetate (Figure 3.6a) compared to the other strains. Additionally, Z2-mCh also produced greater amounts of Lactic acid (Figure 3.6b) and succinic acid (Figure 3.6c) compared to the other strains. The production of these compounds may be a contributor to the reduced growth rates observed after induction.. Citric acid (Figure 3.4b) has a different trend than the other organic acids quantified, as it is a component in the Hf.1 medium (Table 2.8). The concentration of citric acid decreases with the bacterial growth, indicating that the strains do not produce and secrete citric acid. The citric acid curve of Z2-mCh stands out, which is due to the growth inhibition and decreasing biomass observed after induction.



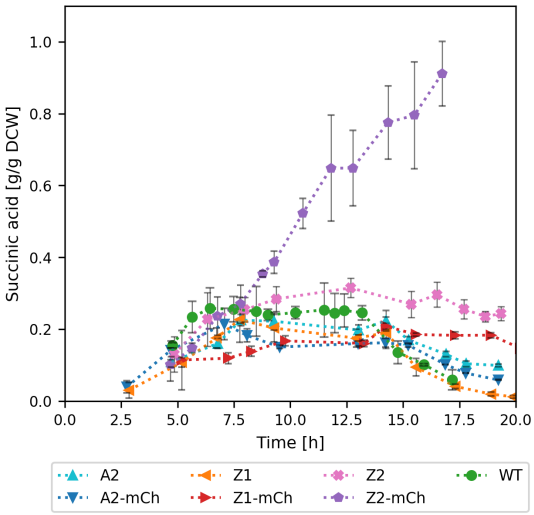
(a) Acetic acid



(b) Citric acid



(c) Lactic acid



(d) Succinic acid

Figure 3.4: Extracellular concentration [g/g DCW] of acetic acid (a), citric acid (b), lactic acid (c) and succinic acid (d) plotted against cultivation time for WT *E. coli* BL21 and the recombinant strains A2, A2-mCh , Z1, Z1-mCh, Z2, and Z2-mCh. The acids were quantified in samples of culture broth using the HPLC procedure described in Subsection 3.1.4.

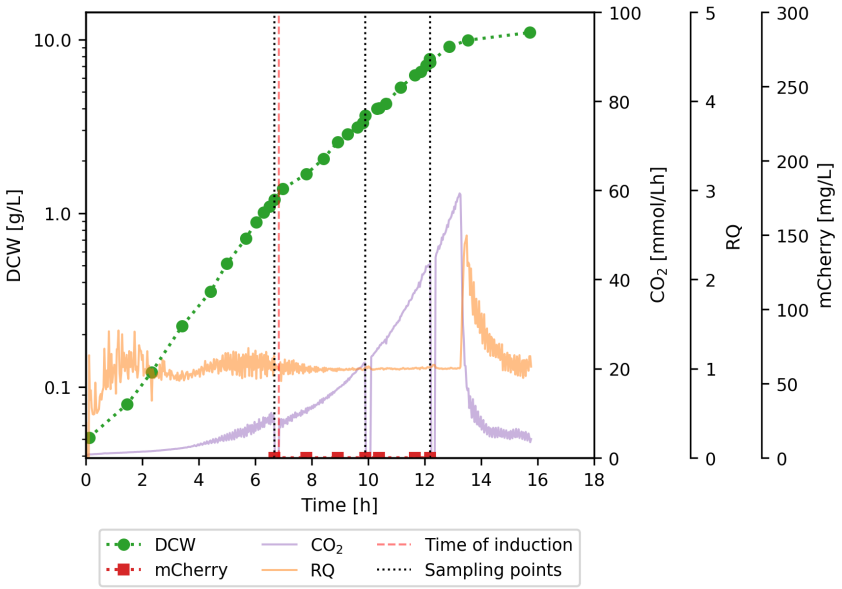
3.2 The effects of expression vector presence on the amino acid and TCA cycle metabolite pools in *E. coli*

The quantification of metabolites was performed to study how the central carbon metabolism reacted to the metabolic burden imposed by the expression vector. All strains could not be studied due to limited time and capacity. The strains *E. coli* BL21 WT, A2, A2-mCh, Z2, and Z2-mCh were selected for the analysis, as it would be interesting to compare the central carbon metabolism of recombinant strains that are not subjected to much metabolic burden (A2 and A2-mCh), with strains that are severely affected (Z2 and Z2-mCh). WT was included so the metabolism of the recombinant strains could be compared to the one of WT.

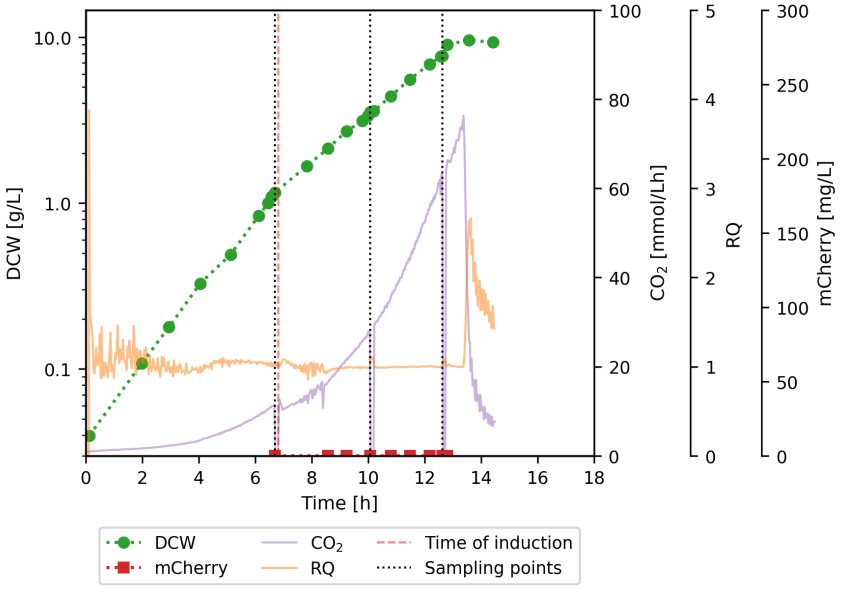
The recombinant strains were analyzed in one biological replica, while the WT strain was analyzed in biological triplicates (WT-I, WT-II, and WT-III) to examine the repeatability between cultivations. Endometabolome samples were taken in at least 4 technical replicas at biomass concentration 1.3 g DCW/L (T1) right before induction, 3.5 g DCW/L (T2) and 7.5 g DCW/L (T3) for all strains except Z2-mCh. Due to Z2-mCh's decreasing trend in biomass concentration after induction, it was sampled at biomass concentration 1.3 g DCW/L (T1) as the other strains, 30 min after induction (T2), and 4.5 h after induction (T3), which is when T2 was sampled for Z2, the corresponding negative control strain. Amino acids, TCA intermediates, pyruvate, and lactate were quantified. Phosphorylated metabolites were supposed to be quantified by capIC-MS/MS, however, this could not be carried out as the equipment was under maintenance. Therefore, the focus will be on the changes observed in the AAs and TCA metabolite pools.

3.2.1 Bioreactor cultivation with endometabolome sampling

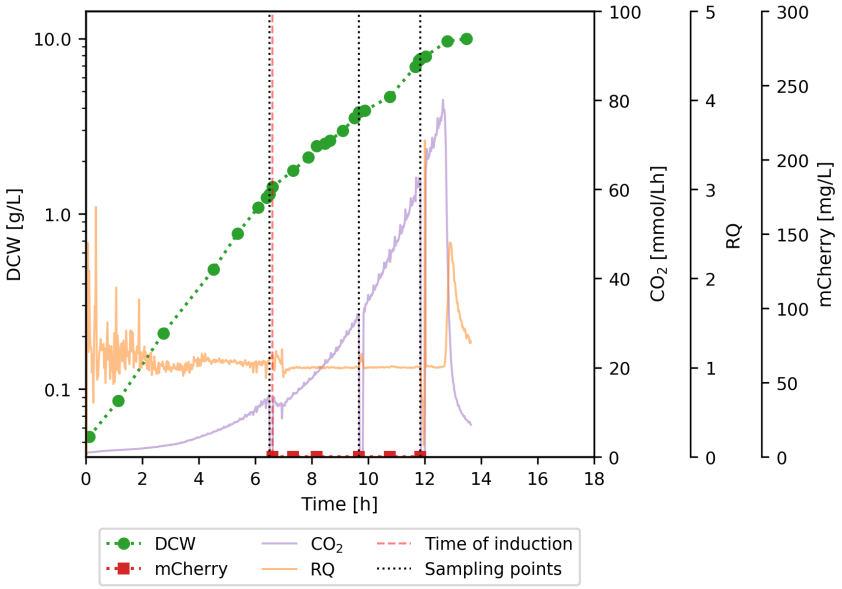
The growth curves are given in Figure 3.5, which also includes mCherry concentration, CO_2 production, and RQ throughout the cultivations. The growth rates are listed in Table 3.3, together with the biomass concentration at the sampling points. The growth curves (Figure 3.5) and estimated growth rates (Table 3.3) shows that the strains acted similarly as observed during the growth characterization cultivations (Section 3.1). The growth rates of the WT replicas were reduced with 30 to 37% after induction, while they were approximately halved for A2 and A2-mCh. Z2 had a greater decrease in growth rate at 54%, which is less than what was observed during the previous cultivation at 60% (Table 3.2). The growth of Z2-mCh was completely inhibited after induction, followed by a decreasing trend in biomass concentration, which is the same trend as previously observed.



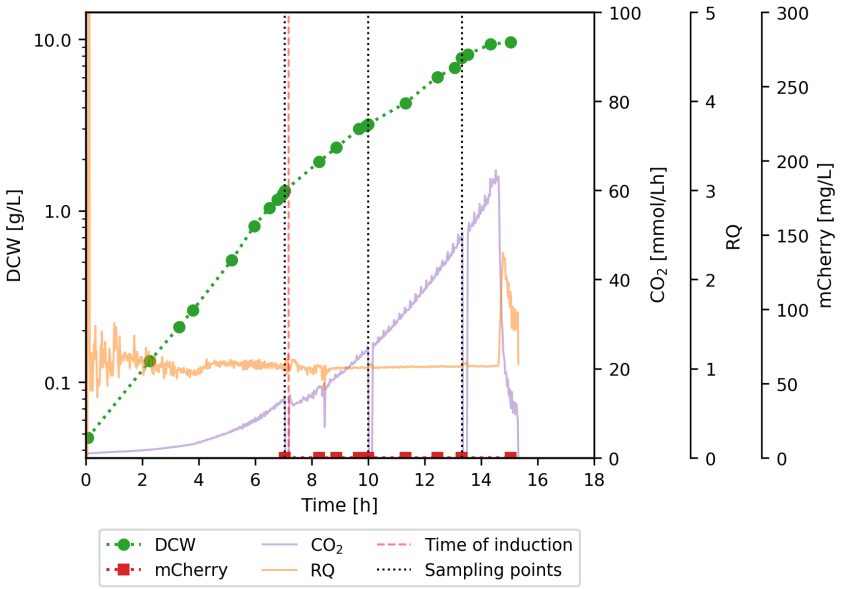
(a) *E. coli* BL21 WT-I



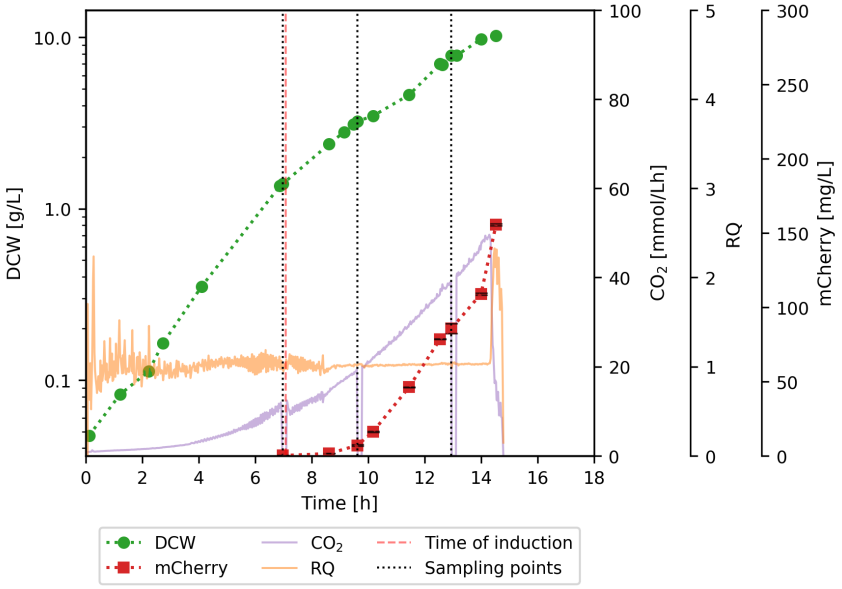
(b) *E. coli* BL21 WT-II



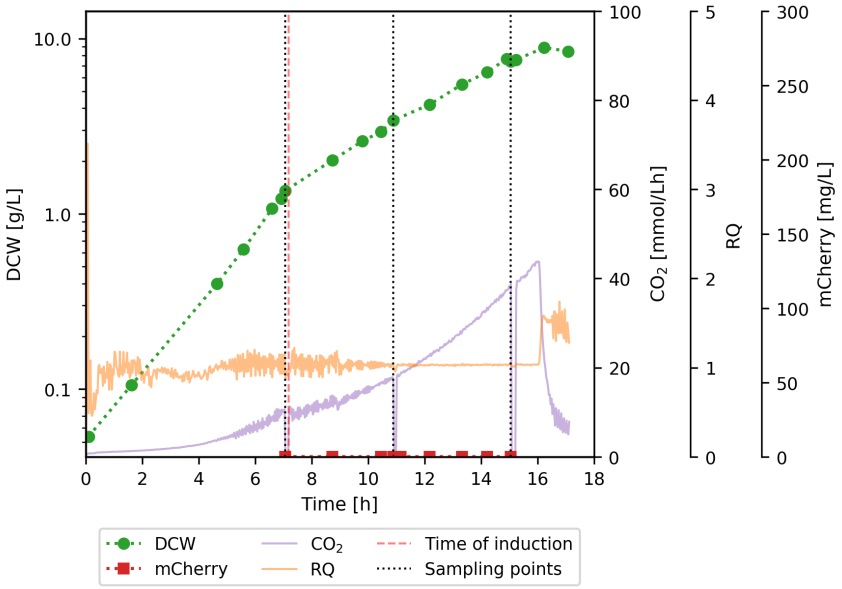
(c) *E. coli* BL21 WT-III



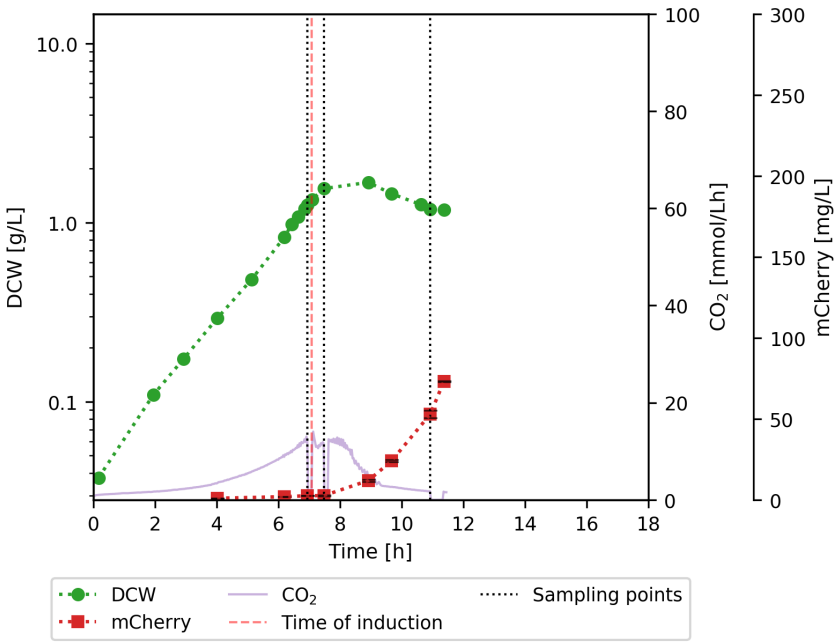
(d) *E. coli* BL21 A2



(e) *E. coli* BL21 A2-mCh



(f) *E. coli* BL21 Z2



(g) *E. coli* BL21 Z2-mCh. The RQ is not included in the plot as the data had high variations, possibly due to the early entry of stationary phase. Due to Z2-mCh's decreasing trend in biomass concentration after induction, it was sampled at the biomass concentration 1.3 g CDW/L (T1) as the other strains, 30 min after induction (T2), and 4.5 h after induction (T3), which is when T2 was sampled for Z2, the corresponding negative control strain.

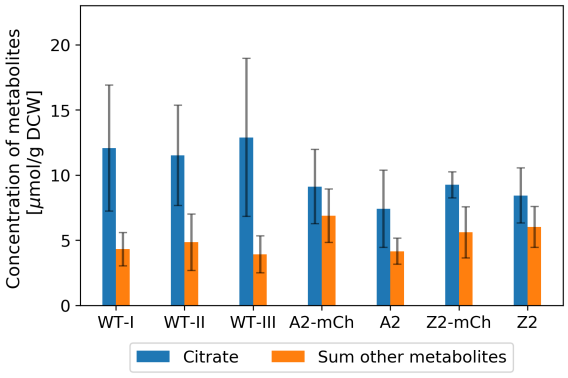
Figure 3.5: Growth curves [g CDW/L] obtained during batch cultivations in bench-scale bioreactors, together with mCherry concentration, CO_2 production, and RQ measured throughout the cultivation. The strains were cultivated in Hf.1 medium (Table 2.8) at 30°C, with continuous adjustment of pH to 7 and DO to 40%. Ampicillin (0.1 mg/mL) was added to recombinant strains. Endometabolome samples were taken as described in Subsection 2.6.1 at three points in the cultivation T1 (~1.3 g DCW), T2 (~3.6 g DCW) and T3 (~7.5 g DCW), indicated by dotted lines. After the first endometabolome sampling, inducer *m*-toluate (1mM) was added, and the temperature was lowered to 25°C. The point of induction is indicated by a red dashed line. The strain corresponding to the growth curve is given in the subfigure captions.

Table 3.3 Biomass concentration [g DCW/L] at sampling point T1, T2, and T3 , and the strains growth rate (μ) before and after induction with *m*-toluate (1 mM) and lowering of temperature, and the % reduction of growth rate after induction

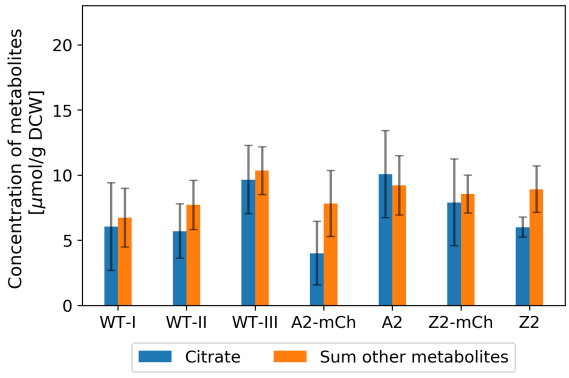
Strain	**DCW T1**	**DCW T2**	**DCW T3**	**Growth rate, μ[h^{-1}]**		
				pre-induction	**post-induction**	**% reduction**
WT-I	1.20	3.65	7.40	0.53	0.37	30
WT-II	1.16	3.54	7.69	0.49	0.33	33
WT-III	1.42	3.80	7.64	0.51	0.32	37
A2	1.31	3.20	7.79	0.50	0.25	50
A2-mCh	1.39	3.24	7.89	0.50	0.26	48
Z2	1.36	3.92	7.43	0.46	0.21	54
Z2-mCh	1.25	1.55	1.19	0.49	0.00	100

3.2.2 Citric acid in the medium contaminated the quantifications of intracellular citrate

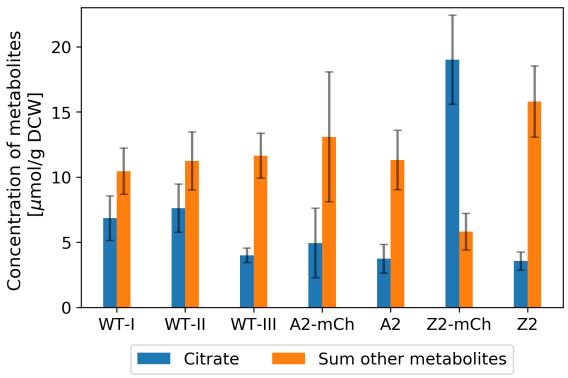
During the data processing of the metabolomic data, it was observed that the concentration of citrate was abnormally high in all endometabolome samples (Figure 3.6). The concentration of citrate at T1 is significantly higher than the combined concentration of all other quantified metabolites for all strains. The citrate concentration is also unusually high in the samples from T2 and T3 as well, although not as extreme as in T1. These abnormally high citrate concentrations contradict the findings of a previous study performed by (Bennett *et al.*, 2009), where the intracellular glutamate concentration alone far exceeded the concentration of citrate in *E. coli*. These abnormal results are most likely caused by the fact that citric acid is a component in the Hf.1 medium (Table 2.8) the strains were cultivated in, and insufficient washing of the cells during endometabolome sampling. The citrate data were therefore excluded from the further analysis of metabolomic data.



(a) Sampling point T1



(b) Sampling point T2



(c) Sampling point T3

Figure 3.6: The concentration [µmol/g DCW] of citrate, compared with the combined concentration [µmol/g DCW] of all other metabolites quantified. The metabolites were quantified in the endometabolome samples from sample point T1 (a), T2 (b), and T3 (c). The samples were obtained from the cultivation of the biological triplicates of the WT *E. coli* BL21 strain, and the one biological replica of the recombinant strains A2, A2-mCh, Z2, and Z2-mCh. The endometabolome samples were taken as described in Subsection 2.6.1, in at least four technical replicas. The metabolites were quantified by RP LC-MS/MS as described in Subsections 2.6.2 and 2.6.3. The metabolomic data was processed as described in Subsection 2.6.5.

3.2.3 Test of reproducibility between biological replicas of WT *E. coli* BL21

The metabolomic analysis was performed on biological triplicates of the WT *E. coli* BL21 strain (WT-I, WT-II, and WT-III), to test the reproducibility of the trends observed in metabolite pools between different cultivations. The biological triplicates were compared in a 2D score plot (Figure 3.7) bar charts illustrating the overall concentration of quantified AA and OA (Figure 3.8), and a heat map illustrating $\log_2$ fold changes in the biological triplicates relative to each other (Figure 3.9).

The 2D score plot (Figure 3.7) accounts for 74.47% of the variation in metabolomic data between the biological triplicates, where principal component (PC) 1 explains 55.61% of the variance and PC2 explains 18.86%. Sampling point T1 differs significantly from T2 and T3 when they are compared based on PC1. This trend is expected as T1 was before induction, while T2 and T3 were after. The differences of T1 compared to T2 and T3 are caused by the lowered temperature, not the addition of *m*-toluate, as the WT strain does not contain an expression vector. T1 also appears to have the most variation between WT-I, WT-II, and WT-III based on PC2. The biological triplicates are closely grouped at sampling point T3. Sampling point T2 shows an interesting trend where WT-I and WT-II are closely grouped, while WT-III is grouped with the triplicates at sampling point T3. However, it was not expected to observe much difference between the triplicates at sampling point T2 and T3, as both samples were taken after the temperature was lowered.

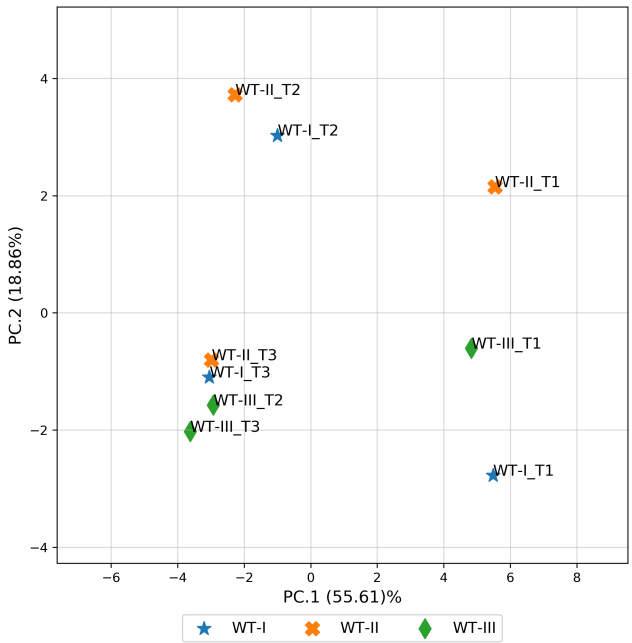
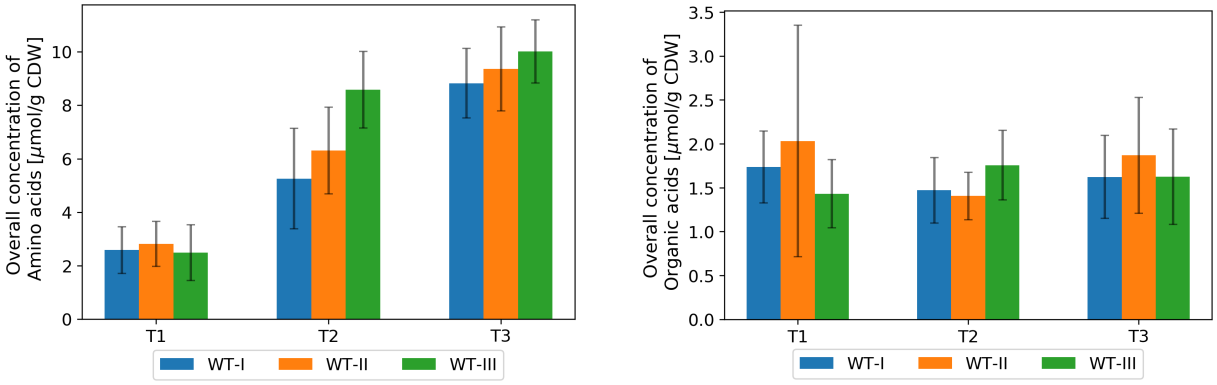


Figure 3.7: A 2D score plot from PCA of the metabolite concentrations quantified in the endometabolome samples taken during the biological triplicate cultivations of WT *E. coli* BL21 (WT-I, WT-II and WT-III). The endometabolome samples were taken as described in Subsection 2.6.1, in at least four technical replicas. The metabolites were quantified by RP LC-MS/MS, as described in Subsections 2.6.2 and 2.6.3.The metabolomic data was processed as described in Subsection 2.6.5. The PCA was performed on the metabolomic data normalized by sum and autoscaled, using the Python script given in Appendix F.2.

The overall concentrations are fairly similar between WT-I, WT-II, and WT-III in most sampling points (Figure 3.8) . An exception is WT-III at sampling point T2, which has a higher overall concentration of AAs than WT-I and WT-II, with a concentration more similar to the biological triplicates at T3. This trend coincides with the observations in the 2D scores plot (Figure 3.7) where WT-II at T2 were closely grouped with all biological replicas at T3. The overall concentrations of the biological replicas at all the other sampling points have similar concentrations with overlapping uncertainties. An interesting observation is that the amino acid pool increases with cultivation time. The increase from T1 to T2 indicates that the cells react to the temperature being lowered from 30°C to 25 °C by producing more amino acids. The increase from T2 to T3 may be a result of the cell signaling that it is approaching the stationary phase. In contrast, the TCA cycle metabolite pool stays relatively between the sampling points.



(a) Overall concentration of amino acids.

(b) Overall concentration of organic acids.

Figure 3.8: The overall concentration [µmol/g DCW] of amino acids (a) and organic acids (b), quantified in endometabolome samples taken during the biological triplicate cultivations of *E. coli* BL21 (WT-I, WT-II, and WT-III). The endometabolome samples were taken as described in Subsection 2.6.1, in at least four technical replicas. The AAs and OAs were quantified by RP LC-MS/MS, as described in Subsections 2.6.2 and 2.6.3, respectively. The metabolomic data was processed as described in Subsection 2.6.5.

The biological replicas had almost identical relative metabolite concentrations, only with variations in valine and isocitrate (Figure 3.9). The ICit MS data had a low signal-to-noise ratio (S/N) and relatively high variation between technical replicas, meaning the ICit concentration have larger uncertainties compared to the other metabolites, and should therefore be interpreted with caution for future results. It was decided not to exclude any of the biological WT replicas for further analysis, as they were very similar when compared in the 2D score plot (Figure 3.7), bar chart (Figure 3.8), and heatmap (Figure 3.9). The average of the biological triplicates (WT-I, WT-II, and WT-III) at each sampling point (T1, T2, and T3) will be used to further analyze metabolomic data.

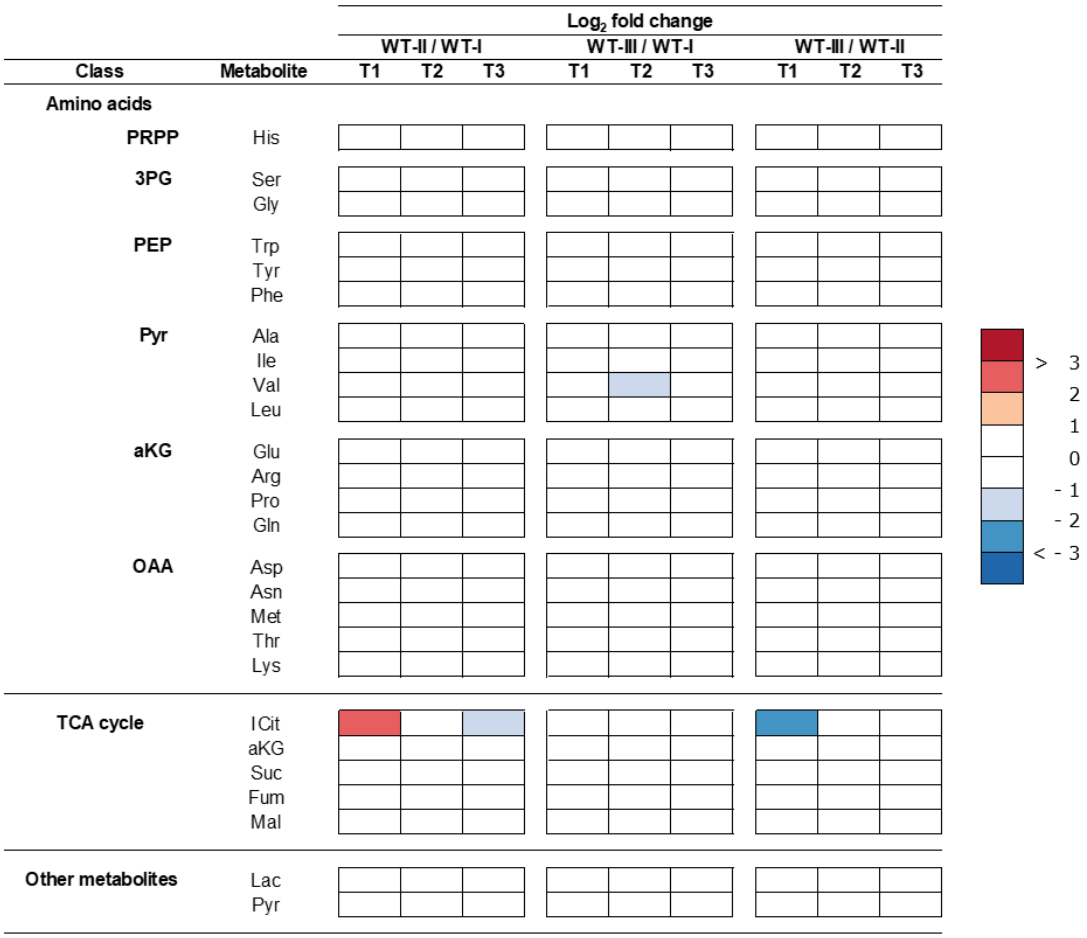
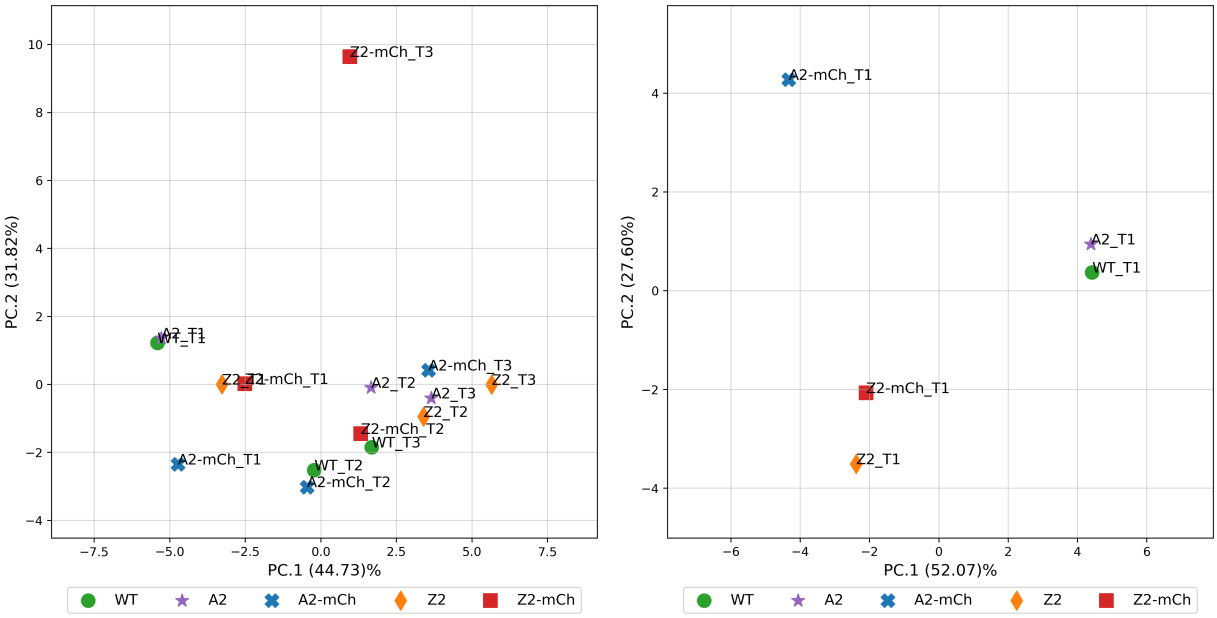


Figure 3.9: Heat map illustrating the $\log_2$ fold change of the quantified metabolites from the endometabolome samples at all sampling points (T1, T2, and T3), in *E. coli* BL21 WT-II relative to WT-I, WT-III relative to WT-I, and WT-III relative to WT-II. The metabolites are categorized in amino acids, TCA cycle intermediates and other metabolites, with amino acids being further categorized by their precursor metabolite in the central carbon metabolism. The heat map is visualized with a discrete color map, where red indicates a relative increase in metabolite concentration, while blue indicates a reduction. The minimum and maximum values of the color map are set to -3 and 3. Metabolite abbreviations are given in Appendix A.

3.2.4 Principal component analysis of the metabolomic data

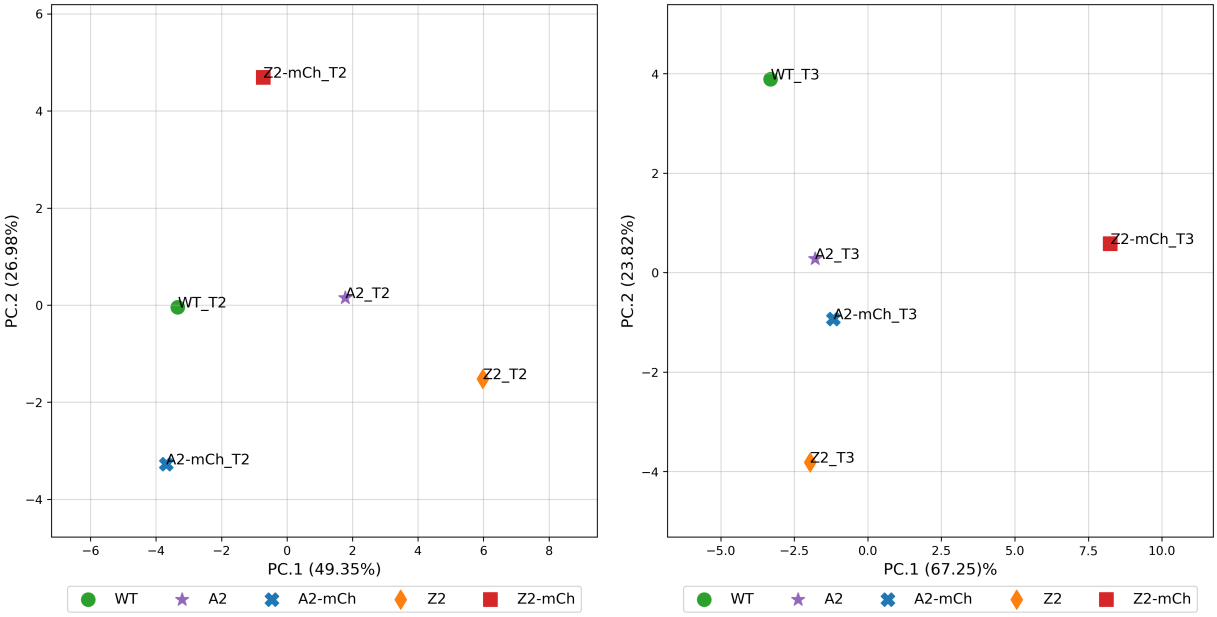
The PCA generated the 2D score plots shown in Figure 3.10. The figure includes 2D score plots of metabolomic data from all sampling points (Figure 3.10a), T1 (Figure 3.10b), T2 (Figure 3.10c), and T3 (Figure 3.10d). The 2D score plot with all sampling points (Figure 3.10a) shows that samples from T1 are separated from the T2 and T3 samples, based on PC1. This separation is most likely caused by the fact that T1 was before induction, while T2 and T3 were after. Samples from T2 and T3 are clustered relatively closely, with most T2 samples being slight to the left in PC1 and/or lower on PC2 compared to samples of the same strain at T3. An exception to this trend is Z2-mCh sampled at T3. This sample stands out from all other endometabolome samples, both in the score plot of all sampling points (Figure 3.10a) and the score plot with only T3 samples (Figure 3.10d). This isolation of Z2-mCh coincides with the cultivation results. Z2-mCh had complete growth inhibition after induction, while the other strains reacted differently by maintaining growth at a lower growth rate (Table 3.3). An interesting observation from the score plot with all sampling points (Figure 3.10a) is that Z2-mCh sampled at T2 is closely clustered with the T2 and T3 samples of the other strains. This is unexpected as the Z2-mCh sample from T2 was taken at a completely different time point in the cultivation compared to the other strains. The T2 sample of Z2-mCh was taken 30 min after induction, while the other strains' T2 sample was taken roughly three to four hours after induction.

The T1 2D score plot (Figure 3.10b) shows that A2 and WT are grouped and that Z2-mCh is closely located to Z2, while A2-mCh stands out. In the T2 2D score plot (Figure 3.10c), none of the strains seem to cluster together, which may indicate metabolic similarities between the strains at T2 with no clear outlier. However, it can be seen that the two mCherry producing strains are on the complete opposite sides of PC2 in the 2D score plot. As previously described, the 2D score plot of T3 (Figure 3.10d) shows that Z2-mCh distinctly stands out from the other strains on the basis of PC1, which explains most of the variation in the data set, at 67.25%. This separation of Z2-mCh from the other strains is most likely caused by the different behavior after induction, as described above. Another observation is that the sample of Z2 at T3 is located on the opposite side of PC2 compared to WT. The large difference between WT and Z2 is expected, as previous results indicate that Z2 is exposed to a metabolic burden after induction, as observed by the low growth rates post-induction (Figures 3.1 and 3.5)



(a) 2D score plot of all sampling points

(b) 2D score plot of sampling point T1



(c) 2D score plot of sampling point T2.

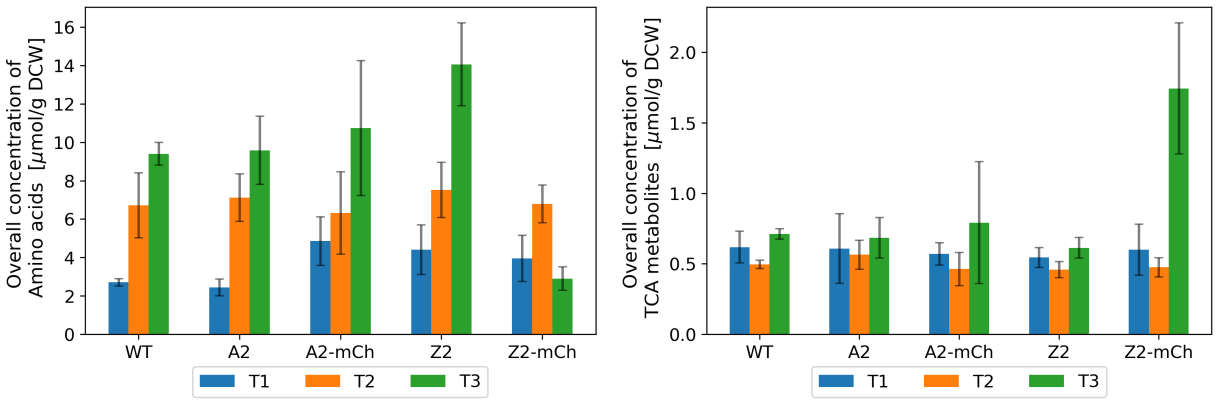
(d) 2D score plot of sampling point T3.

Figure 3.10: 2D score plots from PCA of the metabolite concentrations in the endometabolome samples taken during the cultivations of *E. coli* BL21 WT, A2, A2-mCh, Z2, and Z2-mCh. The 2D score plots are of metabolomic data from all sampling points (a), T1 (b), T2 (c), and T3 (d). The metabolomic data of WT is given as the average of biological triplicates, while the recombinant strains are from one biological replica. The endometabolome samples were taken as described in Subsection 2.6.1, in at least four technical replicas. The metabolites were quantified by RP LC-MS/MS, as described in Subsections 2.6.2 and 2.6.3.The metabolomic data was processed as described in Subsection 2.6.5. The PCA was performed on autoscaled metabolomic data normalized by sum, using the Python script given in Appendix F.2.

3.2.5 Trends in overall concentrations of amino acids and TCA cycle intermediates between sampling points

Bar charts were generated to illustrate the changes in the overall concentration of amino acids and TCA intermediates between sampling points. When comparing the overall AA concentration (Figure 3.11a), Z2-mCh clearly stands out from the other strains. The AA concentration in the other strains increases with cultivation time, while the AA concentration in Z2-mCh is at its lowest at T3. This observation indicates that while the other strains accumulate AA with time, Z2-mCh gets depleted. The low AA concentration at T3 may be due to high-level mCherry expression, as Z2-mCh produces significantly more mCherry per g DCW than A2-mCh (Figure 3.2). The bar chart also shows that A2 has an almost identical overall AA concentration to WT at all sampling points. Other trends that can be observed is that A2-mCh, Z2, and Z2-mCh have slightly higher AA concentration at T1, and Z2 have a higher concentration of AA at T3, compared to the other strains.

Z2-mCh also stands out from the other strains when comparing the concentration of TCA cycle intermediates (Figure 3.11b). WT, A2, A2-mCh, and Z2 have a relatively constant concentration of TCA intermediates between the sampling points, with a slightly lower level at T2 compared to T1 and T3. In contrast, Z2-mCh has a significantly higher concentration of TCA cycle intermediates at T3 compared to the other sampling points and the other strains. The accumulation of TCA cycle intermediates in Z2-mCh at T3 may indicate that the TCA cycle is a factor contributing to the growth inhibition observed after induction during the cultivations of Z2-mCh (Figures 3.1 and 3.5).



(a) Overall concentration of amino acids.

(b) Overall concentration of TCA cycle metabolites.

Figure 3.11: The overall concentration [µmol/g DCW] of amino acids (a) and TCA cycle metabolites (b), quantified in the endometabolome samples taken during the cultivations of *E. coli* BL21 WT, A2, A2-mCh, Z2, and Z2-mCh. The endometabolome samples were taken as described in Subsection 2.6.1, in at least four technical replicas. The AAs and TCA cycle metabolites were quantified by RP LC-MS/MS, as described in Subsections 2.6.2 and 2.6.3, respectively.The metabolomic data was processed as described in Subsection 2.6.5.

3.2.6 Heat map illustrating $\log_2$ ratio in metabolite levels between recombinant strains and WT

The heat map (Figure 3.12) was generated to illustrate the $\log_2$ fold changes of metabolite levels in the recombinant *E. coli* BL21 strains A2, A2-mCh, Z2, and Z2-mCh relative to the WT strain. It can be observed that the relative metabolite levels in A2, A2-mCh, and Z2 are almost identical to that of WT in all sampling points, with only a few variations. This observation indicates that the metabolic burden these expression vector places on the host are not large enough to initiate the host to reroute the metabolism of amino acid and TCA cycle metabolites. In contrast, Z2-mCh has quite large differences in the relative metabolite levels compared to WT at sampling point T3. Z2-mCh has an accumulation of isocitrate (ICit), α-ketoglutarat (aKG), and succinate (Suc) in the TCA cycle. The concentration of ICit is over eight times as high in Z2-mCh relative WT, aKG is over four times as high, and Suc is over twice as high. In contrast, the relative concentration of almost half of the amino acids is decreased in Z2-mCh compared to WT. The decrease is mainly in the amino acids that are synthesized from precursors in the TCA cycle, proline (Pro), glutamine (Gln), asparagine (Asn), methionine (Met), threonine (Thr), and lysine (Lys). The amino acids valine (Val) and tyrosine (Tyr) are also less abundant in Z2-mCh relative to WT. The relative increase in TCA intermediates and decrease in amino acids coincides with the observations in the bar chart illustrating the overall concentration of amino acids (Figure 3.11a) and TCA intermediates (Figure 3.11b). The bar charts show that Z2-mCh has a significantly lower total amino acid pool at T3 compared to the other strains and a significantly higher TCA intermediate pool.

The heat map indicates that Z2-mCh has an accumulation of intracellular lactate, as seen by the relative concentration being over four times as high in Z2-mCh compared to WT. This observation corresponds by the HPLC quantification of extracellular products (Figure 3.4). Z2-mCh had a significantly higher concentration of lactic acid (Figure 3.6b), the conjugate acid of lactate. The observations from the heat map may indicate that the TCA is a limiting bottleneck in the energy generating pathways in Z2-mCh. There is an accumulation of metabolites in the TCA cycle, which can be a sign of limited capacity. Additionally, the concentration of the overflow metabolite lactate was accumulated in the cells and excreted in higher amounts compared to the other strains.

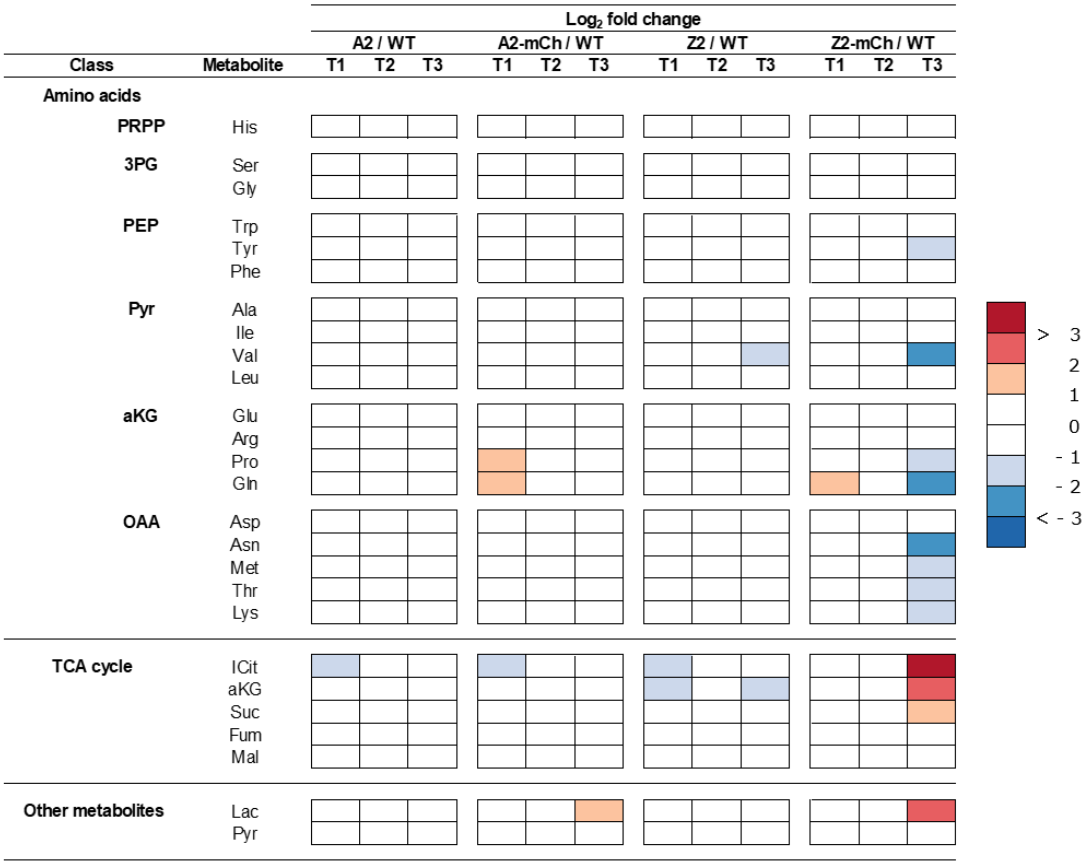
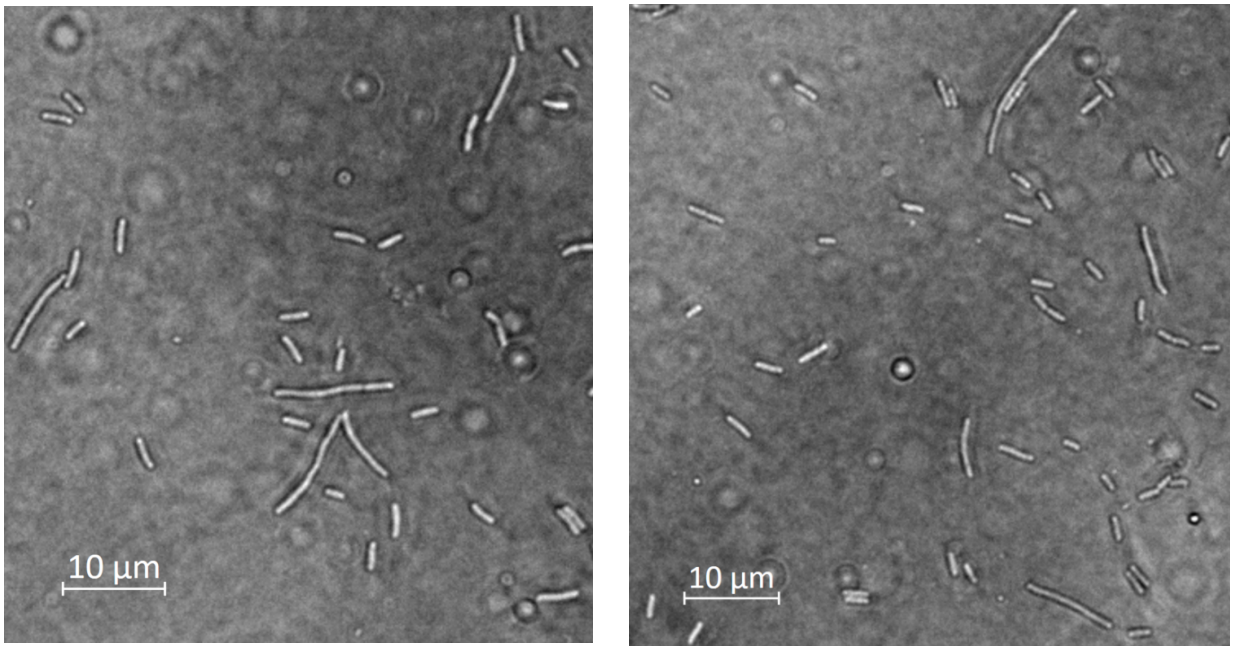


Figure 3.12: Heat map illustrating the $\log_2$ fold change of quantified metabolites from all sampling points (T1, T2, and T3), in the recombinant *E. coli* BL21 strains A2-mCh, A2, Z2 and Z2-mCh, relative to the WT strain. The metabolites are categorized in amino acids, TCA cycle metabolites, and other metabolites, with the amino acids being further categorized into their precursor metabolite, PRPP, 3PG, PEP, Pyr, aKG, or OAA. The heat map is visualized with a discrete color map, where red indicates a relative increase in metabolite concentration, while blue indicates reduction. The minimum and maximum values of the color map is set to -3 and 3. Metabolite abbreviations are given in Appendix A.

3.3 Comparison of cellular length by microscopy

Filamentation is used to describe the abnormal bacterial growth where the cells continue to grow without dividing, resulting in elongated cells with multiple chromosomal copies (Jaimes-Lizcano *et al.*, 2014). A cellular length study was performed on the strains WT, A2, A2-mCh, Z2, and Z2-mCh to examine the occurrence of filamentation. The study was performed by measuring the length of at least 100 cells in two independent cultures of each strain, both before and after induction (Section 2.7). Unexpectedly, filamentous cells were observed in all strains, including WT, both before and after induction. Images of smaller filamentous cells observed in WT cultures before and after induction is shown in Figure 3.13, and an image of a longer filamentous cell observed in a Z2-mCh culture after induction is shown in Figure 3.14.



(a) *E. coli* BL21 WT at T1

(b) *E. coli* BL21 WT at T2

Figure 3.13: Phase contrast images of WT *E. coli* BL21. (a) Cells taken from a culture at OD_{600} 1 grown in Hi medium (Table 2.6) at 30°C. (b) Cells taken from the same culture at OD_{600} 2.5, after addition of *m*-toluate (1 mM) and lowered temperature to 25°C at OD_{600} 1. Both images show presence of filamentous cells.

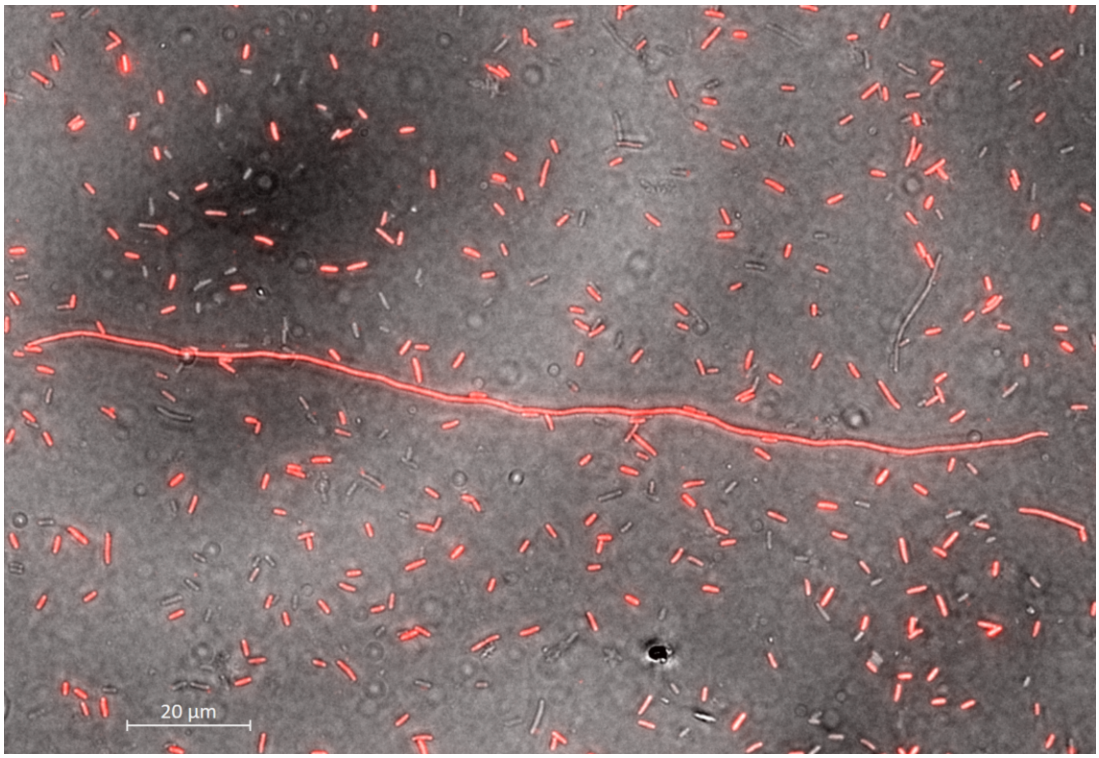
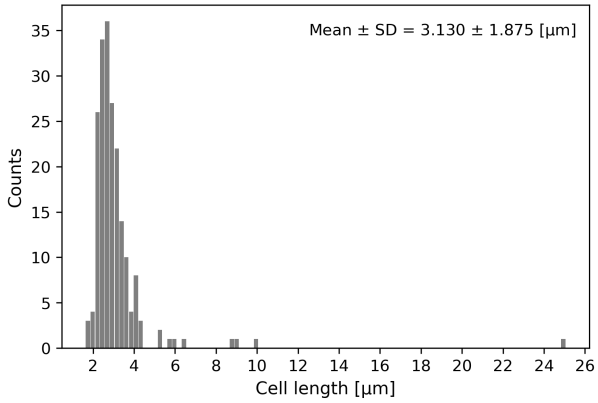
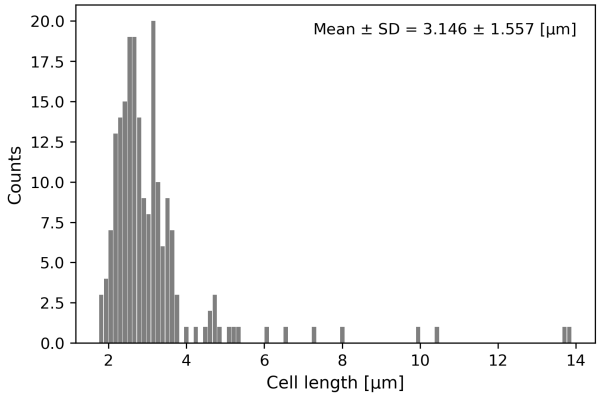


Figure 3.14: Combined fluorescence and phase contrast of *E. coli* BL21 Z2-mCh. The strain was grown in Hi medium (Table 2.6) at 30°C to OD_{600} 1, after which mCherry expression was induced by addition of *m*-toluate and the temperature was lowered to 25°C. The image were taken once the cell density reached OD_{600} 2.5. The image show a large filamentous cell. The red fluorescence is a caused by expression of the reporter protein mCherry.

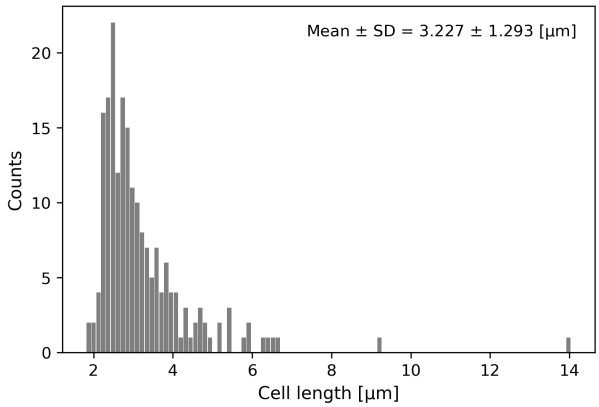
Statistical tests were performed on the microscopy data as shown in Appendix G. The significance level was set to 0.05 for all tests. First, independent t-tests were used to compare the mean cellular length between the biological replicas of the same strain, before and after induction, separately. The tests concluded that the biological replicas did not have significantly different mean cellular lengths. The data of the two biological replicas were combined for all strains at each time point and used to make the histograms (Figure 3.15) illustrating distribution of measure cell lengths, and to determine the mean cellular lengths (Table 3.4). The histograms (Figure 3.15) show that most cells had lengths between 2 and 4 µm. Recombinant strains seem to have a greater occurrence of filamentous cells after induction compared to WT. This can be observed by the higher amounts of cells with a length above 4 µm, and the increased range in cell sizes measured at T2. Additionally, the recombinant strains had more filamentous cells after induction than before.



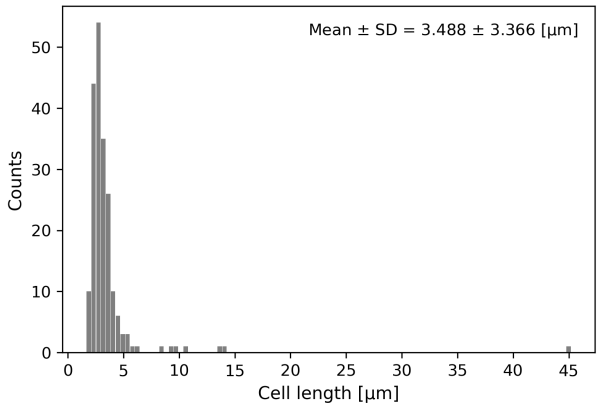
(a) *E. coli* BL21 WT T1



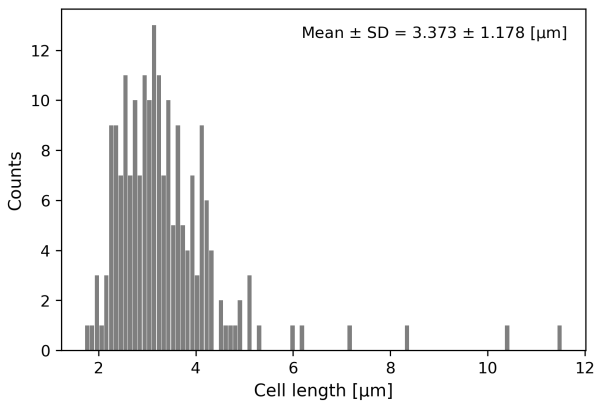
(b) *E. coli* BL21 WT T2



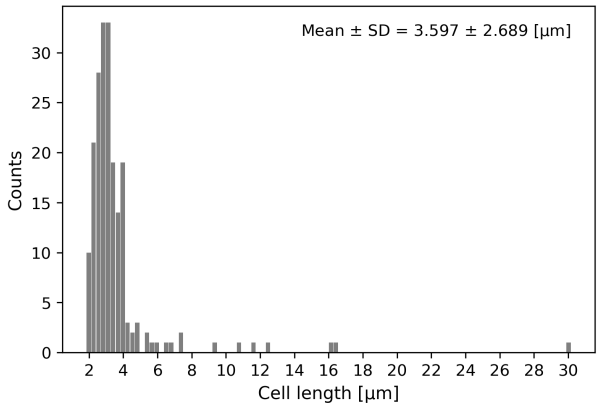
(c) *E. coli* BL21 A2 T1



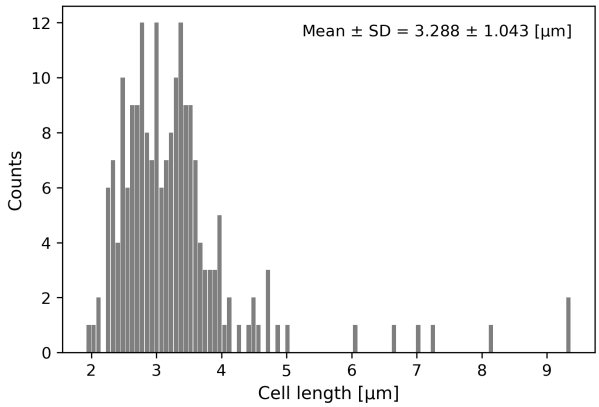
(d) *E. coli* BL21 A2 T2



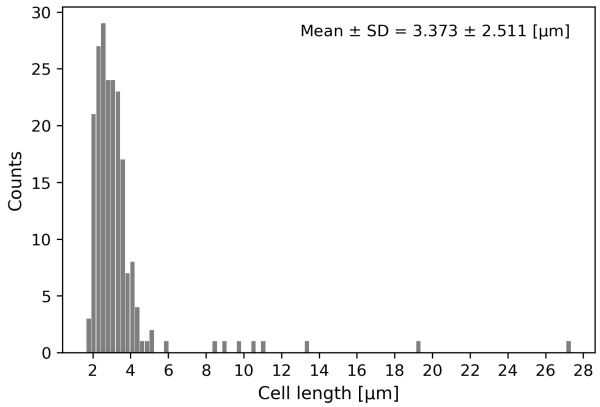
(e) *E. coli* BL21 A2-mCh T1



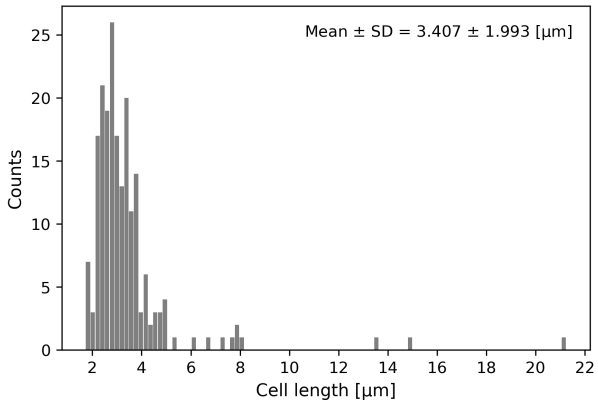
(f) *E. coli* BL21 A2-mCh T2



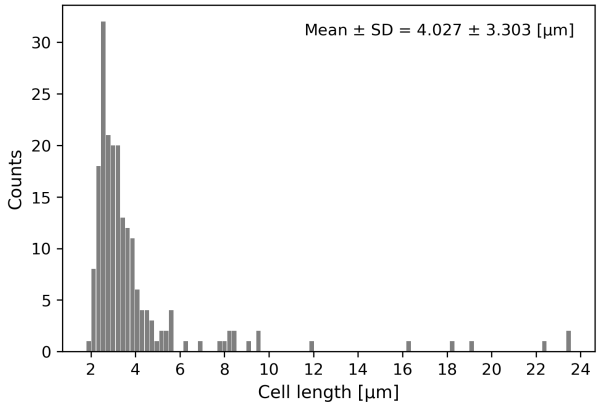
(g) *E. coli* BL21 Z2 T1



(h) *E. coli* BL21 Z2 T2



(i) *E. coli* BL21 Z2-mCh T1



(j) *E. coli* BL21 Z2-mCh T2

Figure 3.15: Histograms illustrating the distribution of the cellular length of WT *E. coli* BL21 and the recombinant strains A2, A2-mCh, Z2, and Z2-mCh. The cells were grown in Hi medium (Table 2.6) at 30°C until biomass reached OD_{600} 1, after which *m*-toluate (1 mM) was added to induce the recombinant strains, and the temperature was lowered to 25°C. The cellular lengths of 100 cells in two independent cultures per strain were measured at two time points in the cultivation, at OD_{600} 1 before induction (T1) and OD_{600} 2.5 after induction. The mean cellular length and standard deviation (SD) are shown in the upper right corner of the histograms. The cells were mounted onto agarose pads and phase contrast images were taken by a microscope. The strain and time point corresponding to the histograms are given in the subfigure captions.

Table 3.4 The mean cellular length and corresponding standard deviation (SD) of each strain before and after induction.

Strain	Mean ± SD Before induction	Mean ± SD After induction
WT	3.13 ± 1.88	3.15 ± 1.56
A2	3.23 ± 1.29	3.49 ± 3.37
A2-mCh	3.37 ± 1.18	3.60 ± 2.69
Z2	3.29 ± 1.04	3.37 ± 2.51
Z2-mCh	3.41 ± 1.99	4.03 ± 3.30

Paired t-tests were used to compare the strains' mean cellular length before induction with the mean cellular length after induction. The tests indicated that there was a significant difference in the mean cellular length of Z2-mCh before and after induction, as indicated with the P-value of 0.0248. The other strains had P-values above 0.05, meaning no significant difference was detected in mean cellular length before and after induction. One-way analysis of variance (One-way ANOVA) was used to compare the mean cellular length of all strains before induction and after induction, seperately. The test showed no significant differences in mean cellular length between the strains before induction with a P-value of 0.372. The test of the cellular lengths after induction resulted in a significant difference in the mean cellular length of at least two strains, indicated by a P-value of 0.026. Tukey's HSD identified that WT and Z2-mCh had significantly different mean cellular lengths after induction, with a P-value of 0.013133. All other strains compared had P-values over 0.05, thus indicating the statistical equal mean cellular length.

3.4 Control cultivations

Control cultivations were performed to test other parameters that may affect growth of bacterial cultures. It was tested if the potential toxicity of m-toluate affected the growth of *E. coli* BL21, as shown in Subsection 3.4.1. The effect of temperature on the growth rates of *E. coli* BL21 Z2 and Z2-mCh was studied, as shown in Subsection 3.4.2.

3.4.1 Test of inducer toxicity

Previous studies has observed that exposure to toxic compounds affects the growth of *E. coli* (Heipieper *et al.*, 1991; Zaldivar and Ingram, 1999). To test if the inducer *m*-toluate affects the growth of *E. coli*, WT *E. coli* BL21 where cultivated at a constant temperature with addition of *m*-toluate. The growth curve (Figure 3.16) shows that WT had relativly linear exponential growth throughout the cultivation. The growth rates were determined to be 0.47 h^{-1} both before and after addition of *m*-toluate, indicating that *m*-toluate did not affect the growth of WT *E. coli* BL21. The figure shows a drop in CO_2 flow around nine hours into the cultivation, which was caused by excess foaming.

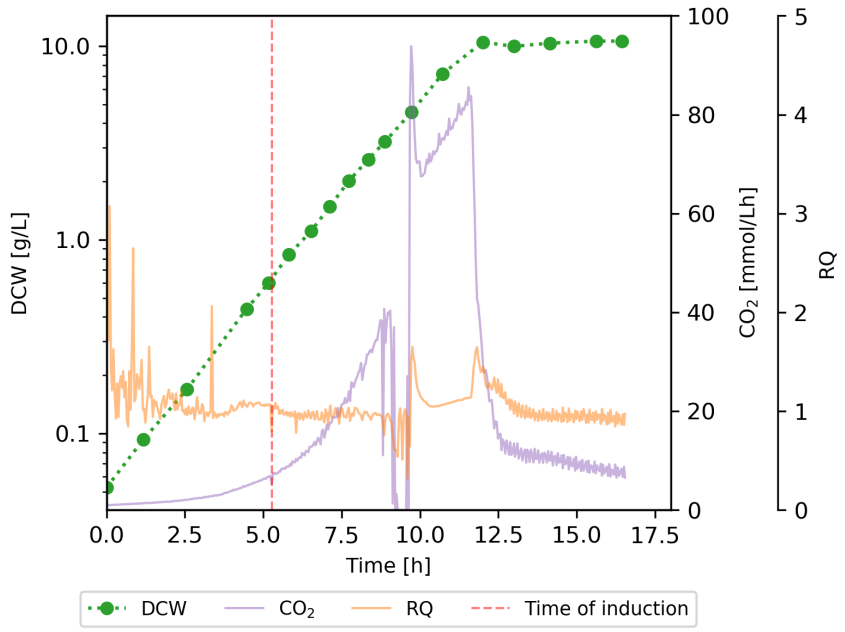


Figure 3.16: Growth curve obtained during batch cultivation of *E. coli* BL21 WT in a bench-scale bioreactor, together with CO_2 production and RQ measured throughout the cultivations. The strain was cultivated in Hf.1 medium (Table 2.8) at 30°C, with continuous adjustment of pH to 7 and DO to 40%. The inducer *m*-toluate (1 mM) was added when the biomass concentration reached 0.5 g DCW/L (~OD_{600} 1), indicated by a red dashed line. The drop in CO_2 flow observed around 9 h into the cultivation is caused by excess foaming.

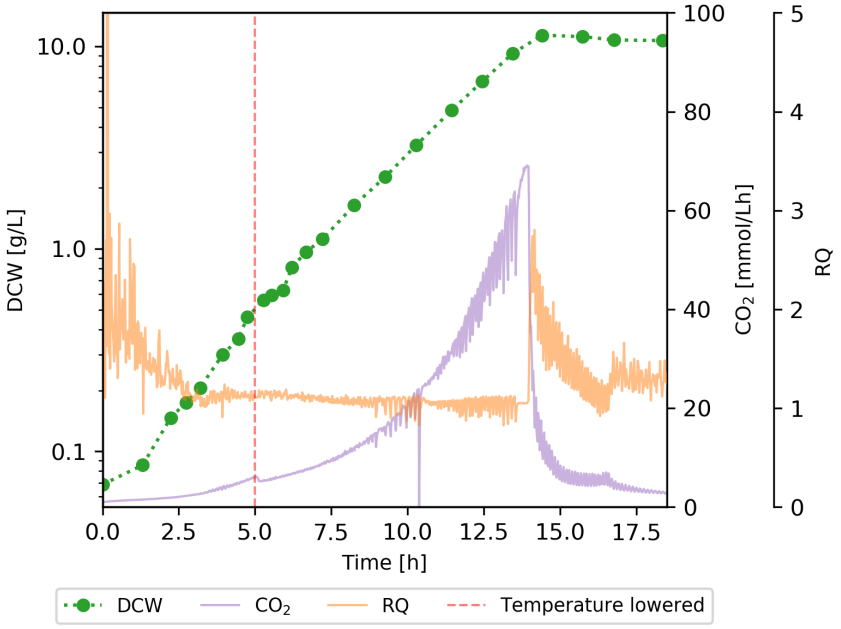
3.4.2 Growth of Z2 and Z2-mCh without addition of inducer

Previous cultivations showed that *E. coli* BL21 Z2 and Z2-mCh were two of the strains that were most severely affected by induction (Figures 3.1 and 3.5). A cultivation without addition of *m*-toluate, with lowering of temperature was performed to study the contribution of the lowered temperature on the reduced growth rates, and to determine the highest biomass concentration these strains could reach in batch cultivations.

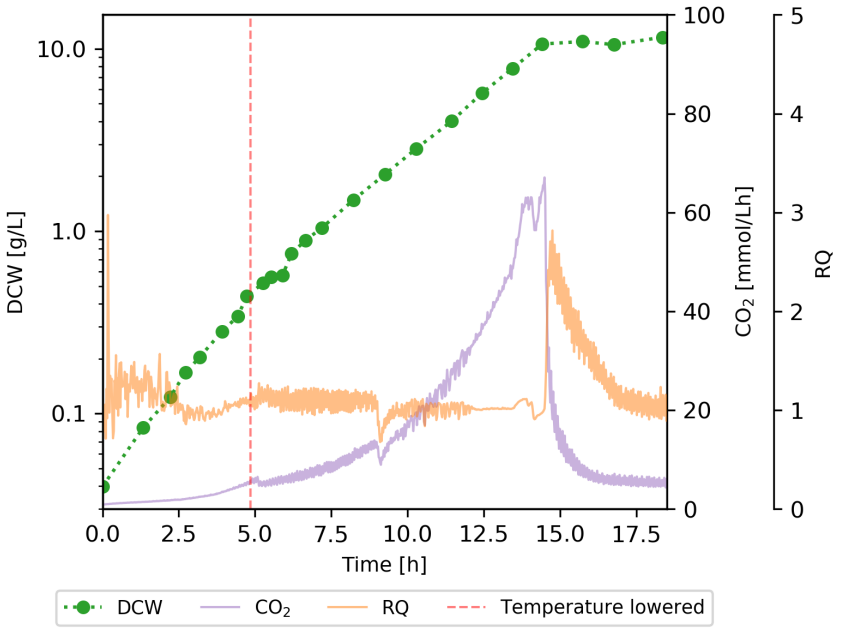
The results from the cultivations (Figure 3.17) are significantly different compared to the cultivations with addition of *m*-toluate (Figures 3.1 and 3.5). The growth rates after lowered temperature for both Z2 and Z2-mCh (Table 3.5) are within the range of the ones observed for the WT strain at 32 to 37 h^{-1} (Tables 3.2 and 3.3). Additionally, both Z2 and Z2-mCh reach significantly higher biomass concentrations at 11.6 g DCW/L, compared to the previously observed of Z2 at 9.1 g DCW/L and growth inhibition for Z2-mCh (Table 3.2).

Table 3.5 Summery of the growth curves shown in Figure 3.17. The table includes the highest biomass concentration [g CDW/L], and the strains growth rate (μ) before and after the temperature was lowered from 30°C to 25°C.

Strain	**Maximum OD_{600}**	**Growth rate, μ[h^{-1}] 30°C**	**25 °C**
Z2	11.6	0.46	0.34
Z2-mCh	11.6	0.45	0.33



(a) *E. coli* BL21 Z2.



(b) *E. coli* BL21 Z2-mCh.

Figure 3.17: Growth curve obtained during batch cultivation of *E. coli* BL21 Z2 (a) and Z2-mCh (b) in a bench-scale bioreactor, together with CO_2 production and RQ measured throughout the cultivations. The strains were cultivated in Hf.1 medium (Table 2.8) with ampicillin (0.1 mg/mL) at 30°C, with continuous adjustment of pH to 7 and DO to 40%. When the strains reached biomass concentration 0.5 g DCW/L (~OD_{600} 1), the temperature was lowered to 25°C, indicated by a red dashed line. The drops in CO_2 flow and RQ is caused by the system being opened to add antifoam.

Chapter 4

Discussion

The presence of expression vectors imposes a metabolic burden on the host (Glick, 1995; Soriano *et al.*, 1999). Plasmid maintenance and expression of plasmid-encoded genes drain the cell of its resources, leading to metabolic and physiological consequences. The severity of these consequences is inversely correlated with the degree of metabolic burden the expression vector imposes on the host cell (Glick, 1995; Hoffmann and Rinas, 2004). In other words, the severity of the consequences increases with factors including PCN, expression vector size, and the expression cassette strength (Glick, 1995; Hoffmann and Rinas, 2004; Zwick *et al.*, 2012). The physiological and metabolic consequences observed during the experiments performed in this thesis were decreased growth rates, reduced biomass yields, increased cell filamentation, adjustment of the central carbon metabolism, and increased overflow metabolism.

4.1 Reproducibility of results between cultivations

The bioreactor cultivations performed in this thesis had some small variations in growth rates between biological replica of the strains WT *E. coli* BL21 and Z2 (Tables 3.2 and 3.3). Some variations were expected as cultivation processes of microbes are prone to high batch-to-batch variations (Jenzsch *et al.*, 2006; Gnoth *et al.*, 2007). Small variations in cultivation conditions can lead to significant differences in biomass and product formation if an RP is produced. Factors such as inoculum size, which cannot be completely avoided, have been shown to lead to batch-to-batch variations (Gnoth *et al.*, 2007).

Three biological replicas of WT *E. coli* BL21 were analyzed with RP LC-MS/MS quantification of amino acids and organic acids to test the reproducibility of trends observed in metabolite pools between biological replicas. The WT strain was selected for this test as it does not contain any expression vector that can potentially lead to higher variability. Minor variations in the overall concentration of the quantified AAs and OAs were observed between the biological replicas (Figures 3.7 and 3.8). However, the relative distribution of the metabolites

was almost identical between the biological replicas, except for isocitrate (Figure 3.9). During the processing of the metabolic data, it was observed that the metabolite isocitrate low S/N, resulting in high variability between technical replicas. CapIC-MS/MS data can be used to quantify isocitrate, the metabolite should therefore be quantified and compared when the capIC-MS/MS analysis has been performed.

4.2 Physiological consequences observed in *E. coli* BL21 transformed with high-level expression vectors

4.2.1 Reduced growth rates

Reduction in growth rate is one of the most commonly observed consequences of expression vector presence. The metabolic burden makes recombinant cells grow slower than nontransformed cells, as housekeeping and growth-related genes have to compete over the cellular resources (Neubauer *et al.*, 2003). During the cultivations, it was observed that the recombinant strains had slower growth rates than the WT *E. coli* BL21 strains after induction (Figures 3.1 and 3.5) (Tables 3.2 and 3.3).

Additionally, the recombinant strains harboring expression vectors with a higher theoretical metabolic burden in the form of combinations of the stronger expression cassettes, higher PCN or the mCherry encoding gene, or both (Z2, Z1-mCh, and Z2-mCh) had slower growth rates compared to the other strains (A2, A2-mCh, and Z1). These observations coincide with the reported trend that a higher metabolic burden leads to more severe consequences in the host (Glick, 1995; Hoffmann and Rinas, 2004). During the growth characterization cultivations (Figure 3.1) (Table 3.2), it was observed that the growth rate of Z1 decreased with 38% after induction, A2 and A2-mCh decreased by approximately half, Z1-mCh and Z2 decreased with around 60%, and Z2-mCh entered complete growth inhibition. The significant difference in reaction to induction in A2 and A2-mCh compared to Z2 and Z2-mCh shows that the strength of the expression cassette has a significant impact on the metabolic burden the expression vector places on the cell. These strains have the same PCN of 20 and both a mCherry producing and control strain. Apart from this, no clear trends were observed in the other expression cassette factors, mCherry producing compared to control strain and PCN.

The difference in metabolic burden between the mCherry producing strains and their corresponding control strain lacking the mCherry encoding gene may depend on the strength of the expression vector. A2 and A2-mCh with the weaker WT A-type expression cassette have almost identical growth rates before and after induction, with both strains having approximately a 50% reduction in growth rate after induction. In contrast, the strains harboring expression vectors with the stronger mutated Z-type expression cassette had significant differences between the mCherry producing strain and the control strain after induction. Z1 is the recombinant strain that was the least affected by induction, with only a 38% reduction in

growth rate, while Z1-mCh's growth rate was reduced by 62%, which means the difference in reduction of growth rate after induction is around 24%. Z2's growth rate was reduced by 60% after induction, while Z2-mCh had complete growth inhibition. These observations indicate that the mCherry expression does not contribute to the metabolic burden when controlled under the weaker A-type expression cassette, but it does when controlled under the stronger Z-type expression cassette.

Li and Rinas (2020) had a similar observation during their study of growth rate reduction resulting from expression vector presence. They compared the growth rate of strains harboring expression vectors with two different proteins, the easily foldable green fluorescent protein (GFP) and human basic fibroblast growth factor (hFGF-2), which has a tendency to form inclusion bodies in *E. coli*. Each recombinant protein had a strain expressing the protein and a strain lacking ribosome binding site(s) and thus could not express the protein. They observed that the synthesis of recombinant proteins only added to the metabolic burden if there were folding regulated problems during the production (e.g. inclusion bodies). If the protein folded properly, the strains expressing the protein had similar growth rates to the control strains. Based on these results, they hypothesized that the metabolic burden observed by growth inhibition mainly results from elevated transcription or transcript levels and improper folding of proteins.

The mCherry producing strains with the strong Z-type expression cassette may be subjected to a cellular stress response upon induction. The expression cassette is strong, which may lead to expression rates faster than the capacity of the folding machinery of *E. coli*, resulting in the formation of inclusion bodies that induce a heat shock-like stress response, and thereby increase the metabolic burden. Accumulation of inclusion bodies can be tested using SDS-page by comparing the soluble and insoluble cell fraction of the mCherry producing strains. Comparison of transcription levels of genes associated with the heat-shock response using transcriptomics may also give an indication that a heat-shock-like stress response has been initiated. It is also possible that the high-level mCherry expression levels drained the producing cells of amino acids, resulting in the induction of a stringent-like stress response. This can be tested by comparing the transcription level of genes associated with the stringent stress response by transcriptomics. The results from the growth rate study may indicate that the transcription and level of transcripts or both are the main factor of the reduced growth rates observed unless the protein synthesis induces a stress response.

When comparing the strains based on PCN, an interesting trend is observed. Z1 (PCN 5) has a less reduced growth rate after induction compared to A2 (PCN 20) and A2-mCh (PCN 20), while Z1-mCh (PCN 5) has a significantly higher reduction (Tables 3.2 and 3.3). However, if the mCherry producing strains are not taken into account due to the potential induction of cellular stress responses in Z1-mCh after induction, as described above. Then the expression vector combination of PCN 5 and Z-type expression cassette (Z1) should place a less metabolic burden on the host compared to PCN (20) and A-type expression cassette (A2), based on the

reduction in growth rate after induction (Table 3.2).

The reduced growth rates observed after induction are not only caused by the addition of inducer *m*-toluate. Other factors, including the lowered temperature and potential toxicity of the inducer and expressed protein, may have negatively affected growth rate (Ratkowsky *et al.*, 1982; Zaldivar and Ingram, 1999). A control cultivation was performed to test the toxicity of *m*-toluate for *E. coli* (Subsection 3.4.1). WT *E. coli* BL21 were cultivated at a constant temperature with addition of *m*-toluate (1 mM). The results showed that the growth rate was unaffected by the addition of *m*-toluate, as it was determined to be 0.47 h^{-1} both before and after induction. This observation supports the previous reporting of *m*-toluate being nontoxic to *E. coli* (Gawin *et al.*, 2017). The presence of mCherry should not have affected the growth rate, as observed by the growth rates of A2 and A2-mCh. The strains had very similar growth rates, although one strain produces and accumulates mCherry intracellularly and the other strain does not.

Temperature is one of the key factors that influence the growth of bacteria (Ratkowsky *et al.*, 1982). The cultivation temperature was lowered from 30°C to 25°C upon induction with *m*-toluate. To test the effect the lowered temperature had on the growth rate, Z2 and Z2-mCh were cultivated without the addition of *m*-toluate, with lowering of the temperature (Subsection 3.4.2). Z2 and Z2-mCh were selected for the control cultivations as they are two of the strains that were the most severely affected by induction (Figures 3.1 and 3.5) (Tables 3.2 and 3.3). The control cultivation resulted in significantly higher growth rates after induction compared to when *m*-tolaute is added. Both strains had growth rates of at around 0.33 h^{-1} after induction, compared to around 0.20 h^{-1} and complete growth inhibition, which were observed for Z2 and Z2-mCh, respectively when *m*-toluate was added. A growth rate of 0.33 h^{-1} is within the range of growth rates observed for WT at 25°C at 0.32 to 0.37 h^{-1} (Tables 3.2 and 3.3). The similar growth rate of Z2 and Z2-mCh to WT indicates that the difference in growth rates after induction between the strains is caused by induction of the expression cassette, as *m*-toluate showed to be nontoxic, and the observation of Z2 and Z2-mCh having comparable growth rates as WT when the inducer is not added.

4.2.2 Reduced biomass formation

Another easily observable consequence of expression vector presence is reduced biomass formation (Glick, 1995; Hoffmann and Rinas, 2004; Rozkov *et al.*, 2004). During cultivation experiments, it was observed that Z2 and Z2-mCh reached lower biomass concentrations than the other strains (Figure 3.2). Z2 reached a biomass concentration of OD_{600} 9.08 g DCW/L and Z2-mCh had complete growth inhibition at 1.55 g DCW after induction, while the other strains reached biomass concentrations around 10 g DCW/L. In the control cultivations where Z2 and Z2-mCh were grown without addition of *m*-toluate, both strains reached biomass concentrations around 11.6 g DCW/L (Table 3.5). This may indicate that induction of the expression vectors led to lower biomass formation. It was surprising that Z2 had reduced

biomass yield, as the strain does not contain the mCherry encoding gene. However, it has previously been observed that strains containing plasmids only carrying resistance markers ("empty plasmid") can produce significantly lower biomass yields compared to plasmid-free cells (Rozkov *et al.*, 2004).

As previously described, Z2-mCh had complete growth inhibition after induction, followed by a decreasing trend in biomass concentration (Figures 3.1g and 3.5g). It is possible that Z2-mCh entered the "viable but nonculturable" state, where the cells exhibit metabolic activities but are unable to divide (Colwell *et al.*, 1985; Andersson *et al.*, 1996; Hoffmann and Rinas, 2004). *E. coli* has shown to enter this state when starved (Kaprelyants *et al.*, 1993; Rozkov *et al.*, 2004), but also under high-level RPP (Hoffmann and Rinas, 2004). A flow cytometry experiment performed by Soriano *et al.* (1999) on high-level RP expressing *E. coli* controlled by the strong T7 promoter showed that the cells had an immediate inability to proliferate after induction, and the slow growth in biomass concentration after was determined to be caused by increasing cell size and not by proliferation. In the growth curves of Z2-mCh, the biomass has a slight increase after induction, followed by a decreasing trend. According to the study performed by Soriano *et al.* (1999), it is possible that the cell proliferation was immediately inhibited after induction, and the slight increase in biomass is caused by increased cell size.

The decreasing biomass concentration of Z2-mCh after induction (Figures 3.1g and 3.5g) may be a result of cell lysis, as previous studies have observed that high-level RPP can lead to cell lysis (Newton *et al.*, 2016; Sun *et al.*, 2021). The extracellular protein concentration in supernatant culture broth samples was quantified using a Bradford protein assay (Figure 3.3) to test this hypothesis. Z2-mCh had a significantly higher extracellular protein concentration after induction compared to the other strains. On DCW basis, the concentration was almost six times as high as the other strains. Interestingly, Z2 also had a high extracellular protein concentration compared to WT, A2, A2-mCh, Z1, and Z1-mCh. The high extracellular protein concentrations indicate that cell lysis was a contributor to the low biomass formation observed for Z2 and the main contributor to the decreasing biomass of Z2-mCh after induction (Figures 3.1 and 3.5).

4.2.3 Filamentation of *E. coli*

High-level protein expression has shown to cause drastic morphological changes of *E. coli*, with increased cell filamentation (Jeong and Lee, 2003; Lederer *et al.*, 2011). During the cellular length comparison study, it was observed that all cells, including the WT, had occurrences of filamentous cells both before and after induction. It was not expected to observe such a high occurrence of filamentation before induction of recombinant strains, and especially not in the WT strain. A possible reason for the high occurrence of filaments may have been a result of the deficiency of proteases in *E. coli* BL21. *E. coli* BL21 is deficient in the proteases Lon and OmpT (Waegeman and Soetaert, 2011; Rosano and Ceccarelli, 2014), and it has pre-

viously been reported that inactivation of Lon can lead to filamentation of *E. coli* (Baneyx, 1999).

During the cellular length study, it was observed that Z2-mCh had higher mean cell length after induction compared to before induction (Figure 3.15). Filamentous cells as long as 170 µm were observed (Figure 3.14). Paired t-test concluded that Z2-mCh had a significantly longer mean cell length after induction than before, and the independent t-test concluded that Z2-mCh had a significantly longer mean cell length than WT after induction (Appendix G). The other recombinant strains seemed to have longer mean cell length and a higher occurrence of filamentous cells after induction than before (Figure 3.15), but the statistical tests concluded that the difference in mean cell length was insignificant (Appendix G). However, the length of more cells should be measured before a conclusion can be drawn. Only 200 cells were counted for each strain at each time point from two biological replicas. A higher number of cells of more biological replicas should be measured before any conclusions can be drawn. Another method that can be used to measure the cellular length is Flow cytometry. The technique allows for high-throughput, automated measurements of cell length and can measure the length of the larger population, which can give more reliable results.

4.3 Metabolic consequences observed in *E. coli* BL21 transformed with high level expression vectors

As previously mentioned, the presence of expression vectors places a metabolic burden on the host cell. The increased energy demand may cause the cell to adjust its metabolism (Hoffmann and Rinas, 2004). The aim of the intracellular metabolite quantification was to perform metabolic profiling analysis on the central carbon metabolism of *E.coli* harboring induced expression vectors. This includes quantification of metabolites in the glycolysis, pentose phosphate pathway (PPP), phosphorylated nucleosides, TCA cycle, and amino acids. The phosphorylated metabolites could not be quantified due to equipment malfunction. Hence, only amino acids, TCA cycle intermediates, lactate, and pyruvate could be quantified. The focus will therefore be on the changes on the amino acids and TCA cycle intermediates. It should be noted that quantification of metabolites alone does not give a full insight into the metabolism, as it does not give information of pathway activity. Metabolomics should, therefore, by paired with fluxomics to give further information on the metabolism pathway activity (Jang *et al.*, 2018). However, the quantified metabolite data should be sufficient to hypothesize trends based on observed concentration for further studies and to be tested with further biological replicas.

Although the focus will be on amino acids and TCA intermediates, it can be mentioned that previous fluxomics studies of recombinant producing *E. coli* has been reported to have increased flux through the glycolysis and glyoxylate shunt and decreased flux through the anabolic pathways, including PPP and gluconeogenesis (Özkan *et al.*, 2005; Wittmann *et al.*,

2007). Wittmann *et al.* (2007) also observed decreased levels of ATP and increased levels of ADP and AMP.

4.3.1 Selection of endometabolome sampling points

The endometabolome sampling points were decided based on the growth curves obtained during the growth characterization cultivations (Figure 3.1). The first sampling point was decided to be right before induction at 1.3 g DCW/L (OD_{600} ~ 2.5), to compare the metabolite distribution between the cells before induction. The second sampling point was decided to be around biomass concentration 3.6 g DCW/L (OD_{600} ~ 7), which was when A2-mCh started to produce mCherry (Figure 3.1c). The sampling point was selected to compare the metabolism between the strains at the early production phase. The last sampling point was decided to be around biomass concentration 7.5 g DCW/L (OD_{600} ~ 15), which was in the late exponential growth phase right before the stationary phase, to compare the metabolism at the late production phase.

4.3.2 Citric acid in the medium contaminated the quantification of intracellular citrate

Citric acid has many properties that make it favorable as a growth medium component, including high water solubility and good chelating and buffering properties (Behera *et al.*, 2021). Citric acid is a component in the Hf.1 medium (Table 2.8) used for the bioreactor cultivations. During the fast-filtration step of the endometabolome sampling, the growth medium was washed off with MQ-H_2O, and the citric acid should therefore not have been an issue. However, it was observed that all samples had an abnormally high intracellular concentration of citrate 3.6). Some samples had citrate concentrations over twice as high as the combined concentration of all other metabolites quantified. These abnormally high citrate concentrations contradict the findings of a study performed by Bennett *et al.* (2009), where they observed that several amino acids had significantly higher concentrations than citrate, in glucose fed *E. coli* in exponential growth (Bennett *et al.*, 2009).

Previous studies have shown that citric acid has the ability to bind proteins by ionic interactions in aqueous solutions at neutral pH (Xu *et al.*, 2015; Chanphai and Tajmir-Riahi, 2020, 2021). It is possible that medium citric acid bound to proteins on the cell surface, which were strong enough to withstand the washing step during endometabolome sampling. To prevent this from happening in further metabolome studies, a different defined medium should be used for the cultivations, or citric acid should be substituted in the Hf.1 medium. *E. coli* is incapable of utilizing citrate under aerobic conditions, as the strain lacks a functional citrate transporter in aerobic environments (Pos *et al.*, 1998). It is therefore unlikely that *E. coli* imported and metabolized citrate during the bioreactor cultivations.

4.3.3 Comparison of the intracellular concentrations of amino acids and TCA cycle intermediates between the strains analyzed

The quantified intracellular amino acid and TCA intermediate pools gave some interesting insight. A2, A2-mCh, and Z2 had very similar trends in amino acid and TCA cycle intermediate pools as the WT *E. coli* BL21 at all sampling points. The overall concentration of amino acids increased with cultivation time with all these strains (Figure 3.11a). This indicates that the strains adjusted to the lowering of temperature by increasing the intracellular amino acid pools, or that the protein synthesis rate was reduced, resulting in higher AA concentrations. The increase in AA from T2 to T3 may be a result of the cells already having started to adapt their metabolism towards the stationary phase. Strains might have stopped producing proteins and proliferation, resulting in higher intracellular AA levels. In contrast, the intracellular concentration of TCA intermediates remained relatively constant throughout the cultivation (Figure 3.11b), indicating that the TCA cycle was not much affected by the lowering of temperature. A2, A2-mCh, and Z2 also had almost identical relative concentrations of the intracellular metabolites quantified (Figure 3.12). This may indicate that the metabolic burden elicited by the expression vectors in A2, A2-mCh and Z2 were not severe enough to cause the host to reroute its central carbon metabolism.

Z2-mCh has similar trends to the other strains at sampling points T1 and T2. This is surprising as the sampling point of T2 was 30 min after induction for Z2-mCh, and 3 to 4 hours for the other strains. This may indicate that the strains used less than 30 min to adapt their metabolism to the lowered temperature. Another observation is that the growth inhibition Z2-mCh is subjected to may happen later than 30 min after induction. This makes it possible that Z2-mCh produced mCherry before growth the observed inhibition, and that the mCherry biosynthesis caused a cellular stress response in Z2-mCh. At sampling point T3, the amino acid and TCA intermediate pools of Z2-mCh were significantly different compared to the other strains. This will be further discussed in the subsequent subsections.

4.3.4 The concentration of amino acids in Z2-mCh were significantly reduced at the last sampling point

The concentration of amino acids was significantly lower in Z2-mCh than the other strains at sampling point T3 (Figure 3.11a). This may be a consequence of the high-level mCherry expression, as it was observed that Z2-mCh produced much greater amounts of mCherry on a g/g DCW basis compared to the other mCherry producing strains (Figure 3.2b). As previously mentioned, Z2-mCh had a complete growth inhibition around an hour after induction (Figures 3.1g and 3.5g). A possible explanation for the growth inhibition rapid consumption of amino acids for mCherry production in Z2-mCh lead to depleted amino acid levels resulting in induction of a stringent-like stress response. Previous studies have observed that high-level recombinant protein production led to up-and down-regulation of several genes similarly as during the stringent stress response (Haddadin and Harcum, 2005), and also accumulation of

the alarmone (p)ppGpp, which is the trigger that induces the stringent stress response (Andersson *et al.*, 1996; Cserjan-Puschmann *et al.*, 1999). These findings imply that the high rate utilization of amino acids during RPP can induce a stringent-like stress response. This may have been the cause of the growth inhibition of Z2-mCh observed after induction.

The amino acids that have a relatively less abundance in Z2-mCh compared to WT are mainly the ones synthesized from precursors in the TCA cycle (Figure 3.12). The amino acids glutamic acid (Glu) and arginine (Arg), which are synthesized from α-ketoglutarat (aKG), and asparagine (Asp) synthesized from oxaloacetate (OAA) are exceptions, as they have relatively similar concentrations in Z2-mCh compared to WT. Interestingly, Chang *et al.* (2002) observed that the genes in the biosynthetic pathway of arginine were upregulated in glucose depleted *E. coli*, which they had concluded to have initiated the stringent stress response. Haddadin and Harcum (2005) had a similar observation when they studied the transcriptome of high-level expressing *E. coli*, with the same genes in the biosynthesis pathway of arginine being upregulated in the RP producing *E. coli*. Another interesting observation is the relatively high concentration of asparagine compared to the other amino acids synthesized from oxaloacetate. Özkan *et al.* (2005) compared the fluxome of the central carbon metabolism between an induced recombinant protein producing *E. coli* with the uninduced strain. They observed that the induced strain had an increased flux from oxaloacetate to asparagine and that the flux through the asparagine branch was mainly redirected to the formation of fumarate, which reentered the TCA cycle. No clear correlation can be drawn when comparing the amino acid composition of mCherry (Appendix H) with the amino acids that had a relative decrease in Z2-mCh compared to WT (Figure 3.12). Glutamic acid, glycine, and lysine are the most abundant amino acids in mCherry. Of these, only lysine has a relative decrease in Z2-mCh compared to WT. It may be speculated that stress proteins expressed by a potential stress-response, as previously discussed, may have contributed to the observed depletion of amino acids.

4.3.5 Z2-mCh had elevated metabolites levels in the TCA cycle and increased overflow metabolism

The concentration of TCA cycle intermediates were significantly higher in Z2-mCh at T3 compared to the other strains (Figure 3.11b). The TCA cycle has two essential roles in the central carbon metabolism, in respiration the pathway performs complete oxidation of acetyl-CoA, additionally, it supplies intermediates for biosynthesis of several amino acids (Shiloach and Fass, 2005). It can be hypothesized that the increased levels of TCA metabolites could be a result of the increasing energy demand due to the expression of mCherry, as protein synthesis has been reported to be one of the most energy-demanding processes in the cell (Shiloach and Fass, 2005). Another possibility is that the accumulation may also be caused by the cell trying to produce more amino acids as they are utilized for mCherry expression.

Fluxomics studies performed on recombinant producing *E. coli* have had some contradicting observations for the changes of fluxes in the TCA cycle of recombinant expressing *E. coli*. Özkan *et al.* (2005) observed that TCA cycle fluxes were increased when in induced protein-expressing *E. coli* compared to uninduced. In other words, they observed that the fluxes in the TCA cycle were increased when the strains were placed under a higher degree of metabolic burden. In contrast, Wittmann *et al.* (2007) observed that the fluxes in the TCA cycle were significantly lowered after temperature-induced expression of recombinant proteins in *E. coli*. However, these differences may have been caused by the different induction systems. Özkan *et al.* (2005) used an IPTG inducible system, while Wittmann *et al.* (2007) used an inducible temperature promoter, where the induction was performed by increasing the temperature from 30 to 42°C.

The TCA cycle intermediates that had a higher relative concentration in Z2-mCh compared to WT at T3 were isocitrate, α-ketoglutarat, and succinate (Suc) (Figure 3.12). These metabolites are located at the beginning of the TCA cycle, right after the entry of acetyl-CoA (Shiloach and Fass, 2005). This may indicate that there is a bottleneck located at the beginning of the TCA cycle. Another result that strengthens this hypothesis is the results from the quantification of extracellular products during the growth characterization cultivations (Figure 3.4). It was observed that Z2-mCh secreted high amounts of the overflow metabolites acetic acid and lactic acid. The intracellular concentration of lactate, the conjugate base of lactic acid, was also observed to have a relative concentration in Z2-mCh, which was over four times as high as the concentration in WT (Figure 3.12). Shiloach and Fass (2005) have previously hypothesized that there is a bottleneck located around the pyruvate node, which causes overflow metabolism. It can be suspected that there may be another bottleneck located around α-ketoglutarat or succinate, as only the TCA intermediates isocitrate, α-ketoglutarat and succinate had relatively increased levels in Z2-mCh compared to WT (Figure 3.12).

4.4 Future work

The following recommendations will give more insight into the metabolic burden caused by expression vectors and the physiological and metabolic consequences the metabolic burden can cause.

- Perform the capIC-MS/MS analysis to quantify the remaining metabolites in the central carbon metabolism. This will give a more complete insight into how the metabolic burden affects the strain's central carbon metabolism.
- Citric acid should be removed form the bioreactor cultivation medium, as it contaminates the quantification of intracellular citrate levels. The medium should also be optimized as precipitation was observed before inoculation.
- The HPLC method used to quantify glucose should be optimized, as significant variations were observed between technical replicas. Alternatively, another method (e.g.

NMR) should be used to quantify glucose, so substrate consumption rate and substrate-based yields can be calculated and used to describe the metabolic burden.

- Combine metabolic profiling with fluxomics to give more information about the metabolic pathway activities in addition to the concentrations of metabolites. Also, include transcriptomics to study the transcription of genes associated with stress responses to see if the strains have induced cellular stress responses due to the metabolic burden of expression vector presence.
- Conduct metabolic profiling analysis on the Z1-mCh strain. It would be interesting to see how the central metabolism of Z1-mCh adapted to the metabolic burden caused by induction of the expression vector, as this strain had a higher metabolic burden than A2-mCh but not as high as Z2-mCh.
- Perform quality control studies and quantification of recombinant proteins (e.g. using gel-based approaches and western blotting)
- Conduct metabolic profiling on knock-out mutants where genes associated with the heat-shock, stringent, and SOS stress response have been inactivated. More insight into the contribution of stress responses to the metabolic burden may be obtained by comparing this metabolic profiling data with strains lacking knock-out mutations.
- Perform additional biological replicas of the experiments performed to validate the observations discussed in this thesis.

Chapter 5

Conclusions

The experimental data collected in this thesis demonstrated that the metabolic burden expression vectors placed on the cells lead to physiological and metabolic consequences. It was observed that the theoretical degree of metabolic burden implied with increased PCN, expression cassette strength, plasmid size, and expression of the reporter protein mCherry, was reflected in the severity of these consequences. The recombinant strains transformed with expression vectors with the combinations of the stronger expression cassette, higher PCN, or presence of the mCherry encoding gene, or both led to more significant decreases in growth rates after induction and increased secretion of overflow metabolites.

The difference in metabolic burden between the mCherry expressing strain and the control strain lacking the mCherry encoding gene was found to be dependent on the strength of the expression cassette. The mCherry producing strain and control strain containing the weaker expression cassette were found to have almost identical growth rates to each other before and after induction. In contrast, mCherry producing strain and control strain containing the stronger expression cassette had significantly different growth rates after induction. These observations may indicate that the protein synthesis does not contribute to the metabolic burden unless the protein synthesis or expressed protein induces a cellular stress response.

All recombinant strains had physiological consequences of metabolic burden reflected in their reduced growth rate. The strain harboring the expression vector with the strong expression cassette, highest PCN, and the mCherry encoding gene (Z2-mCh) was observed to have complete growth inhibition after induction, which may have resulted from a cellular stress response. The strain was also observed to have a decreasing trend in biomass concentration after induction, potentially caused by cell lysis due to the metabolic burden imposed by the expression vector. Additionally, the strain was also observed to have an increased degree of cell filamentation after induction of mCherry expression.

Of the recombinant strains analyzed with RP LC-MS/MS-based quantification of the intracellular levels of amino acids and TCA cycle intermediates, only Z2-mCh was observed to have significantly different intracellular concentrations from WT *E. coli* BL21. Z2-mCh was observed to have a significantly lower amino acid pool compared to the other strains and significantly higher levels of TCA cycle intermediates. The low amino acid pool may suggest that the amino acids were rapidly utilized in protein synthesis. The increased levels of TCA intermediates may indicate that Z2-mCh adjusted its metabolism towards energy-generating pathways due to the increasing energy demand after the induction of mCherry expression. The strain was observed to have a high extracellular concentration of the overflow metabolites acetic and lactic acid after induction, suggesting that the TCA cycle is a limiting step in respiration.

Bibliography

Abdi, H. and Williams, L. J. (2010) 'Principal component analysis', *Wiley interdisciplinary reviews: computational statistics*, 2(4), pp. 433–459.

Aertsen, A. and Michiels, C. W. (2005) 'Mrr instigates the sos response after high pressure stress in escherichia coli', *Molecular microbiology*, 58(5), pp. 1381–1391.

Andersson, L., Yang, S., Neubauer, P. and Enfors, S.-o. (1996) 'Impact of plasmid presence and induction on cellular responses in fed batch cultures of escherichia coli', *Journal of biotechnology*, 46(3), pp. 255–263.

Arsène, F., Tomoyasu, T. and Bukau, B. (2000) 'The heat shock response of escherichia coli', *International journal of food microbiology*, 55(1-3), pp. 3–9.

Balzer, S., Kucharova, V., Megerle, J., Lale, R., Brautaset, T. and Valla, S. (2013) 'A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in escherichia coli', *Microbial cell factories*, 12(1), pp. 1–14.

Baneyx, F. (1999) 'Recombinant protein expression in escherichia coli', *Current opinion in biotechnology*, 10(5), pp. 411–421.

Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J. R. and Hwa, T. (2015) 'Overflow metabolism in escherichia coli results from efficient proteome allocation', *Nature*, 528(7580), pp. 99–104.

Bedre, R. (2020), 'bioinfokit: Bioinformatics data analysis and visualization toolkit.'. Available from: http://doi.org/10.5281/zenodo.3698145.

Behera, B. C., Mishra, R. and Mohapatra, S. (2021) 'Microbial citric acid: Production, properties, application, and future perspectives', *Food Frontiers*, 2(1), pp. 62–76.

Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J. and Rabinowitz, J. D. (2009) 'Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli', *Nature chemical biology*, 5(8), pp. 593–599.

Blasina, A., Kittell, B. L., Toukdarian, A. E. and Helinski, D. R. (1996) 'Copy-up mutants of the plasmid rk2 replication initiation protein are defective in coupling rk2 replication origins', *Proceedings of the National Academy of Sciences*, 93(8), pp. 3559–3564.

Blatny, J. M., Brautaset, T., Winther-Larsen, H. C., Haugan, K. and Valla, S. (1997) 'Construction and use of a versatile set of broad-host-range cloning and expression vectors based on the rk2 replicon', *Applied and Environmental Microbiology*, 63(2), pp. 370–379.

Bradford, M. M. (1976) 'A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding', *Analytical biochemistry*, 72(1-2), pp. 248–254.

Brown, A. M. (2005) 'A new software for carrying out one-way anova post hoc tests', *Computer methods and programs in biomedicine*, 79(1), pp. 89–95.

Carneiro, S., Ferreira, E. C. and Rocha, I. (2013) 'Metabolic responses to recombinant bioprocesses in escherichia coli', *Journal of biotechnology*, 164(3), pp. 396–408.

Chang, D.-E., Smalley, D. J. and Conway, T. (2002) 'Gene expression profiling of escherichia coli growth transitions: an expanded stringent response model', *Molecular microbiology*, 45(2), pp. 289–306.

Chanphai, P. and Tajmir-Riahi, H. (2020) 'Conjugation of citric acid and gallic acid with serum albumins: Acid binding sites and protein conformation', *Journal of Molecular Liquids*, 299, pp. 112178.

Chanphai, P. and Tajmir-Riahi, H. (2021) 'Locating the binding sites of citric acid and gallic acid on milk β-lactoglobulin', *Journal of Biomolecular Structure and Dynamics*, 39(14), pp. 5160–5165.

Colwell, R., Brayton, P., Grimes, D., Roszak, D., Huq, S. and Palmer, L. (1985) 'Viable but non-culturable vibrio cholerae and related pathogens in the environment: implications for release of genetically engineered microorganisms', *Bio/technology*, 3(9), pp. 817–820.

Cserjan-Puschmann, M., Kramer, W., Duerrschmid, E., Striedner, G. and Bayer, K. (1999) 'Metabolic approaches for the optimisation of recombinant fermentation processes', *Applied microbiology and biotechnology*, 53(1), pp. 43–50.

Davidson, M. W. and Campbell, R. E. (2009) 'Engineered fluorescent proteins: innovations and applications', *Nature methods*, 6(10), pp. 713–717.

Day, R. N. and Davidson, M. W. (2009) 'The fluorescent protein palette: tools for cellular imaging', *Chemical Society Reviews*, 38(10), pp. 2887–2921.

Dean, R. B. and Dixon, W. J. (1951) 'Simplified statistics for small numbers of observations', *Analytical chemistry*, 23(4), pp. 636–638.

Durland, R. H., Toukdarian, A., Fang, F. and Helinski, D. (1990) 'Mutations in the trfa replication gene of the broad-host-range plasmid rk2 result in elevated plasmid copy numbers', *Journal of bacteriology*, 172(7), pp. 3859–3867.

Efstathiou, C. E. (2006) 'Estimation of type i error probability from experimental dixon's "q" parameter on testing for outliers within small size data sets', *Talanta*, 69(5), pp. 1068–1071.

Fang, F. C. and Helinski, D. R. (1991) 'Broad-host-range properties of plasmid rk2: importance of overlapping genes encoding the plasmid replication initiation protein trfa', *Journal of bacteriology*, 173(18), pp. 5861–5868.

Feizollahzadeh, S., Kouhpayeh, S., Rahimmansh, I., Khanahmad, H., Sabzehei, F., Ganjalikhani-Hakemi, M., Andalib, A., Hejazi, Z. and Rezaei, A. (2017) 'The increase in protein and plasmid yields of e. coli with optimized concentration of ampicillin as selection marker', *Iranian journal of biotechnology*, 15(2), pp. 128.

Ferullo, D. J. and Lovett, S. T. (2008) 'The stringent response and cell cycle arrest in escherichia coli', *PLoS genetics*, 4(12), pp. e1000300.

Francis, D. M. and Page, R. (2010) 'Strategies to optimize protein expression in e. coli', *Current protocols in protein science*, 61(1), pp. 5–24.

Ganten, D., Ruckpaul, K., Birchmeier, W., Epplen, J. T., Genser, K. and Gossen, M. (2006) *Recombinant Protein. In: Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, Springer, Berlin, Heidelberg, pp. 1609–1609.
Available from: https://doi.org/10.1007/3-540-29623-9_8485

Gawin, A., Valla, S. and Brautaset, T. (2017) 'The xyls/pm regulator/promoter system and its use in fundamental studies of bacterial gene expression, recombinant protein production and metabolic engineering', *Microbial biotechnology*, 10(4), pp. 702–718.

Gil, A., Siegel, D., Permentier, H., Reijngoud, D.-J., Dekker, F. and Bischoff, R. (2015) 'Stability of energy metabolites—an often overlooked issue in metabolomics studies: A review', *Electrophoresis*, 36(18), pp. 2156–2169.

Glick, B. R. (1995) 'Metabolic load and heterologous gene expression', *Biotechnology advances*, 13(2), pp. 247–261.

Gnoth, S., Jenzsch, M., Simutis, R. and Lübbert, A. (2007) 'Process analytical technology (pat): Batch-to-batch reproducibility of fermentation processes by robust process operational design and control', *Journal of Biotechnology*, 132(2), pp. 180–186.

Greated, A., Lambertsen, L., Williams, P. A. and Thomas, C. M. (2002) 'Complete sequence of the incp-9 tol plasmid pww0 from pseudomonas putida', *Environmental Microbiology*, 4(12), pp. 856–871.

Guisbert, E., Yura, T., Rhodius, V. A. and Gross, C. A. (2008) 'Convergence of molecular, modeling, and systems approaches for an understanding of the escherichia coli heat shock response', *Microbiology and Molecular Biology Reviews*, 72(3), pp. 545–554.

Gupta, S. K. and Shukla, P. (2016) 'Advanced technologies for improved expression of recombi-

nant proteins in bacteria: perspectives and applications', *Critical reviews in biotechnology*, 36(6), pp. 1089–1098.

Haddadin, F. T. and Harcum, S. W. (2005) 'Transcriptome profiles for high-cell-density recombinant and wild-type escherichia coli', *Biotechnology and bioengineering*, 90(2), pp. 127–153.

Hanahan, D. (1983) 'Studies on transformation of escherichia coli with plasmids', *Journal of molecular biology*, 166(4), pp. 557–580.

Harcum, S. W. and Bentley, W. E. (1999) 'Heat-shock and stringent responses have overlapping protease activity in escherichia coli', *Applied biochemistry and biotechnology*, 80(1), pp. 23–37.

Heipieper, H.-J., Keweloh, H. and Rehm, H.-J. (1991) 'Influence of phenols on growth and membrane permeability of free and immobilized escherichia coli', *Applied and Environmental Microbiology*, 57(4), pp. 1213–1217.

Heppert, J. K., Dickinson, D. J., Pani, A. M., Higgins, C. D., Steward, A., Ahringer, J., Kuhn, J. R. and Goldstein, B. (2016) 'Comparative assessment of fluorescent proteins for in vivo imaging in an animal model system', *Molecular biology of the cell*, 27(22), pp. 3385–3394.

Hoffmann, F. and Rinas, U. (2004) 'Stress induced by recombinant protein production in escherichia coli', *Physiological stress responses in bioprocesses*, pp. 73–92.

Hotelling, H. (1933) 'Analysis of a complex of statistical variables into principal components.', *Journal of educational psychology*, 24(6), pp. 417.

Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. and Boyer, H. W. (1977) 'Expression in escherichia coli of a chemically synthesized gene for the hormone somatostatin', *Science*, 198(4321), pp. 1056–1063.

Jaimes-Lizcano, Y. A., Hunn, D. D. and Papadopoulos, K. D. (2014) 'Filamentous escherichia coli cells swimming in tapered microcapillaries', *Research in microbiology*, 165(3), pp. 166–174.

Jang, C., Chen, L. and Rabinowitz, J. D. (2018) 'Metabolomics and isotope tracing', *Cell*, 173(4), pp. 822–837.

Jenzsch, M., Simutis, R. and Lübbert, A. (2006) 'Optimization and control of industrial microbial cultivation processes', *Engineering in Life Sciences*, 6(2), pp. 117–124.

Jeong, H., Kim, H. J. and Lee, S. J. (2015) 'Complete genome sequence of escherichia coli strain bl21', *Genome announcements*, 3(2), pp. e00134–15.

Jeong, K. J. and Lee, S. Y. (2003) 'Enhanced production of recombinant proteins in escherichia coli by filamentation suppression', *Applied and environmental microbiology*, 69(2), pp. 1295–1298.

Johnson, I. S. (1983) 'Human insulin from recombinant dna technology', *Science*, 219(4585), pp. 632–637.

Kaprelyants, A. S., Gottschal, J. C. and Kell, D. B. (1993) 'Dormancy in non-sporulating bacteria', *FEMS microbiology letters*, 104(3-4), pp. 271–285.

Kaur, J., Kumar, A. and Kaur, J. (2018) 'Strategies for optimization of heterologous protein expression in e. coli: Roadblocks and reinforcements', *International Journal of Biological Macromolecules*, 106, pp. 803–822.

Kim, T. K. (2015) 'T test as a parametric statistic', *Korean journal of anesthesiology*, 68(6), pp. 540.

Kirk, O., Borchert, T. V. and Fuglsang, C. C. (2002) 'Industrial enzyme applications', *Current opinion in biotechnology*, 13(4), pp. 345–351.

Kost, T. A., Condreay, J. P. and Jarvis, D. L. (2005) 'Baculovirus as versatile vectors for protein expression in insect and mammalian cells', *Nature biotechnology*, 23(5), pp. 567–575.

Kumar, J., Chauhan, A. S., Shah, R. L., Gupta, J. A. and Rathore, A. S. (2020) 'Amino acid supplementation for enhancing recombinant protein production in e. coli', *Biotechnology and Bioengineering*, 117(8), pp. 2420–2433.

Kusnadi, A. R., Nikolov, Z. L. and Howard, J. A. (1997) 'Production of recombinant proteins in transgenic plants: practical considerations', *Biotechnology and bioengineering*, 56(5), pp. 473–484.

Kvitvang, H. F. and Bruheim, P. (2015) 'Fast filtration sampling protocol for mammalian suspension cells tailored for phosphometabolome profiling by capillary ion chromatography–tandem mass spectrometry', *Journal of Chromatography B*, 998, pp. 45–49.

Kvitvang, H. F., Kristiansen, K. A. and Bruheim, P. (2014) 'Assessment of capillary anion exchange ion chromatography tandem mass spectrometry for the quantitative profiling of the phosphometabolome and organic acids in biological extracts', *Journal of Chromatography A*, 1370, pp. 70–79.

Lagendijk, E. L., Validov, S., Lamers, G. E., De Weert, S. and Bloemberg, G. V. (2010) 'Genetic tools for tagging gram-negative bacteria with mcherry for visualization in vitro and in natural habitats, biofilm and pathogenicity studies', *FEMS microbiology letters*, 305(1), pp. 81–90.

Larson, M. G. (2008) 'Analysis of variance', *Circulation*, 117(1), pp. 115–121.

Lederer, F. L., Günther, T. J., Raff, J. and Pollmann, K. (2011) 'E. coli filament formation induced by heterologous s-layer expression', *Bioengineered Bugs*, 2(3), pp. 178–181.

Li, Z. and Rinas, U. (2020) 'Recombinant protein production associated growth inhibition results mainly from transcription and not from translation', *Microbial cell factories*, 19(1), pp. 1–11.

Majchrzak, M., Bowater, R. P., Staczek, P. and Parniewski, P. (2006) 'Sos repair and dna supercoiling influence the genetic stability of dna triplet repeats in escherichia coli', *Journal of molecular biology*, 364(4), pp. 612–624.

Markets and Markets (2020), 'Protein expression market by type (escherichia coli, mammalian, yeast, pichia, insect, baculovirus and cell-free), products (reagents, competent cells, instruments, services), application, end-user and region - global forecast to 2025', `https://www.marketsandmarkets.com/Market-Reports/protein-expression-market-180323924.html`. Fetched 09.01.2022.

Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M. and Porro, D. (2012) 'Recombinant protein production in yeasts', *Recombinant gene expression*, pp. 329–358.

Meagher, R. B., Tait, R. C., Betlach, M. and Boyer, H. W. (1977) 'Protein expression in e. coli minicells by recombinant plasmids', *Cell*, 10(3), pp. 521–536.

Miki, B. and McHugh, S. (2004) 'Selectable marker genes in transgenic plants: applications, alternatives and biosafety', *Journal of biotechnology*, 107(3), pp. 193–232.

Mishra, P., Singh, U., Pandey, C. M., Mishra, P. and Pandey, G. (2019) 'Application of student's t-test, analysis of variance, and covariance', *Annals of cardiac anaesthesia*, 22(4), pp. 407.

Nakano, K., Rischke, M., Sato, S. and Märkl, H. (1997) 'Influence of acetic acid on the growth of escherichia coli k12 during high-cell-density cultivation in a dialysis reactor', *Applied microbiology and biotechnology*, 48(5), pp. 597–601.

Naylor, L. H. (1999) 'Reporter gene technology: the future looks bright', *Biochemical pharmacology*, 58(5), pp. 749–757.

Neubauer, P., Lin, H. and Mathiszik, B. (2003) 'Metabolic load of recombinant protein production: inhibition of cellular capacities for glucose uptake and respiration after induction of a heterologous gene in escherichia coli', *Biotechnology and bioengineering*, 83(1), pp. 53–64.

Newton, J. M., Schofield, D., Vlahopoulou, J. and Zhou, Y. (2016) 'Detecting cell lysis using viscosity monitoring in e. coli fermentation to prevent product loss', *Biotechnology progress*, 32(4), pp. 1069–1076.

Nielsen, J. (2001) 'Metabolic engineering', *Applied microbiology and biotechnology*, 55(3), pp. 263–283.

Overton, T. W. (2014) 'Recombinant protein production in bacterial hosts', *Drug discovery today*, 19(5), pp. 590–601.

Ow, D. S.-W., Nissom, P. M., Philp, R., Oh, S. K.-W. and Yap, M. G.-S. (2006) 'Global transcriptional analysis of metabolic burden due to plasmid maintenance in escherichia coli dh5α during batch fermentation', *Enzyme and Microbial Technology*, 39(3), pp. 391–398.

Özkan, P., Sariyar, B., Ütkür, F. Ö., Akman, U. and Hortaçsu, A. (2005) 'Metabolic flux analysis of

recombinant protein overproduction in escherichia coli', *Biochemical engineering journal*, 22(2), pp. 167–195.

Palomares, L. A., Estrada-Moncada, S. and Ramírez, O. T. (2004) 'Production of recombinant proteins', *Recombinant gene expression*, pp. 15–51.

Paulsson, J. and Ehrenberg, M. (2001) 'Noise in a minimal regulatory network: plasmid copy number control', *Quarterly reviews of biophysics*, 34(1), pp. 1–59.

Pearson, K. (1901) 'Liii. on lines and planes of closest fit to systems of points in space', *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11), pp. 559–572.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011) 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research*, 12, pp. 2825–2830.

Plotka, M., Wozniak, M. and Kaczorowski, T. (2017) 'Quantification of plasmid copy number with single colour droplet digital pcr', *PLoS One*, 12(1), pp. e0169846.

Pos, K. M., Dimroth, P. and Bott, M. (1998) 'The escherichia coli citrate carrier citt: a member of a novel eubacterial transporter family related to the 2-oxoglutarate/malate translocator from spinach chloroplasts', *Journal of Bacteriology*, 180(16), pp. 4160–4165.

Puetz, J. and Wurm, F. M. (2019) 'Recombinant proteins for industrial versus pharmaceutical purposes: a review of process and pricing', *Processes*, 7(8), pp. 476.

Punt, P. J., van Biezen, N., Conesa, A., Albers, A., Mangnus, J. and van den Hondel, C. (2002) 'Filamentous fungi as cell factories for heterologous protein production', *Trends in biotechnology*, 20(5), pp. 200–206.

Rafailidis, P. I., Ioannidou, E. N. and Falagas, M. E. (2007) 'Ampicillin/sulbactam', *Drugs*, 67(13), pp. 1829–1849.

Ransom, E. M., Ellermeier, C. D. and Weiss, D. S. (2015) 'Use of mcherry red fluorescent protein for studies of protein localization and gene expression in clostridium difficile', *Applied and environmental microbiology*, 81(5), pp. 1652–1660.

Ratkowsky, D. A., Olley, J., McMeekin, T. and Ball, A. (1982) 'Relationship between temperature and growth rate of bacterial cultures', *Journal of bacteriology*, 149(1), pp. 1–5.

Rorabacher, D. B. (1991) 'Statistical treatment for rejection of deviant values: critical values of dixon's" q" parameter and related subrange ratios at the 95% confidence level', *Analytical Chemistry*, 63(2), pp. 139–146.

Rosano, G. L. and Ceccarelli, E. A. (2014) 'Recombinant protein expression in escherichia coli: advances and challenges', *Frontiers in microbiology*, 5, pp. 172.

Rosano, G. L., Morales, E. S. and Ceccarelli, E. A. (2019) 'New tools for recombinant protein production in escherichia coli: A 5-year update', *Protein science*, 28(8), pp. 1412–1422.

Røst, L. M., Brekke Thorfinnsdottir, L., Kumar, K., Fuchino, K., Eide Langørgen, I., Bartosova, Z., Kristiansen, K. A. and Bruheim, P. (2020) 'Absolute quantification of the central carbon metabolome in eight commonly applied prokaryotic and eukaryotic model systems', *Metabolites*, 10(2), pp. 74.

Rozkov, A., Avignone-Rossa, C., Ertl, P., Jones, P., O'Kennedy, R., Smith, J., Dale, J. and Bushell, M. (2004) 'Characterization of the metabolic burden on escherichia coli dh1 cells imposed by the presence of a plasmid containing a gene therapy sequence', *Biotechnology and bioengineering*, 88(7), pp. 909–915.

Sezonov, G., Joseleau-Petit, D. and d'Ari, R. (2007) 'Escherichia coli physiology in luria-bertani broth', *Journal of bacteriology*, 189(23), pp. 8746–8749.

Shaner, N. C., Steinbach, P. A. and Tsien, R. Y. (2005) 'A guide to choosing fluorescent proteins', *Nature methods*, 2(12), pp. 905–909.

Shiloach, J. and Fass, R. (2005) 'Growing e. coli to high cell density—a historical perspective on method development', *Biotechnology advances*, 23(5), pp. 345–357.

Shlens, J. (2014) 'A tutorial on principal component analysis', *arXiv preprint arXiv:1404.1100*, .

Silva, F., Queiroz, J. A. and Domingues, F. C. (2012) 'Evaluating metabolic stress and plasmid stability in plasmid dna production by escherichia coli', *Biotechnology advances*, 30(3), pp. 691–708.

Singha, T. K., Gulati, P., Mohanty, A., Khasa, Y. P., Kapoor, R. K. and Kumar, S. (2017) 'Efficient genetic approaches for improvement of plasmid based expression of recombinant protein in escherichia coli: a review', *Process Biochemistry*, 55, pp. 17–31.

Skinner, S. O., Sepúlveda, L. A., Xu, H. and Golding, I. (2013) 'Measuring mrna copy number in individual escherichia coli cells using single-molecule fluorescent in situ hybridization', *Nature protocols*, 8(6), pp. 1100–1113.

Sletta, H., Nedal, A., Aune, T., Hellebust, H., Hakvag, S., Aune, R., Ellingsen, T., Valla, S. and Brautaset, T. (2004) 'Broad-host-range plasmid pjb658 can be used for industrial-level production of a secreted host-toxic single-chain antibody fragment in escherichia coli', *Applied and environmental microbiology*, 70(12), pp. 7033–7039.

Sletta, H., Tøndervik, A., Hakvåg, S., Aune, T. V., Nedal, A., Aune, R., Evensen, G., Valla, S., Ellingsen, T. and Brautaset, T. (2007) 'The presence of n-terminal secretion signal sequences leads to strong stimulation of the total expression levels of three tested medically important proteins during high-cell-density cultivations of escherichia coli', *Applied and environmental microbiology*, 73(3), pp. 906–912.

Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. and de la Cruz, F. (2010) 'Mobility of plasmids', *Microbiology and Molecular Biology Reviews*, 74(3), pp. 434–452.

Soriano, E., Borth, N., Katinger, H. and Mattanovich, D. (1999) 'Flow cytometric analysis of metabolic stress effects due to recombinant plasmids and proteins in escherichia coli production strains', *Metabolic engineering*, 1(3), pp. 270–274.

Stafsnes, M. H., Røst, L. M. and Bruheim, P. (2018) 'Improved phosphometabolome profiling applying isotope dilution strategy and capillary ion chromatography-tandem mass spectrometry', *Journal of Chromatography B*, 1083, pp. 278–283.

Sun, X.-M., Zhang, Z.-X., Wang, L.-R., Wang, J.-G., Liang, Y., Yang, H.-F., Tao, R.-S., Jiang, Y., Yang, J.-J. and Yang, S. (2021) 'Downregulation of t7 rna polymerase transcription enhances pet-based recombinant protein production in escherichia coli bl21 (de3) by suppressing autolysis', *Biotechnology and Bioengineering*, 118(1), pp. 153–163.

Szenk, M., Dill, K. A. and de Graff, A. M. (2017) 'Why do fast-growing bacteria enter overflow metabolism? testing the membrane real estate hypothesis', *Cell systems*, 5(2), pp. 95–104.

Tan, B., Lu, Z., Dong, S., Zhao, G. and Kuo, M.-S. (2014) 'Derivatization of the tricarboxylic acid intermediates with o-benzylhydroxylamine for liquid chromatography–tandem mass spectrometry detection', *Analytical biochemistry*, 465, pp. 134–147.

Tenaillon, O., Skurnik, D., Picard, B. and Denamur, E. (2010) 'The population genetics of commensal escherichia coli', *Nature reviews microbiology*, 8(3), pp. 207–217.

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.

Thakker, C., Martínez, I., San, K.-Y. and Bennett, G. N. (2012) 'Succinate production in escherichia coli', *Biotechnology journal*, 7(2), pp. 213–224.

Thiese, M. S., Ronna, B. and Ott, U. (2016) 'P value interpretations and considerations', *Journal of thoracic disease*, 8(9), pp. E928.

Toukdarian, A. E. and Helinski, D. R. (1998) 'Trfa dimers play a role in copy-number control of rk2 replication', *Gene*, 223(1-2), pp. 205–211.

Traxler, M. F., Summers, S. M., Nguyen, H.-T., Zacharia, V. M., Hightower, G. A., Smith, J. T. and Conway, T. (2008) 'The global, ppgpp-mediated stringent response to amino acid starvation in escherichia coli', *Molecular microbiology*, 68(5), pp. 1128–1148.

Tripathi, N. K. and Shrivastava, A. (2019) 'Recent developments in bioprocessing of recombinant proteins: expression hosts and process development', *Frontiers in bioengineering and biotechnology*, 7, pp. 420.

Vera, A., González-Montalbán, N., Arís, A. and Villaverde, A. (2007) 'The conformational qual-

ity of insoluble recombinant proteins is enhanced at low growth temperatures', *Biotechnology and bioengineering*, 96(6), pp. 1101–1106.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P. and SciPy 1.0 Contributors (2020) 'SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python', *Nature Methods*, 17, pp. 261–272.

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.

Walsh, G. (2018) 'Biopharmaceutical benchmarks 2018', *Nature biotechnology*, 36, pp. 1136–1145.

Winther-Larsen, H. C., Josefsen, K. D., Brautaset, T. and Valla, S. (2000) 'Parameters affecting gene expression from the pm promoter in gram-negative bacteria', *Metabolic engineering*, 2(2), pp. 79–91.

Wittmann, C., Weber, J., Betiku, E., Krömer, J., Böhm, D. and Rinas, U. (2007) 'Response of fluxome and metabolome to temperature-induced recombinant protein synthesis in escherichia coli', *Journal of biotechnology*, 132(4), pp. 375–384.

Wurm, F. M. (2004) 'Production of recombinant protein therapeutics in cultivated mammalian cells', *Nature biotechnology*, 22(11), pp. 1393–1398.

Wyre, C. J. (2015), Recombinant protein production in Escherichia coli: optimisation of improved protocols, PhD thesis, University of Birmingham.

Xu, H., Shen, L., Xu, L. and Yang, Y. (2015) 'Low-temperature crosslinking of proteins using non-toxic citric acid in neutral aqueous medium: Mechanism and kinetic study', *Industrial Crops and Products*, 74, pp. 234–240.

Zaldivar, J. and Ingram, L. O. (1999) 'Effect of organic acids on the growth and fermentation of ethanologenic escherichia coli ly01', *Biotechnology and bioengineering*, 66(4), pp. 203–210.

Zwick, F., Lale, R. and Valla, S. (2012) 'Strong stimulation of recombinant protein production in escherichia coli by combining stimulatory control elements in an expression cassette', *Microbial cell factories*, 11(1), pp. 1–8.

Appendix A

Metabolite Abbreviations

Table A.1 Metabolic abbreviations together with their full names. The metabolites are divided into the classes Glycolysis, PPP, TCA cycle, Amino acids, Nucleotide phosphates and Other metabolites.

Class	Abbreviation	Metabolite
Glycolysis	1,3BPG	1,3-Biphosphoglycerate
	2PG	2-Phosphoglycerate
	3PG	3-Phosphoglycerate
	DHAP	Dihydroxyacetone phosphate
	F1,6BP	Fructose 1,6-biphosphate
	F6P	Fructose 6-phosphate
	G6P	Glucose 6-phosphate
	GA3P	Glyceraldehyde 3-phosphate
	Glc	Glucose
	PEP	Phosphoenolpyruvate
	Pyr	Pyruvate
PPP	6PG	6-Phosphogluconate
	6PGL	6-Phosphogluconolactone
	E4P	Erythrose 4-phosphate
	R5P	Ribose 5-phosphate
	RL5P	Ribulose 5-phosphate
	S7P	Sedoheptulose 7-phosphate
	X5P	Xylulose 5-phosphate
TCA cycle	AcCoA	Acetyl-CoA
	aKG	α-Ketoglutarate
	Cit	Citrate
	Fum	Fumarate
	ICit	Isocitrate
	Mal	Malate

Continued on next page

Table A.1 – continued from previous page

Class	Abbreviation	Metabolite
	OAA	Oxaloacetate
	Suc	Succinat
	SucCoA	Succinyl-CoA
Amino acids	Ala	Alanine
	Arg	Arginine
	Asn	Asparagine
	Asp	Aspartic acid
	Cys	Cysteine
	Gln	Glutamin
	Glu	Glutamic acid
	Gly	Glycine
	His	Histidine
	Ile	Isoleucine
	Leu	Leucine
	Lys	Lysine
	Met	Methionine
	Phe	Phenylalanine
	Pro	Proline
	Ser	Serine
	Thr	Threonine
	Trp	Tryptophan
	Tyr	Tyrosine
	Val	Valine
Nucleoside Phosphates	ATP	Adenosine triphosphate
	ADP	Adenosine diphosphate
	AMP	Adenosine monophosphate
	GTP	Guanosine triphosphate
	GMP	Guanosine monophosphate
	CTP	Cytidine triphosphate
	CDP	Cytidine diphosphate
	CMP	Cytidine monophosphate
	UTP	Uridine triphosphate
	UDP	Uridine diphosphate
	UMP	Uridine monophosphate
	dATP	Deoxyadenosine triphosphate
	dADP	Deoxyadenosine diphosphate
	dAMP	Deoxyadenosine monophosphate

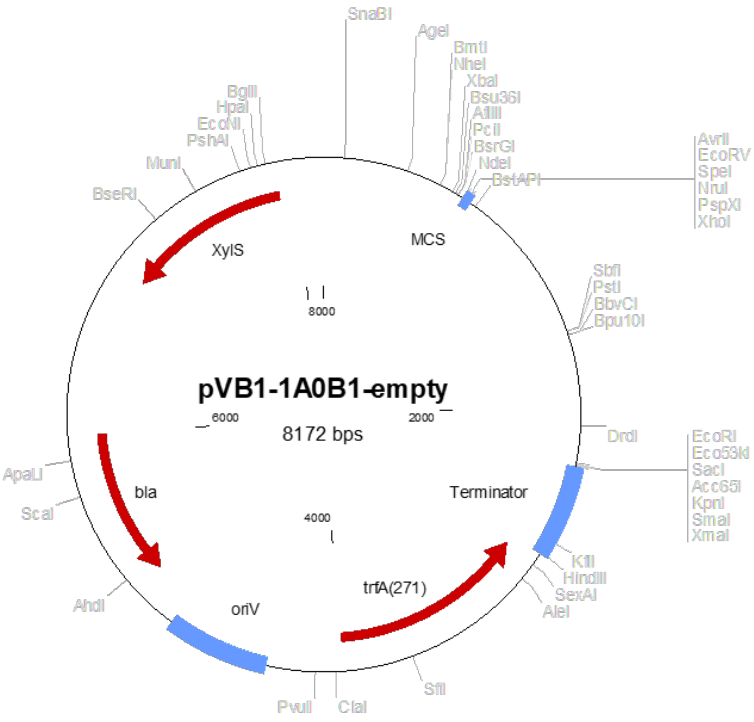
Continued on next page

Table A.1 – continued from previous page

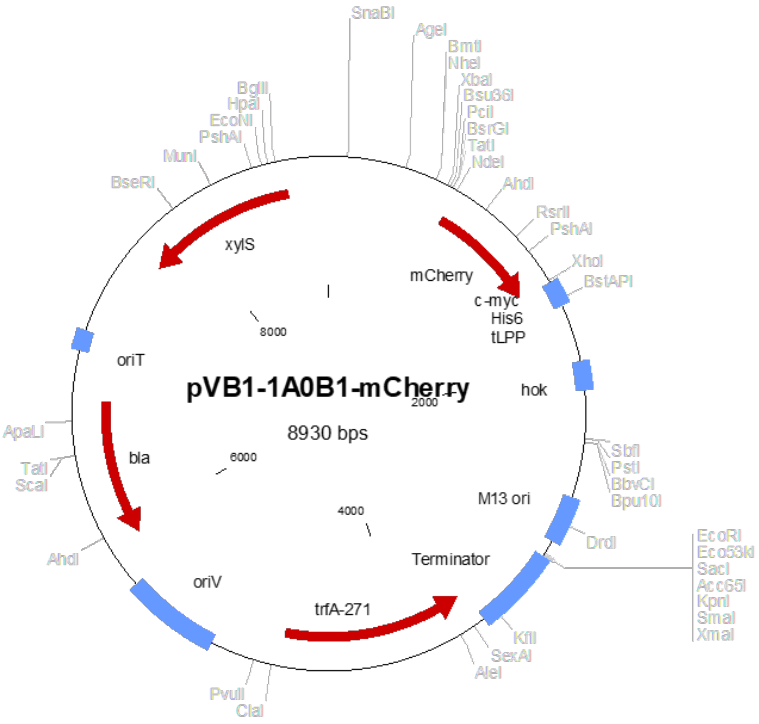
Class	Abbreviation	Metabolite
	dGTP	Deoxyguanosine triphosphate
	dGDP	Deoxyguanosine diphosphate
	dCTP	Deoxycytosine triphosphate
	dUMP	Deoxyuridine monophosphate
	dTTP	Deoxythymidine triphosphate
	dTDP	Deoxythymidine diphosphate
	dTMP	Deoxythymidine monophosphate
Other metabolites	2- IPPMal	2-Isopropylmalate
	2-OHGlc	2-Hydroxyglutarate
	F1P	Fructose 1-phosphate
	G1P	Glucose 1-phosphate
	GAL1P	Galactose 1-phosphate
	GAL6P	Galactose 6-phosphate
	GL3P	Glycerol 3-phosphate
	IA	Itaconic acid
	Lac	Lactate
	M1P	Mannose 1-phosphate
	M6P	Mannose 6-phosphate
	PRPP	Phosphoribosyl pyrophosphate
	UDP-Gal	UDP-galactose
	UDP-Glc	UDP-glucose
	UDP-GlcNac	UDP-N-acetylglucosamine

Appendix B

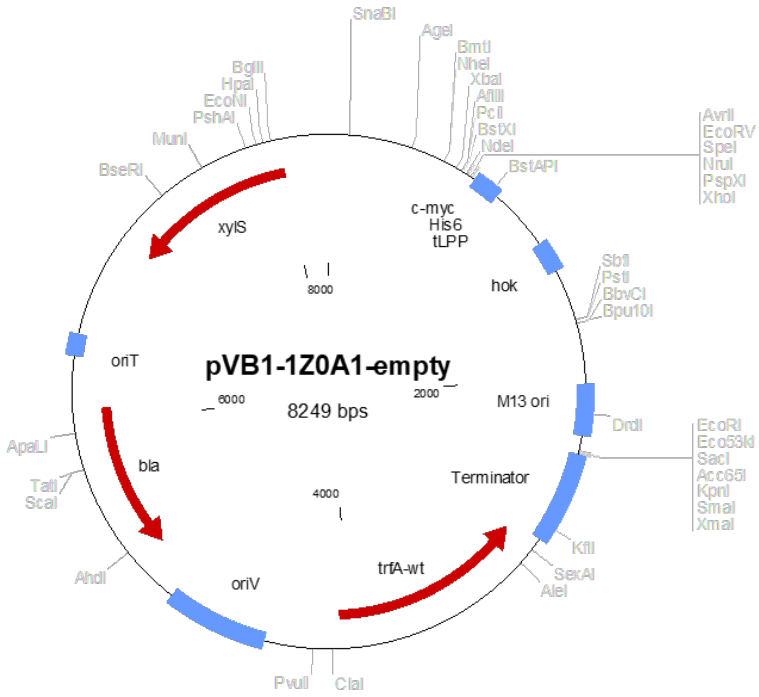
Plasmid maps



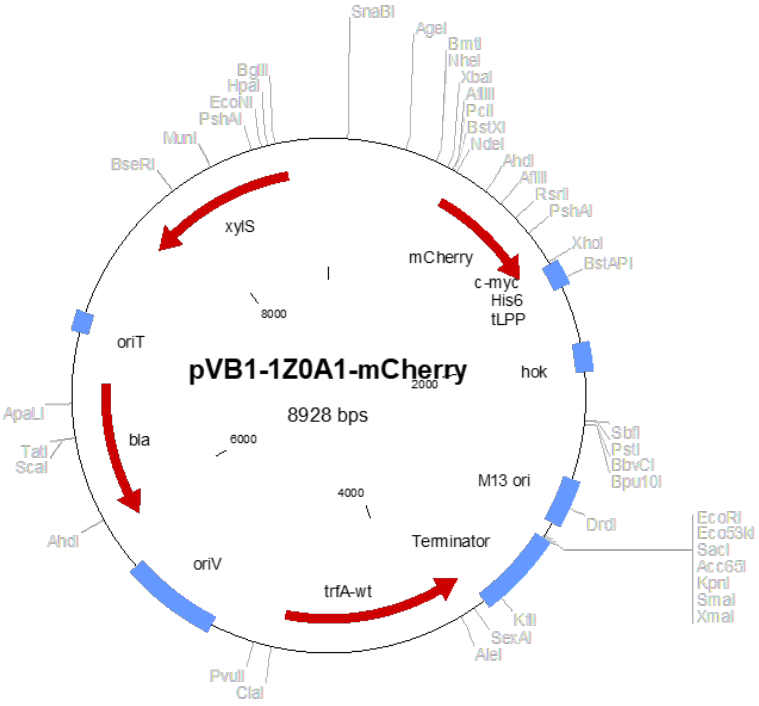
(a) *E. coli* BL21 A2



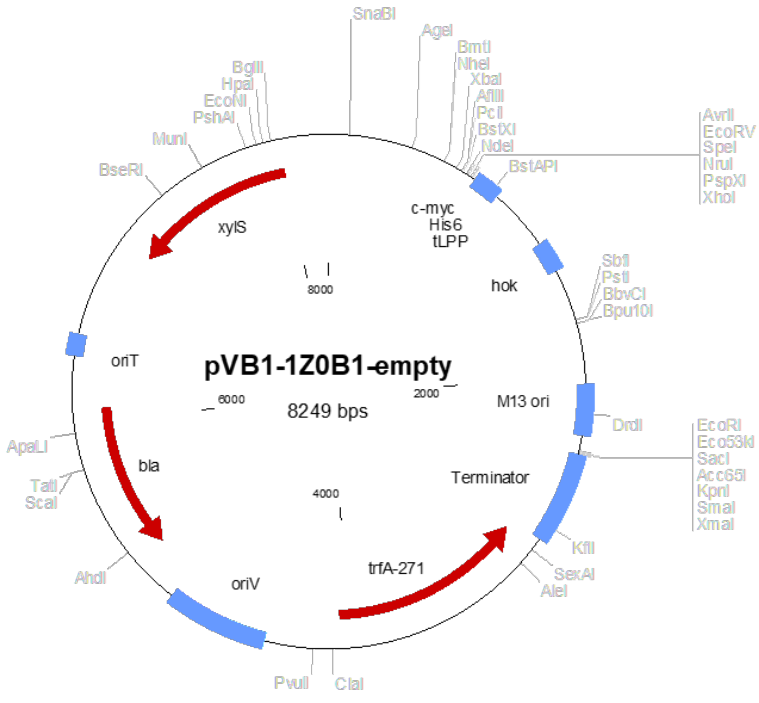
(b) *E. coli* BL21 A2-mCh



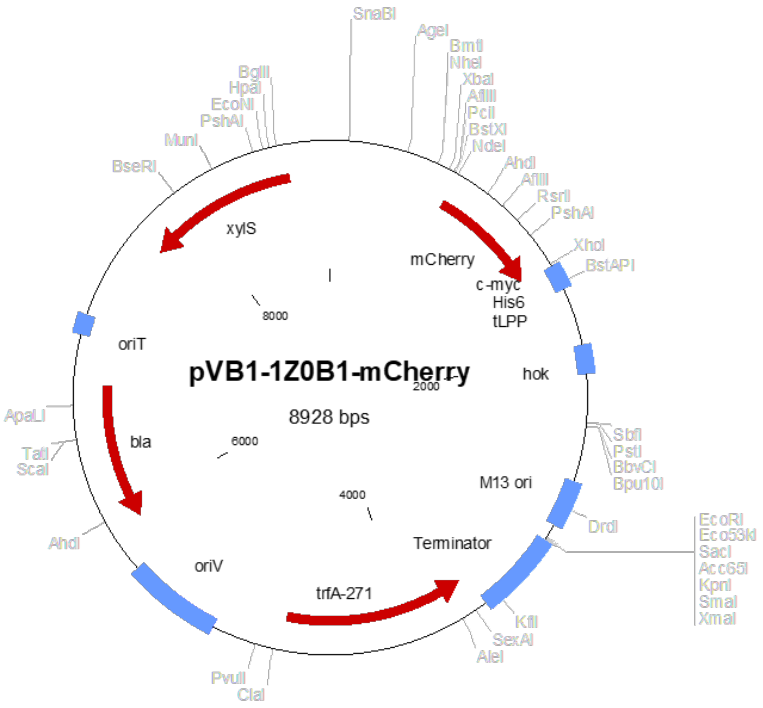
(c) *E. coli* BL21 Z1



(d) *E. coli* BL21 Z1-mCh



(e) *E. coli* BL21 Z2



(f) *E. coli* BL21 Z2-mCh

Figure B.1: Plasmid maps of the expression vectors. All expression vectors contain the *bla* gene encoding ampicillin resistance, *XylS* gene encoding the transcriptional regulator XylS, and a mutated or WT *trfA* gene which controls the PCN of the expression vector, all coloured in red. The mCherry producing strains also have the mCherry encoding sequence, also colored in red. The name of the recombinant strain transformed with the expression vector is given in the Subfigure captions.

Appendix C

Determination of growth rate

The growth rates (μ) were determined by performing exponential regression in Excel on the linear parts of the growth curves from cultivation experiments. These linear parts of the growth curves represents the exponential growth phase. The exponential regression gives a function in the form of $y = A \cdot e^{bx}$, where b represents the bacteria's growth rate.

An example determination of growth rate of *E. coli* is given in Figure C.1. The exponential regression was performed on the linear range before induction of the growth curve obtain during cultivation with endometabolome sampling of *E. coli* BL21 Z2-mCh (Figure 3.5g).

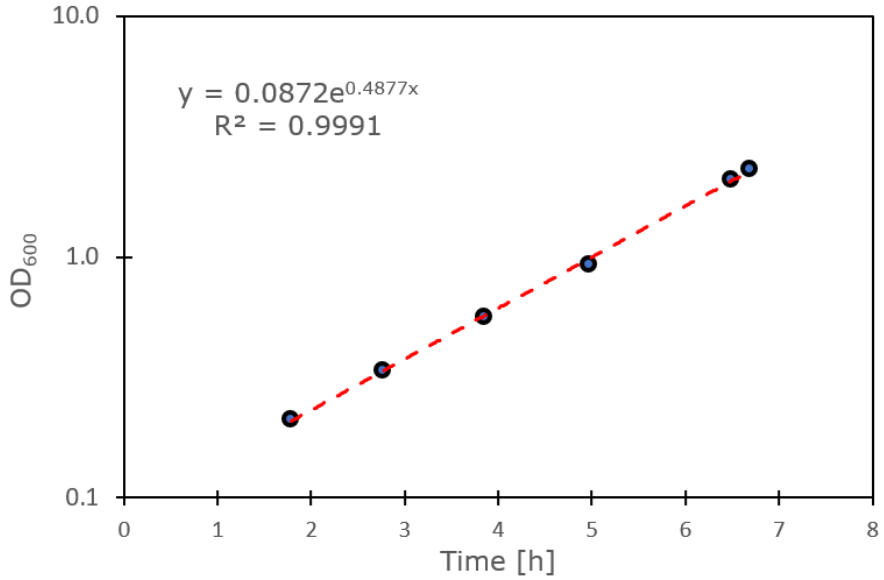


Figure C.1: Example determination of growth rate. The figure shows the results from an exponential regression performed on the linear range before induction of a growth curve, obtained during a cultivation of *E. coli* BL21 Z2-mCh. The exponential regression resulted in the function $y = 0.0872 \cdot e^{0.4877x}$ with $R^2 = 0.9991$, which corresponds to a growth rate of $\mu = 0.4877\ h^{-1} \sim 0.49\ h^{-1}$.

Appendix D

Calibrations

D.1 Biomass calibration

The calibration between OD_{600} and cell dry weight was performed on the strains *E. coli* BL21 WT, Z2 and Z2-mCh, as described in Subsection 2.4.2. The correlations between DCW and OD_{600} were found by performing linear regression of the average DCW measurements as a function of the corresponding average OD_{600} measurements. The blank (Hi medium, Table 2.6) was subtracted from the DCW measurements. The linear regressions analyses were performed with the python script given in Appendix D.3. The results from the linear regression analyses are given in Table D.1.

Table D.1 DCW as a function of OD_{600} for the strains *E. coli* BL21 WT, Z2 and Z2-mCh, together with the R^2. The linear regression analyses were performed with the Python script given in Appendix D.3.

Strain	**DCW as a function of OD_{600}**	**R^2**
WT	y = 0.516x + 0.677	0.990
Z2	y = 0.524x + 0.280	0.991
Z2-mCh	y = 0.507x + 0.670	0.985

The resulting biomass correlations are very similar, thus it was decided to use the WT linear regression function (y = 0.516x + 0.677) for all strains.

D.2 mCherry calibration

The calibration between mCherry concentration and fluorescence was performed as described in Subsection 2.4.3. First, a calibration curve of mCherry concentration as a function of fluorescence was prepared as described in Subsection 2.4.3.1. Then a calibration curve of the fluorescence of a lysed cell suspension as a function of the fluorescence of a cell suspension with whole cells was prepared as described in Subsection 2.4.3.2. The calibration curves were found by performing linear regression with the python script given in Appendix D.3. The results from the linear regression analyses is shown below.

The linear regression of fluorescence lysed cells as a function of fluorescence whole cells resulted in the following function.

$$y = 1.125x - 229.938, \qquad R^2 = 1.000$$

This results shows that the fluorescence of lysed and whole cell suspensions are different.

The linear regression of mCherry concentration [ng/µL] as a function of fluorescence of lysed cells resulted in the following function.

$$y = 0.0006x + 11.039 \ \text{ng/µL}, \qquad R^2 = 0.991$$

The two calibration curves were used to convert fluorescence measurements from the cultivation experiments to an estimation of the concentration of mCherry. An example calculation is shown below from the average fluorescence measured at sampling point T3 from the cultivation with endometabolome sampling (Figure 3.5e) of *E. coli* BL21 A2-mCh.

$$\text{Fluorescence measured during cultivation} = 12693.00$$

$$\text{Fluorescence lysed cell suspension} = 1.125 \cdot (12693.00) - 229.938 = 14049.687$$

$$\text{mCherry concentration} = 0.006 \cdot (14049.687) + 11.039 = 95.3371 \ \text{ng/µL} \approx 95.34 \ \text{µg/mL}$$

D.3 Linear regression Python script

The script uses a function in the SciPy library to perform the linear regression (Virtanen *et al.*, 2020).

```
#-------------------------------------------------------------------------
# Author; Sigurd Tommerberg Humlebrekk
#
# The script is used to perform linear regression analysis
#-------------------------------------------------------------------------

import pandas as pd
from scipy import stats

def linear_regression(Info, x, y):
    # The function performes a linear regression analysis,
    #and prints y as a function of x, and  the R^2 value
    slope, intercept, r_value, p_value, std_err = stats.linregress(x,y)

    if intercept > 0:
```

```
        pm = '+'
    else:
        pm = '-'

    print (Info,'\t' ,'y={:.3f}x'.format(slope),pm,'{:.3f}'.format(abs(
    intercept)),'R\u00b2 = {:.3f}'.format(r_value**2))

print('\nBiomass Calibration:\n')

Biomass_calibration_df = pd.read_csv('220122_Biomass_calibration_test.csv
    ',
                                     delimiter=';')
#Calling on the linear_regression function with Biomass calibration data
linear_regression('WT',
                  Biomass_calibration_df['WT_avg_OD'],
                  Biomass_calibration_df['WT_avg_DCW'])
linear_regression('Z2',
                  Biomass_calibration_df['Z2_avg_OD'],
                  Biomass_calibration_df['Z2_avg_DCW'])
linear_regression('Z2-mCh',
                  Biomass_calibration_df['Z2-mCh_avg_OD'],
                  Biomass_calibration_df['Z2-mCh_avg_DCW'])



print('\nmCherry Calibration:')

mCherry_calibration_df = pd.read_csv('210107_mCherry_calibration.csv',
                                     delimiter=';')
#Calling on the linear_regression function with mCherry calibration data
linear_regression('\nFluorescence whole to lysed cells:\n',
                  mCherry_calibration_df['Fluorescence whole cells'],
                  mCherry_calibration_df['Fluorescence lysed cells'])

linear_regression('\nFluorescence to mCherry concentration [ng/ \u03BC L
    ]:\n',
                   mCherry_calibration_df['mCherry fluorescence'],
                   mCherry_calibration_df['mCherry concentration'])


print('\nBradford standard curve: ')

Bradford_df = pd.read_csv('220205_Bradford_standard_curve.csv',
                                     delimiter=';')
linear_regression('\OD595 to protein concentration [\u03BC g/ \u03BC g]\n
    ',
                   Bradford_df['BSA_OD'],
                   Bradford_df['BSA_Concentration'])

```

```
""" Output

Biomass Calibration:

WT        y=0.516x + 0.677 R2 = 0.990
Z2        y=0.524x + 0.280 R2 = 0.991
Z2-mCh    y=0.507x + 0.670 R2 = 0.985

mCherry Calibration:

Fluorescence whole to lysed cells:
          y=1.125x - 229.938 R2 = 1.000

Fluorescence to mCherry concentration [ng/uL]:
          y=0.006x + 11.039 R2 = 0.991

Bradford standard curve:
\OD595 to protein concentration [ug/ug]
          y=21.241x - 0.026 R2 = 0.991
"""
```

Appendix E

Dixon's Q-test

Dixon's Q-test is a statistical method that can be used to identify potential outliers in data sets with small numbers of replicas or observations (Dean and Dixon, 1951). The method assumes a Gaussian (normal) distribution of the data, and the test can only be performed once on a data set (Dean and Dixon, 1951; Efstathiou, 2006). The test is based on the calculation of the experimental Q-value, which is given by the gap between the replica to its nearest neighbour, divided by the range of the observed values (Dean and Dixon, 1951). If the data set is composed of N replicas, the replicas are sorted based on size ($x_1 < x_2 < ... < x_{N-1} < x_N$) then the test is performed on the lowest (x_1) and highest (x_N) value as shown in Equation E.1.

$$Q = \frac{x_1 - x_2}{x_N - x_1} \text{ or } Q = \frac{x_N - x_{N-1}}{x_N - x_1} \tag{E.1}$$

If the experimental Q-value surpasses the critical value Q_{crit} then the replica can be rejected with a probability based on the selected confidence level (Dean and Dixon, 1951; Efstathiou, 2006). The critical value Q_{crit} depends on the number of replicas in the data set and the chosen confidence level (Rorabacher, 1991). For the Q-tests performed in this master thesis, the confidence level was chosen to be 95%, and the sub-sample sets were composed of 3 to 5 replicas. The critical values used for the tests performed in this master thesis were taken from (Rorabacher, 1991). The critical values are listed in Table E.2. An example of Dixon's Q-test is given in Table

Table E.1 The critical values used to identify potential outliers in the endometabolome sub-sample sets, using two tailed Dixon's Q-test with a confidence level of 95 %. The values are taken from (Rorabacher, 1991).

Number of replicas	Critical value
3	0.970
4	0.829
5	0.710

Table E.2 Example of Dixon's Q-test done on the sample replicas of *E. coli* BL21 Z2-mCh T2. The table shows sum of all AA and OA, Gap, Range, Q-value, critical value and if it is an outlier or not, for all replicas.

Sample ID	Sum all AA + OA	Gap	Range	Q-value	Critical Value	Outlier Y/N
Z2mCh_T2_A	13.54475295	0.49191657	9.36605357	0.052521221	0.71	N
Z2mCh_T2_C	14.03666952					
Z2mCh_T2_D	15.41351924					
Z2mCh_T2_E	15.88983596					
Z2mCh_T2_B	22.91080652	7.02097056	9.36605357	0.749618877	0.71	Y

Appendix F

PCA analysis of metabolic data

F.1 Principal component analysis

Principal component analysis (PCA) is a simple, non-parametric multivariate statistical technique (Abdi and Williams, 2010; Shlens, 2014). The technique has roots back to 1901, when Person described a method to find "lines and planes of closest fit to systems of points in space" (Pearson, 1901). Since then, PCA has been further developed, and the modern PCA method has been said to be formalized by Hotelling, who also introduced the term principal component (Hotelling, 1933; Abdi and Williams, 2010). The technique takes in a data set of observations described by dependent, inter-correlated variables. PCA isolates the important information from this data set by performing linear combinations of the original variables, and represents this information in the form of orthogonal variables, termed principal components (PC) (Abdi and Williams, 2010; Shlens, 2014). The first PC (PC1) describes the largest possible variation of the data set. PC2 is computed to be orthogonal to PC1, and to explain the largest possible variance in the data set not described by PC1. The patters of similarity between observations and variables can be displayed as points in a map (Abdi and Williams, 2010). In this thesis, the patterns of similarity in metabolic data between the different strains were displayed in the form of 2D score plots.

The goal of PCA is to extract the most important information from the data set, simplify the data by compressing the dimensions and size of the data set, and to analyze the structure of the variables and observations (Abdi and Williams, 2010; Shlens, 2014). The technique does have some assumptions. Firstly, PCA assumes linearity in the data set. Secondly, it is assumed that large variances have important structure, while lower variance represents noise. Lastly, it is assumed that the principal components are orthogonal (Shlens, 2014).

F.2 PCA python script

The PCA analysis was performed using the python script shown below. The script uses the Scikit-learn library to perform the PCA analysis (Pedregosa *et al.*, 2011).

```
#-----------------------------------------------------------------
#Author; Sigurd Tommerberg Humlebrekk.

#This script takes in a data set of metabolic data of the strains E. coli
    BL21 WT, A2, A2-mCh, Z2 and Z2-mCh. The script performs PCA on the
   metabolic data, generates a 2D score plot, and prints out the 10
   metabolites contributing the most to the variance in PC1 and their
   loading scores.
#-----------------------------------------------------------------

import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA


def pca_function(df, features):

    #Separating out the features and autoscale the data
    x = df.loc[:, features].values
    x = StandardScaler().fit_transform(x)

    #Performes a PCA analysis on the autoscaled data set
    pca = PCA(n_components=2)
    principalComponents = pca.fit_transform(x)
    principalDf = pd.DataFrame(data = principalComponents,
                               columns = ['PC.1', 'PC.2'])

    #Adding the saple names and groupings into the data set
    Df_PCA = pd.concat([principalDf, df[['Group', 'Sample']]], axis = 1)
    return pca, Df_PCA


def loading_scores(pca, metabolites):
    # The function prints out the 10 features that contributes the most
   to the variance in PC1, and their loading scores.

    loading_scores = pd.Series(pca.components_[0], index=metabolites)
    sorted_loading_scores = loading_scores.abs().sort_values(ascending=
   False)
    top_10_metabolites = sorted_loading_scores[0:10].index.values
    print(loading_scores[top_10_metabolites])


```

```
def Make_figure(pca, finalDf, targets, color, markers, save_name):
    # The function makes a 2D score plot from the results of the PCA analysis, and saves it as a png with the input save_name

    fig = plt.figure(figsize = (8,8))
    ax = fig.add_subplot(1,1,1)

    for target, color,marker in zip(targets,colors,markers):
        indicesToKeep = finalDf['Group'] == target
        ax.scatter(finalDf.loc[indicesToKeep, 'PC.1']
               , finalDf.loc[indicesToKeep, 'PC.2']
               , c = color
               , s = 120
               , marker = marker)

    for i in range(len(finalDf['Sample'])):
        ax.annotate(s=finalDf['Sample'][i], xy = (finalDf['PC.1'][i] ,
                                                  finalDf['PC.2'][i]), size=12)

    ax.set_xlim(min(finalDf['PC.1'])-3.5,max(finalDf['PC.1'])+3.5)
    ax.set_ylim(min(finalDf['PC.2'])-1.5,max(finalDf['PC.2'])+1.5)

    ax.set_xlabel('PC.1 ({:.2f})%'.format(pca.explained_variance_ratio_[0]*100)
                  , fontsize=13)
    ax.set_ylabel('PC.2 ({:.2f}%)'.format(pca.explained_variance_ratio_[1]*100)
                  , fontsize=13)
    ax.legend(targets,loc='lower center', bbox_to_anchor=(0.5,-0.155), ncol=5
              , fontsize ='large')
    ax.grid(alpha=0.5)

    plt.savefig(save_name,dpi=300,bbox_inches='tight')
    plt.show()


# The features in the data set
metabolites = ['Lac','Pyr','Mal','Suc', 'Fum', 'ICit', 'aKG','Gly', 'Ala', 'Ser','Pro', 'Val', 'Thr', 'Leu', 'Ile', 'Asn', 'Asp', 'Gln', 'Glu', 'Met', 'His', 'Phe', 'Arg', 'Tyr', 'Trp', 'Lys',]

# Groupings, colors and markers to generate the 2D score plot
groups = ['WT', 'A2', 'A2-mCh', 'Z2', 'Z2-mCh']
colors = ['tab:green', 'tab:purple', 'tab:blue','tab:orange','tab:red']
markers = ['o', '*', 'X', 'd', 's']

# Data file
```

```
df = pd.read_csv('220203_All_strains_PCA.csv', delimiter = ';', header=0)

# Calling the functions
pca, df_PCA = pca_function(df, metabolites)
loading_scores(pca,metabolites)
Make_figure(pca, df_PCA, groups, colors, markers,
            '220208_All_strains_PCA.png')
```

Appendix G

Statistical analysis of microscopy data

G.1 Hypothesis testing

There are two types of hypotheses in hypothesis testing, the null hypothesis (H_0) which states that both means are statistically equal, and the alternative hypothesis (H_1) which states that the means are not statistically equal (Mishra *et al.*, 2019). The P-value describes the probability that the effect is at least as large as observed if H_0 is true. Therefore, the lower the P-value, the more evidence for rejecting H_0 (Thiese *et al.*, 2016). In this thesis, the level of significance was set to 0.05, which means P-values below 0.05 leads to rejection of H_0.

G.2 T-test

T-test is a statistical method for comparing the means of two groups (Kim, 2015). T-test can be divided into three categories, one sample t-test, independent t-test and paired t-test. One sample t-test is used to test if a sample from a parent group has a significantly different mean from the parent group. Independent t-test is used when the groups tested are independent of each other, while the paired t-test is used when the groups tested are dependant of each other (Mishra *et al.*, 2019). The assumptions for t-tests are that the groups compared are normally distributed, the two groups have equal variance, and the data used for the t-test are either sampled independently from the two groups (independent t-test), or the samples should be fully paired (paired t-test) (Kim, 2015). In this thesis, independent t-test was conducted to test for statistically significant between two independent biological replicas of strains, and paired t-test was performed to test for significant difference between the time points T1 and T2 of the same strain.

G.3 One-way Analysis of variance

Analysis of variance (ANOVA) is a statistical method that can be used to compare the means of three or more groups (Mishra *et al.*, 2019). One-way ANOVA is a form of ANOVA where 3 or more independent groups are compared based on one fixed effect or variable. One-way ANOVA can be used to test H_0, which states that all groups have the same mean (Larson, 2008). The statistical method is based on four assumptions. Firstly, all observations are independent. Secondly, the data can be represented by a statistical model with additive components. Thirdly, all groups follow a normal distribution, and lastly, all groups have identical variance (Larson, 2008; Mishra *et al.*, 2019). The results from ANOVA is a common P-value, where a significant P-value indicate that at least two groups in the data set have significant different means from eachother. The method does not indicate which groups have significant different means, therefore a pair-wise multiple comparisons (post hoc) test must be performed to identify the significant pair(s) (Mishra *et al.*, 2019). One such post hoc test is Tukey's method, also termed the honestly significant differences (HSD) method. Tukey's HSD compares the means of all groups to every other group (Brown, 2005). In this thesis, one-way ANOVA was used to determine if there were statistical differences in the means between the strains at both time points (T1 and T2), and Tukey's HSD was used to determine which pair(s) of strains had significant different means.

G.4 Results from the statistical analysis of microscopy data

The microscopy procedure used to measure the cell length of *E. coli* WT, A2, A2-mCh, Z2 and Z2-mCh is described in Section 2.7. Two independent biological replicas (A and B) was studied for each strain. The cell length of at least 100 cells per replica was measured both before induction at OD 1 (T1), and after induction at OD 2.5 (T2). The level of significance for all statistical tests were set to 0.05. If the resulting P-values is below 0.05, then H_0 is rejected. The python script used to perform all the statistical tests is given in Appendix G.5.

First, independent t-tests were used to compare the mean cell length between the biological replicas of the same strain at both time points. The resulting p-values are given in Table G.1. No significant difference was observed in mean cell length between the biological replicas were observed. The data of the two biological replicas were combined for each time point, and paired t-test was used to test for significant difference in mean cell length between T1 and T2. The resulting p-values are given in Table G.1. The paired t-test comparing T1 and T2 of Z2-mCh resulted in the P-value 0.0248. This P-value is lower than the level of significance at 0.05, thus meaning Z2-mCh had significantly different mean cell length at T2 compared to T1. All other strains had statistically equal mean cell length at T1 and T2, as their paired t-test resulted in P-values above 0.05.

One way analysis of variance (One-way ANOVA) was then used to compare the mean cell length of all strains at T1, and at T2. The test showed no significant differences in mean cell length between the strains at T1, with a P-value of 0.371532. At T2, the test resulted in a significant difference in mean cell length of at least two strains, with a P-value of 0.026174. Tukey's HSD was then performed to identify which strains had significant different mean cell length from each other by comparing all strains against each other. The P-values from Tukey's HSD is shown in Table G.2. The test showed that WT and Z2-mCh had significant different mean cell length at T2, with a P-value of 0.013133. All other strains compared had P-values over 0.05, thus had statistical equal mean cell length.

Table G.1 The P-values from t-tests performed on data from the cell length study. The table shows the P-values from independent t-test of biological replicas A and B, at T1 and T2, and the P-values from the paired t-test of T1 and T2. A P-value below 0.05 means the groups compared have a significantly different mean cell length.

Strain	**Independent t-test of Replica A and B at T1**	**Independent t-test of Replica A and B at T2**	**Paired t-test of T1 and T2**
WT	0.7100	0.9175	0.9259
A2	0.3090	0.2129	0.2936
A2-mCh	0.1647	0.8261	0.2803
Z2	0.1798	0.7646	0.6536
Z2-mCh	0.3883	0.7264	0.0248

Table G.2 The results from Tukey's HSD on the mean cell length of all strains at T2. The table shows the strains compared and their corresponding P-value. A P-values of 0.9 should be interpret as P-value equal to or greater than 0.9.

Strains	**compared**	**P-value**
WT	A2	0.705405
WT	A2-mCh	0.481113
WT	Z2	0.900000
WT	Z2-mCh	0.013133
A2	A2-mCh	0.900000
A2	Z2	0.900000
A2	Z2-mCh	0.293265
A2-mCh	Z2	0.900000
A2-mCh	Z2-mCh	0.523871
Z2	Z2-mCh	0.126740

G.5 Python script for statistical test of cell length data

The python script used to perform the statistical tests of microscopy data from the cell length study. Functions from the SciPy library was used to perform the independent t-tests and paired t-tests (Virtanen *et al.*, 2020). Functions from the Bioinfokit library (Bedre, 2020) was used to perform one-way ANOVA and Tukey's HSD.

```
#-------------------------------------------------------------------------
# Author: Sigurd Tommerber Humlebrekk
#-------------------------------------------------------------------------
#The script is used for statistical testing of data from a cell length
    study of the strains E. coli BL21 WT, A2, A2-mCh, Z2, and Z2-mCh.
#An independent t-test is used to compare the mean cell length of two
    biological replicas of the same strain, at the same timepoint.
#A paired t-test is used to compare the strain's mean cell length between
     the two timepoints
#One-way ANOVA is used to compare the mean cell length of all strains at
    both timepoints. If the analysis identifies significant different mean
     cell length, then Tukey's HSD can be used to compare all strains to
    find which strains have significant different mean cell length from
    each other.
#-------------------------------------------------------------------------

import pandas as pd
import scipy.stats as stats
from bioinfokit.analys import stat

df = pd.read_csv('220207_Morphology_data.csv', delimiter=';')


strains = ['WT', 'A2', 'A2-mCh', 'Z2', 'Z2-mCh']

def Independent_ttest(strain):
     # Conducts independant t-test and prints the p-value
     # Null hypothesis, H0 = the biological replicas have identical mean
    cell lengths
     # If the p-value is smaller than the threshold of 5% (0.05),
     # then the Null hypothesis is rejected

    t_T1, p_value_T1 = stats.ttest_ind(df[strain+'_T1_A'],
                                       df[strain+'_T1_B'])
    t_T2, p_value_T2 = stats.ttest_ind(df[strain+'_T2_A'],
                                       df[strain+'_T2_B'])

    H0_T1 = 'H0 not rejected'
    H0_T2 = 'H0 not rejected'

    if p_value_T1 < 0.05:
        H0_T1 = 'H0 rejected'
```

```
35
36      if p_value_T2 < 0.05:
37          H0_T2 = 'H0 rejected'
38
39      print('\n', strain,
40            '\n T1 P-value = {0:.4f},'.format(p_value_T1), H0_T1,
41            '\n T2 P-value = {0:.4f},'.format( p_value_T2), H0_T2)
42
43  print('Independent t-test of biological replicas:')
44  for strain in strains:
45      Independent_ttest(strain)
46
47
48  def paired_ttest(strain):
49       # Conducts paired t-test and prints the p-value
50       # Null hypothesis, H0 = the strains have identical mean cell length
     at both timepoints
51       # If the p-value is smaller than the threshold of 5% (0.05),
52       # then the Null hypothesis is rejected
53
54      t, p_value = stats.ttest_rel(
55                           df[strain+'_T1_A'].append(df[strain+'_T1_B']),
56                           df[strain+'_T2_A'].append(df[strain+'_T2_B']))
57
58      H0 = 'H0 not rejected'
59      if p_value < 0.05:
60          H0 = 'H0 rejected'
61
62      print('\n', strain,
63            ' P-value = {0:.4f},'.format(p_value), H0)
64
65  print('\nPaired t-test of time points:')
66  for strain in strains:
67      paired_ttest(strain)
68
69
70
71  df_ANOVA_T1 = pd.read_csv('220207_ANOVA_T1.csv', delimiter=';')
72  df_ANOVA_T2 = pd.read_csv('220207_ANOVA_T2.csv', delimiter=';')
73
74  res = stat()
75
76  # One-way ANOVA and Tukey's HSD:
77
78  res.anova_stat(df=df_ANOVA_T2, res_var='Value', anova_model='Value ~ C(
     Strain)')
79  res.tukey_hsd(df=df_ANOVA_T2, res_var='Value', xfac_var='Strain',
     anova_model='Value ~ C(Strain)')
80
```

```
print('\n Results from one-way ANOVA\n',res.anova_summary,
'\n\n Results from Tukey HSD\n',res.tukey_summary)

"""
Output:

Independent t-test of biological replicas:

 WT
 T1 P-value = 0.7100, H0 not rejected
 T2 P-value = 0.9175, H0 not rejected

 A2
 T1 P-value = 0.3090, H0 not rejected
 T2 P-value = 0.2129, H0 not rejected

 A2-mCh
 T1 P-value = 0.1647, H0 not rejected
 T2 P-value = 0.8261, H0 not rejected

 Z2
 T1 P-value = 0.1798, H0 not rejected
 T2 P-value = 0.7646, H0 not rejected

 Z2-mCh
 T1 P-value = 0.3883, H0 not rejected
 T2 P-value = 0.7264, H0 not rejected

Paired t-test of time points:

 WT  P-value = 0.9259, H0 not rejected

 A2  P-value = 0.2936, H0 not rejected

 A2-mCh  P-value = 0.2803, H0 not rejected

 Z2  P-value = 0.6536, H0 not rejected

 Z2-mCh  P-value = 0.0248, H0 rejected

 Results from one-way ANOVA
                df        sum_sq    mean_sq         F    PR(>F)
C(Strain)     4.0     85.108092  21.277023  2.771302  0.026174
Residual    995.0   7639.239941   7.677628       NaN       NaN

 Results from Tukey HSD
    group1  group2      Diff     Lower     Upper   q-value   p-value
0       WT      A2  0.341772 -0.415442  1.098987  1.744369  0.705405
1       WT  A2-mCh  0.450783 -0.306432  1.207997  2.300746  0.481113
```

```
2       WT       Z2  0.227077 -0.530138  0.984291  1.158974  0.900000
3       WT  Z2-mCh  0.881282  0.124067  1.638496  4.497967  0.013133
4       A2  A2-mCh  0.109010 -0.648204  0.866225  0.556376  0.900000
5       A2       Z2  0.114696 -0.642519  0.871910  0.585395  0.900000
6       A2  Z2-mCh  0.539509 -0.217705  1.296724  2.753598  0.293265
7  A2-mCh       Z2  0.223706 -0.533508  0.980921  1.141772  0.900000
8  A2-mCh  Z2-mCh  0.430499 -0.326715  1.187714  2.197221  0.523871
9       Z2  Z2-mCh  0.654205 -0.103009  1.411420  3.338993  0.126740
"""
```

Appendix H

Amino acid composition of mCherry

Amino acid sequence

MVSKGEEDNMAIIKEFMRFKVHMEGSVNGHEFEIEGEGEGRPYEGTQTAKLKVTKGGPLP FAWDIL-
SPQFMYGSKAYVKHPADIPDYLKLSFPEGFKWERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLRGT-
NFPSDGPVMQKKTMGWEASSERMYPEDGALKGEIKQRLKLKDGGHYD AEVKTTYKAKKPVQLP-
GAYNVNIKLDITSHNEDYTIVEQYERAEGRHSTGGMDELYK

Amino acid composition

Table H.1 Amino acid composition of mCherry (Ransom *et al.*, 2015).

Amino acid	One letter code	Number of residues	Percentage [%]
Alanine	A	11	4.66
Cysteine	C	0	0.00
Aspartic acid	D	14	5.93
Glutamic acid	E	24	10.17
Phenylalanine	F	10	4.24
Glycine	G	25	10.59
Histidine	H	6	2.54
Isoleucine	I	10	4.24
Lysine	K	24	10.17
Leucine	L	13	5.51
Methionine	M	10	4.24
Asparagine	N	7	2.97
Proline	P	12	5.08
Glutamine	Q	8	3.39
Arginine	R	8	3.39
Serine	S	12	5.08
Threonine	T	12	5.08
Valine	V	15	6.36
Tryptophan	W	3	1.27
Tyrosine	Y	12	5.08

Sigurd Tømmerberg Humlebrekk

NTNU
Norwegian University of
Science and Technology